Technische Universität München
TUM School of Computation, Information and Technology

Novel Analytical and Learning-based Image Processing Techniques for Dose Reduction in Interventional X-ray Imaging

Sai Gokul Hariharan

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

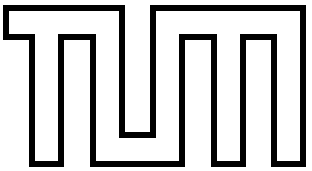
genehmigten Dissertation.

Vorsitz: Prof. Dr. Hans Michael Gerndt

Prüfer*innen der Dissertation:

1. Prof. Dr. Nassir Navab
2. Assistant Prof. Mathias Unberath, Ph.D.

Die Dissertation wurde am 13.12.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 18.05.2023 angenommen.



FAKULTÄT FÜR INFORMATIK

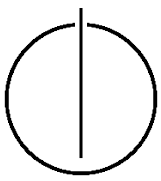
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Doctoral Thesis in Informatics

Novel Analytical and Learning-based Image Processing Techniques for Dose Reduction in Interventional X-ray Imaging

Neuartige analytische und lernbasierte Bildverarbeitungsverfahren zur Röntgendosisreduktion für die interventionelle Röntgenbildgebung

Author:	Sai Gokul Hariharan
Supervisors:	Prof. Dr. Nassir Navab
	Prof. Dr. Norbert Strobel
Mentor:	PD Dr. Markus Kowarschik
Date:	November 27, 2022



I hereby declare that this thesis is entirely the result of my own work. I have used only the resources that I have acknowledged in the references.

Forchheim, November 27, 2022 Sai Gokul Hariharan

Acknowledgements

I am deeply grateful to my supervisors, Prof. Dr. Nassir Navab and Prof. Dr. Norbert Strobel, my advisors PD Dr. Markus Kowarschik and Dr. Christian Kaethner, for providing me with an opportunity to conduct my research on developing image processing techniques for enabling X-ray dose reduction in interventional X-ray imaging under the joint aegis of Chair for Computer Aided Medical Procedures, Technische Universität München, Munich, Germany and Advanced Therapies, Siemens Healthineers AG, Forchheim, Germany. I owe a lot to Nassir and Markus for their continuous support. The discussions I had with Nassir during the workshops conducted by the Chair for Computer Aided Medical Procedures were crucial in providing direction to my thesis. I am extremely thankful to Norbert, Markus and Christian who always spared some of their precious time to offer me meticulous guidance. They have spent a lot of their time in brainstorming ideas and reviewing papers. The discussions we had were always encouraging and had a huge impact on the course of this work and on the project in general. I enjoyed our discussions very much and I am glad that I could work with them. Norbert has been much more than a supervisor during the years of my Doctoral as well as Master's thesis. He was always available when I had questions and I got to learn a lot from him.

The feedback and advice from my supervisors, advisors as well as colleagues, especially Prof. Dr. Shadi Albarqouni, at Technische Universität München and, Boris Stowasser, Dr. Markus Lendl, Dr. Richard Obler, Stefan Böhm, Dr. Philipp Bernhardt, PD Dr. Stanislav Taschenov, Anton Nekovar, Dr. Sebastian Schafer, Dr. Heiko Rohdjess, Dr. Michael Mannhart, Dr. Annette Birkhold, Dr. Sebastian Bauer, Dr. Günter Lauritsch, Andreas Berting, Heinz Horneggar Dr. Mathias John, Dr. Francesco Pisana, Dr. Uwe Fabricius, Sebastian Kaeppler and Coung Nguyen, Siemens Healthineers AG have played a major role in enabling me to achieve the goal. Thanks to the Siemens Healthineers Test Center in Forchheim for their technical support and assistance.

I would like to thank the clinical partners Universitäts Klinikum Erlangen, Erlangen, German Heart Centre Munich, Munich, Columbia University Medical Center, New York, Klinikum Barmherzige Brüder, Regensburg, and University of Wisconsin Hospital and Clinics, Madison, for the clinical data which have been crucial for evaluation purposes.

I am thankful to Prof. Dr. Rebecca Fahrig and Dr. Heiko Rohdjess for believing in me and employing me in the department Innovations, Advanced Therapies, Siemens Healthineers AG, Forchheim, Germany, and encouraging me to continue working on the topic of X-ray dose reduction and image quality improvement. Finally, I would like to express my gratitude towards my aunt, Dr. C. T. Indra for financially supporting me throughout my studies. I am deeply indebted to my parents - Hariharan Gopalan and Dr. Meenakshi Hariharan - my wife Sridivya Vajapeyajula, my relatives and all my friends for their motivation, encouragement and support. I thank Sridivya for her patience and mental support throughout my doctoral research.

Abstract

X-rays can be used for imaging a patient's anatomy as well as for navigating medical devices during an intervention. However, they may expose patients as well as clinical staff to a non-negligible amount of ionizing radiation. To keep the potential consequences to a minimum, the X-ray dose should be kept as low as reasonably achievable. Unfortunately, lowering the X-ray dose also results in a reduced signal-to-noise-ratio (SNR), i.e., a reduced image quality. To improve the situation in such cases, it is possible to apply denoising techniques that carefully consider the associated image characteristics.

The denoising of X-ray images corrupted by signal-dependent quantum noise as well as signal-independent electronic noise is often approached using noise variance stabilization techniques, such as the generalized Anscombe transform (GAT). This image preprocessing technique facilitates the use of denoising algorithms designed for white Gaussian noise. In order to perform the GAT, the system gain and the additive white Gaussian noise component need to be known. Unfortunately, predicting these parameters on-the-fly while acquiring X-ray images can be complicated as they are impacted by the patient, X-ray spectrum as well as the imaging mode of the detector. As a solution, a data-driven noise estimation method relying on a linear Poisson-Gaussian noise model is introduced in this thesis for deriving the parameters required for performing the GAT.

Subsequently, two spatio-temporal denoising methods are presented. They utilize the above mentioned noise model to exploit the non-local self-similarity in X-ray image sequences. The methods make use of temporally coherent and redundant structures across subsequent frames to reduce the noise in the images. This way, images with a higher SNR, which are otherwise associated with higher dose acquisitions, are obtained. Since low-dose images are severely corrupted by correlated noise, the Poisson-Gaussian noise model is used for identifying similar patches or pixels. The matched pixels are then processed using constrained low-rank approximations in order to reduce the associated noise.

Even though well-engineered analytical denoising methods significantly improve the image quality, applying them on high-resolution X-ray images in real-time requires cost-performance trade-offs. As a solution, a novel real-time capable learning-based denoising strategy is presented in the final part of this thesis. It utilizes noise-corrupted instances alone during the training phase. The data required for training the denoising network is generated using a model-based noise simulation approach proposed in this thesis. It takes into account the system gain, electronic noise and detector blur. During a thorough analysis it has been found that there are interesting similarities between the proposed learning-based denoising strategy and denoising via thresholding of sub-band coefficients.

A thorough evaluation of the proposed denoising methods suggests that X-ray dose can be significantly reduced without sacrificing clinically relevant information. This opportunity for X-ray dose reduction will be highly beneficial to patients and clinical practitioners. To determine and validate the dose reduction potential for existing and new applications, further clinical evaluations are required. Nevertheless, low-noise images can be obtained at acceptable X-ray dose levels. This will open doors for a variety of procedures that were not performed in the past due to the requirement of high exposure levels. Beyond clinical benefits, the dose reduction capability also enables the use of lower power X-ray tubes that do not require elaborate cooling solution, effectively facilitating lighter and more mobile X-ray imaging devices. This might improve the overall patient access to X-ray imaging.

Contents

1. Introduction

1.1. Motivation

X-ray imaging is one of the oldest and the most established imaging modality. It is widely used for the visualization of the interior of the human body, both for diagnostic and interventional purposes. In diagnostics, static images or exposures are, for example, acquired to assist in the diagnosis and treatment regarding bone structure and breast cancer. On the other hand, in an interventional setting, dynamic imaging also known as fluoroscopy or radioscopy is an essential component for many minimally invasive interventional procedures, surgeries in particular, as they facilitate a faster recovery for patients, reduce the risk of infection and also lower healthcare costs. However, X-ray imaging also exposes patients and the clinical staff, in an interventional setting, to ionizing radiation. In particular, complex procedures that last for several hours as well as the repetition of procedures (e.g., due to repeated attempts) may lead to severe irradiation of the patients and also to high accumulation of X-ray dose levels in the clinicians over a period of time. Exposure to such radiation may result in harmful effects such as hair loss, skin injury or even cancer [9]. In addition to the potential health risks caused by the ionizing radiation associated with X-ray imaging, the use of strong iodinated contrast agents for visualizing a patient's vasculature puts a burden on their kidneys [10]. Therefore, the applied radiation dose and the contrast agent must be kept as low as reasonably achievable. Although it is possible to lower the X-ray dose and the concentration of the contrast agent, the resulting loss in image quality may not be acceptable for carrying out the procedure safely without further image enhancement. Therefore, the application of advanced image processing methods to maintain the required image quality at lower X-ray dose levels and reduced contrast agent concentrations is a very important field of research.

The transition of the century-old technique from film-based to digital imaging using semiconductor flat-panel detectors has made it possible to process images before presenting them to clinicians [11, 12, 13]. In addition, technological advancements in the past decades have enabled real-time processing of X-ray images acquired during live procedures. These developments have played a major role in the design of algorithms ultimately facilitating low-dose X-ray imaging. In the case of fluoroscopic procedures, involving successive frames of X-ray images, the obvious ways to reduce the X-ray exposure are to either reduce the frame rate or to lower the X-ray exposure per frame. However, decreasing the frame rate leads to temporal undersampling resulting in jerkiness [14]. In order to smoothen the appearance of the X-ray sequences, temporal interpolation methods have been applied [14]. Continuous visualization of flowing contrast is an important requirement to analyze anatomical structures such as blood vessels. Therefore, it has been investigated whether dose reduction for such applications can be achieved by changing the X-ray intensity on a per-frame basis and then performing a weighted combination of multiple X-ray images to improve the signal-to-noise ratio [15]. Apart from this, the use of

spatial denoising of single images to improve the signal-to-noise ratio before visualization is considered to be a suitable alternative to reduce the X-ray dose.

Prominent analytical denoising approaches often utilize self-similar structures in images and typically involve sophisticated thresholding techniques. Variational approaches based on partial differential equations are also widely used for denoising. Although existing state-of-the-art analytical approaches have yielded good results for high- and standard-dose levels, their performance drops significantly for applications performed with low- and very low-dose settings. These approaches can be improved by taking into account the imaging and noise model associated with X-ray images. However, applying them in real-time involves cost-performance trade-offs. On the other hand, learning-based approaches have been shown to outperform traditional denoising techniques. However, these methods are not directly applicable to X-ray imaging as they usually require spatially aligned pairs of noisy and (ideally) noise-free images whose acquisition in a clinical scenario is practically close to impossible due to ethical concerns and patient motion.

Considering the aforementioned health risks for patients as well as clinical staff due to X-ray imaging and the shortcomings of state-of-the-art image processing techniques to allow for comparable dose reduction in real time, this thesis aims to show that analytical and learning-based approaches that take into account the noise model associated with X-ray images can significantly lower the X-ray dose and yet retain the information needed for clinical decision making.

1.2. Scientific Focus and Contributions

Image quality plays a significant role in clinical imaging. The main scientific focus of this work is to develop novel image processing methods for improving image quality to facilitate X-ray imaging at reduced radiation levels, particularly in an interventional setting. To this end, analytical and learning-based approaches that take into account the noise model associated with X-ray images are proposed. The noise model is derived by accurately modeling the noise characteristics of an X-ray system. In the following list, the major scientific contributions are summarized:

- A data-driven approach for estimating the parameters required to stabilize the noise variance in X-ray images has been proposed in [1]. The method takes into consideration an X-ray imaging model to estimate the variance of signal-dependent noise that is needed for performing a noise variance stabilization. The method has been evaluated using phantom images with respect to beam hardening and denoising performance.

- In [2], a patch-based spatio-temporal denoising algorithm that exploits non-local self similarity in X-ray sequences has been presented. The method carefully takes into account the characteristics of noise associated with X-ray images in order to achieve robust patch matching and thereby improving the performance of denoising based on low-rank approximation.

- In [3], a constrained weighted rank-one approximation for performing the spatio-temporal denoising of digital subtraction angiography sequences – acquired at low

X-ray dose level and contrast concentration – has been presented. The method utilizes the low-rank nature of a spatially aligned temporal sequence where variation is introduced by the flow of contrast agent through a vessel tree of interest. The resulting images are free from artifacts and the inherent characteristics of images is also preserved.

- A novel noise simulation method based on an X-ray imaging model has been proposed in [4]. The method makes use of the system parameters associated with low- and high-dose X-ray image acquisitions, such as system gain and electronic noise, to preserve the noise characteristics of low-dose images. The realistic simulations are crucial for developing robust learning-based denoising algorithms.

- A novel learning-based strategy for denoising low-dose X-ray images that is based on model-based simulations of low-dose X-ray images during the training phase is presented in [5]. In addition, a data-driven normalization step that increases the robustness of the proposed approach to varying amounts of signal-dependent noise associated with different X-ray image acquisition protocols is presented.

- In [6], a thorough analysis on understanding and preventing image degradation when utilizing a learning-based denoising approach is presented. Moreover, it has been shown that viewing the results from the perspective of denoising via thresholding of sub-band coefficients can be very beneficial to get a better understanding of the learning-based denoising strategy proposed in [5].

In addition to the core contributions, peer-reviewed workshop publications [7, 8] have been made.

1.3. Outline

This thesis is a collection of articles that have been published in peer-reviewed international scientific journals and at conferences. It is structured in three parts and consists of five main chapters. Part I contains Chapters 2 and 3 that focus on the technical background of X-ray imaging and image denoising, respectively. The state-of-the-art options, available at the time of writing, are also presented. The scientific contributions are presented in Part II. They include Chapters 4, 5, and 6.

In Chapter 4, a method for estimating the X-ray imaging parameters from X-ray images is introduced. The estimated parameters can be used to perform a noise variance stabilization on the X-ray images to render images with a constant noise level. This approach serves as a basis for the novel denoising methods introduced in the following chapters.

In Chapter 5, two novel analytical spatio-temporal denoising approaches are presented. The first approach improves upon existing patch-based denoising approaches by carefully using the characteristics of noise present in X-ray images to achieve robust patch matching in ultra low-dose fluoroscopic images as well as noise reduction by computing a low-rank approximation. The second denoising approach is designed specifically for neurovascular digital subtraction angiography sequences. The method involves the application of an iterative weighted low-rank approximation to stacks of spatially aligned images. The

weights prevent the contribution of temporally mismatched pixels, caused by the inflow and outflow of contrast agent, towards the low-rank approximation.

In Chapter 6, a novel learning-based X-ray image denoising strategy is presented. The method involves the use of realistic simulations of low-dose X-ray images in the training phase. A model-based method for simulating realistic low-dose images is presented. This is followed by the learning-based denoising strategy of using noise variance-stabilized simulated images along with an initial evaluation of the method on low-dose X-ray images. In addition, an improved noise simulation approach is presented and the denoising strategy is thoroughly evaluated. The potential pitfalls of the method are analyzed along with the ways to prevent them.

Finally, in Part III, all the methods presented in this thesis are summarized and discussed.

1.4. Summary

Motivated by the necessity to reduce the X-ray dose and minimize the health risks due to X-ray radiation exposure, the objective of this thesis is to show that X-ray dose can be significantly lowered using well designed image denoising algorithms. The scientific contributions of this thesis include the development and evaluation of several image processing algorithms that have been published in peer-reviewed international journals and conference proceedings. The contributions are presented in different chapters along with the technical background to medical X-ray imaging and image denoising. Using the proposed methods, a significant reduction in the X-ray dose and the amount of iodinated contrast agent can be achieved. If the proposed methods are implemented in a clinical setting, they could have a direct impact on reducing the accompanied health risks for patients and clinical staff. Apart from clinical benefits, the dose reduction capability can enable the use of low power X-ray tubes that do not require sophisticated cooling solutions. This might facilitate lighter, more mobile and cost effective imaging systems. Such systems will improve the access of patients to X-ray imaging.

Part I.

Background

2. Medical X-ray Imaging

2.1. Introduction

For more than a century, the use of X-rays have had a significant impact on the field of medical imaging as they enabled the visualization of the patient's internal anatomy. In this chapter, the history of X-ray imaging is presented briefly. This is followed by the physical principles of X-ray imaging and their applications in the medical field that are relevant for this thesis. A comprehensive overview of these topics can be found in [16, 17, 18, 19].

2.2. A Brief History of X-ray Imaging

Inspired by the experiments of the physicist Philipp Lenard, Wilhelm Conrad Roentgen (1845-1923), a German physicist, was investigating the behavior of cathode rays (electrons) in air-evacuated cathode ray tubes with short platinum electrodes fitted at the ends at his laboratory in Würzburg, Germany. He discovered that passing a high-voltage discharge through the tube resulted in the production of radiation that could penetrate objects. He used the term 'X' to represent this unknown type of radiation. He also noticed a faint light due to the fluorescence of a small piece of paper coated with Barium platinocyanide where the brightness of the fluorescence differed depending on the material and the thickness of the objects put in front of it [20]. In recognition of his outstanding contribution to science, he was awarded the first Nobel Prize for Physics on December 10, 1901 [21].

Soon after Roentgen's discovery, the investigation of the unknown rays dominated the field of research. Charles Glover Barkla (1877-1944) discovered that X-rays were plane polarizable and that an irradiated compound emits secondary radiation. Moreover, he found the secondary spectrum to be unique for every element, independent of temperature, structure and chemical composition. For this discovery – of the characteristic Roentgen radiation of the elements – he was awarded the Nobel Prize for Atomic Physics in 1917 [22]. Herman Haga and Cornelius Werd found that X-rays could be diffracted and Arnold Sommerfeld performed a mathematical analysis of the diffraction and thereby showed that their results could be explained in terms of aperiodic impulses [23]. Max von Laue, Walter Friedrich and Paul Knipping succeeded in verifying the diffraction of X-rays and proved that they belong to a category of radiation called electromagnetic radiation [11]. William Henry Bragg and William Charles Bragg proceeded with the first measurements in X-ray spectroscopy [23]. They used an X-ray tube and a collimating slit to produce the incoming rays. In addition, they utilized various minerals such as quartz, rock salt, iron, pyrite, zincblende and calcite as three-dimentional gratings and a photographic plate or an ionization chamber as a detector. By 1913, crystal analysis with X-rays had become a standard technique. The results gave insight into the structure of the crystals as well as the nature of the anode that produced the X-rays. In 1921, Albert Einstein was

awarded the Nobel prize for his insights on the photoelectric effect, the most important way of generating contrast with X-rays in medical imaging. Another important person in this context is Arthur Holly Compton, who was awarded the Nobel prize for his work on the Compton effect, the second most important interaction of X-ray photons with matter in the energy range used for medical imaging.

Since its discovery, the focus of X-rays has been on applications, especially in the medical field. Its potential in this field was already pointed out by Roentgen who produced and published an X-ray photograph of his wife's hand. Within a month of Roentgen's announcement, doctors were using X-rays to locate bullets in human bodies and studied broken bones [23]. The first use of X-rays under clinical conditions was by John Hall-Edwards in Birmingham, England, on 12 January 1896, when he radiographed a needle stuck in the hand of an associate [24]. Around this time, doctors also used X-rays to diagnose sarcoma of the tibia in the right leg of a boy [23]. Hall-Edwards, J. R. Ratcliffe and J. H. Clayton were the first to use X-rays in a surgical procedure to guide the removal of a needle in a woman's hand [25]. Henry Cattel confirmed the importance of X-ray imaging for the diagnosis of kidney stones and cirrhotic livers [23]. Roentgen also showed its potential for material testing by taking an X-ray image of a sporting gun and pointing out the flaws in the gun barrel [26]. Several weeks after Roentgen's discovery, Ivan Romanovich Tarkhanov observed that X-rays not only enable to take pictures, but also affect the living function when he irradiated frogs and insects with X-rays [27]. Therefore, X-rays were also considered to be applicable for medial therapeutic applications. Inspired by reading that an engineer working with X-rays lost his hair, Leopold Freund used X-rays to remove the strong hair growth of a five year old girl [28]. At this period, the dangers of X-rays due to the biological interaction, were still underestimated. Problems such as skin injuries were reported as early as 1896. Elihu Thomson deliberately exposed his fingers to an X-ray tube over a period of time-half an hour per day for several days – and suffered pain, swelling and blistering. Then he documented the harm caused by X-rays to his body [29]. His documentation led to the first norms on radiation protection in 1896. In 1927, Hermann Joseph Muller showed that X-rays not only cause observable damage to the body but also lead to artificial mutations of the genes which affect one's progeny. For this discovery, he was awarded the Nobel Prize in Physiology or Medicine in 1946 [30].

On February 2, 1896, Joseph John Thomson introduced the term 'ionizing radiation' as X-rays were capable of splitting air molecules into electrically charged particles. The measure of the ionizing ability of X-rays is called exposure. The roentgen (R) is a (now obsolete) traditional unit of exposure in charge/kilogram of air. Today, the term 'dose' has three different definitions and physical units taking into consideration the ionization, energy exchange and biological effect of the radiation [11]. Absorbed dose is the concentration of energy deposited in the human tissue due to exposure. It is measured in terms of gray (Gy) which has the unit joules/kilogram. Equivalent dose addresses the impact that the radiation has on human tissue. It is measured in terms of the unit sievert (Sv). For diagnostic radiation, the absorbed dose in Gy is equal to the equivalent dose in Sv. Effective dose is a calculated value based on the absorbed dose to the organs, the harm level of the radiation and sensitivities of each organ to radiation. Effective dose is also measured in Sv [16].

2.3. Physical Principles of X-ray Imaging

In this section, the basic working principle and the physics behind the generation and detection of X-rays, relevant to this thesis, are presented.

2.3.1. X-ray Generation

X-rays are generated in an evacuated X-ray tube enclosing two electrodes, known as the cathode and the anode, when electrons are emitted at the cathode and accelerated towards the anode by high voltage and then stopped within the anode. The generation happens due to two physical processes: (i) a strong deceleration of incoming electrons in the electric field of an atomic nucleus, referred to as Bremsstrahlung and (ii) characteristic radiation in the electron shell of atoms [11]. In the following sections, these fundamental principles of X-ray generation are described in more detail.

X-ray Tube

In an X-ray tube, the electrons that are used for generating the X-rays are produced by heating a filament attached to the cathode using a small voltage and a high current. These electrons are then accelerated towards a target, a small region in the anode, by an acceleration voltage V_t. This is caused by a negative potential at the filament and a positive potential at the anode. If there were gas molecules within the tube, the electrons would react with them, slow down and produce secondary electrons [16]. In order to prevent this reaction, the two electrodes of the X-ray tube have to be sealed in vacuum. The usage of vacuum in the X-ray tubes also provides the advantage of controlling the amount and the speed of the accelerated electrons independently. To prevent the electric discharge between the electrodes, specific shapes and sizes are used for the construction of the tubes and the connecting wires are sealed inside the glass wall of the X-ray tube [16]. The basic elements of an X-ray tube are shown in Fig. 2.1(a).

Typically, for medical applications, V_t is in the range of 30 to 130 kV. The current is around 10 to 1000 mA depending on the amount of X-ray flux that needs to be generated. The target may be water-cooled as only around one percent of the power used for accelerating the electrons is converted into X-rays, whereas, the remaining 99 percent end up in the heating of the target.

The number of electrons generated at the filament by thermal emission depends on the work function or the Fermi energy level W of the metal and the temperature T of the filament. Owen Williams Richardson gave a quantitative description of the relationship between the emission current density J of electrons that are produced and the W and T with the Richardson-Dushman-equation

$$J = A_o T^2 e^{\frac{-W}{k_B T}}, \tag{2.1}$$

where A_0 is the Richardson constant and k_B is the Boltzmann constant [31]. As per this equation, the number of electrons generated has an exponential dependency on W and T. For medical imaging, it is important to ensure a constant value for J. For a metal, W is constant, whereas, T can vary. If the acceleration voltage is increased, more of the generated electrons reach the target. A saturation is reached when all the electrons that

have left the metal are accelerated to the target. At such a state, a higher value for T would result in more electrons that consistently reach the target. In other words, at such a state, the parameter T can be used to deliver a consistent beam current. High-power X-ray tubes in medical imaging are designed such that all the generated electrons reach the target.

The filament may comprise a wire made of tungsten of about 0.2 mm in diameter. The wire may be coiled to form a vertical spiral of about 0.2 cm in diameter and 1 cm or less in length. The wires connected to the cathode supply a voltage of about 10V and a current of about 3 to 5A to heat the filament. The temperature T used for heating the wire and produce electrons ranges between 2200°C and 3000°C [16]. Though alloys of tungsten are more efficient in emitting electrons, tungsten is still used in tubes because of its ductility and high melting point (3370°C). Due to the repulsive forces between the electrons, the stream of electrons spreads out and bombards a large area of the anode. This undesirable effect is prevented by the usage of cathode focusing cups made of nickel. The task of the focusing cup is to direct the stream of electrons to a small region of the anode as shown in Fig. 2.1(a).

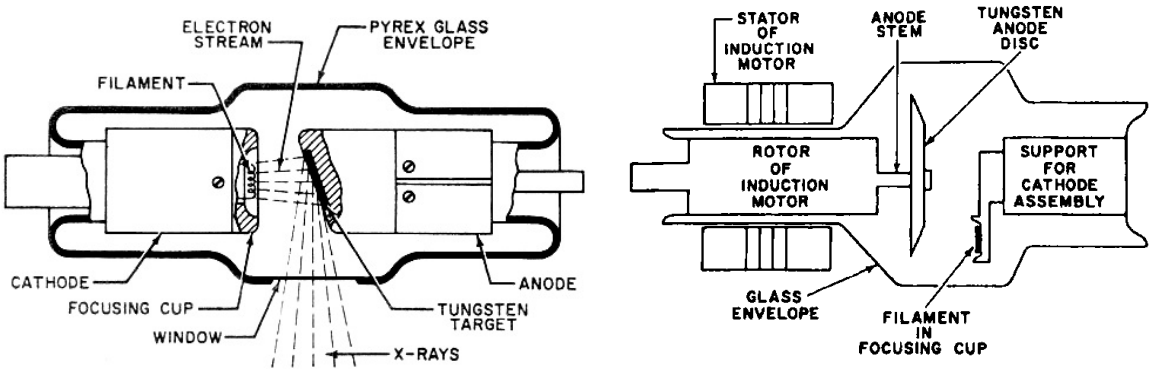


Figure 2.1.: Visualization of an X-ray tube with (a) a stationary anode [32] and (b) a rotating anode [16]. The focusing cup and the inclined anode in (a) minimize the region of the anode bombarded by the electrons from the cathode when looked at from the outside along the beam direction (focal spot). The heating of the anode caused due to the narrowing of the region bombarded can be overcome by using a rotating anode (b).

As tungsten has a high melting point as well as the ability to absorb and dissipate heat rapidly, it is also used as anode material for radiography and computed tomography (CT) applications. Anodes made of molybdenum and rhodium are used in applications such as mammography that require lower energies and a discrete spectrum.

Even though tungsten has good thermal characteristics, it cannot withstand the heat of repeated and continuous exposures. Moreover, the region of the target that is bombarded by electrons is reduced – by inclining the anode by an angle of about 6 to 20° – since a smaller focal spot produces better radiographic details. Focusing X-rays on a small region subjects the anode to significant heating. To prevent the tungsten anode from melting, an-

odes in the form of a large disc made of tungsten or its alloy, rotating at a speed of about 3000 rpm, have been introduced (Fig. 2.1(b)) [16]. This allows for the heat load to be distributed over a ring around the rotating anode. In addition, the target is also cooled by a liquid behind to remove the energy efficiently, thereby substantially increasing the total power that can be applied to an X-ray tube from around 1 kW to 150 kW and facilitating repeated X-ray pulses for longer periods like in fluoroscopy applications due to better heat spreading. Rotating anodes are used mainly in angiography due to lengthy procedures performed at high dose levels. Rotating anodes are heavier and larger, compared to stationary anodes. As a consequence, the compact mobile C-arm models, that are used for low dose and shorter X-ray examinations, usually employ stationary anodes.

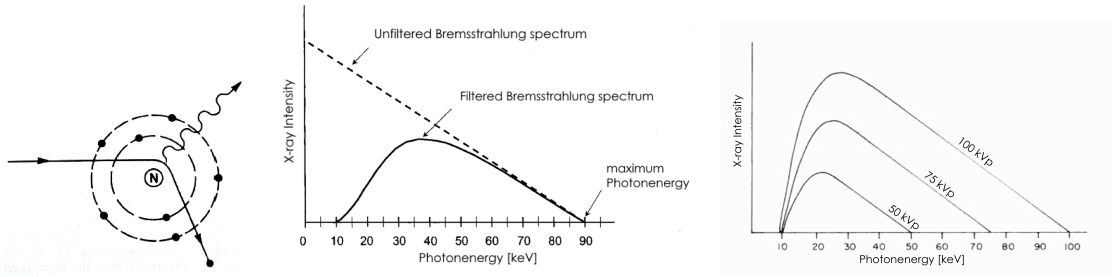
Currently, the use of liquid metal to replace the solid target is under research [33]. The idea is to hit an alloy containing indium and gallium, which is liquid at room temperature, by the electron beam to generate X-rays. This allows for adding more power per square millimeter on a fixed X-ray target as the liquid metal can be continuously replaced with new material. However, this is not yet used in medical imaging as the characteristic emission lines for gallium and indium are relatively lower than required. The emission lines for gallium are at 9.25 kilo electron volt (keV) and 10.26 keV, for indium they are at 24.2 keV and 27.2 keV, whereas for tungsten they are at 59.3 keV and 67.2 keV. If the liquid metal could be replaced with a material having a higher atomic number Z, then it could be potentially used in medical imaging.

In conventional X-ray tubes, switches used for initiating and stopping an exposure are subjected to large changes in the voltage applied between the cathode and the anode. As an alternative, grid controlled X-ray tubes make use of the focusing cup as an additional electrode, to control the flow of electrons from the filament to the target. The voltage across the filament-grid, produces an electric field along the path of the electron beam, that pushes the electrons even closer. This voltage can be made to act as a switch, to turn the tube current on and off, as a high voltage completely reduces the tube current. Since the focusing cup and the filament are placed close to each other, the voltage required to cut off the flow of electrons is not very large. For example, for a 0.3 mm focal spot tube operating at 105 kilo volt peak (kVp) about 1500 V between the filament and the cup are required to cut the electrons off [16].

Grid controlled X-ray tubes can significantly reduce the absorbed dose for the patients [34, 35]. In a grid-controlled tube, an additional electrode, known as the grid electrode, is included in the tube between the cathode and anode. The electrons accelerated from the filament to the target are made to pass through the grid electrode. When a negative voltage relative to the cathode is applied to the grid electrode, the electrons emitted from the target will get repelled and pushed back into the filament [36]. Therefore, precise control over the intensity of the X-ray beam in each pulse can be achieved using this technology. In other words, the electron flow and the X-ray production can be switched off when the resulting X-ray beam gets absorbed by the patient and will not have the required energy to reach the detector and contribute to the imaging process.

Principles related to the the generation of X-rays

Bremstrahlung and characteristic emission of X-ray fluorescence are the two main physical principles that produce X-rays at the target.



(a) (b) (c)

Figure 2.2.: The graphic (a) illustrates the generation of X-rays using Bremsstrahlung [16]. In (b), the Bremsstrahlung spectra with and without filtering are shown for an acceleration voltage of V_t = 90 kVp. In (c), the filtered Bremsstrahlung spectra corresponding to different acceleration voltages are presented.

In Bremsstrahlung, as shown in Fig. 2.2(a), incoming electrons are attracted towards the positive charge of the nucleus of the target atom. This results in the deflection of the electron from its original direction. When the electron changes direction, it loses speed and energy. The lost kinetic energy is emitted in the form of photons. As the loss of kinetic energy depends on the distance between electron and atomic core, the spectrum of the generated X-rays is continuous. In other words, the closer the incoming electron is to the atomic nucleus, the higher is the energy of the emitted X-ray photon. The maximum radiated energy of the emitted X-ray photon E_{photon} is equal to the kinetic energy $E_{\text{KE(electron)}}$ of the incoming electron. This can be represented as

$$E_{\text{photon}} = E_{\text{KE(electron)}}. \tag{2.2}$$

The variable E_{photon} can also be represented as

$$E_{\text{photon}} = h f_{\text{max}}, \tag{2.3}$$

where h is Planck's constant and f_{max} represents the maximum frequency of X-ray emission. The parameter f_{max} is related to the minimum wavelength of the emitted X-rays λ_{min} as follows

$$f_{\text{max}} = \frac{c}{\lambda_{\text{min}}}, \tag{2.4}$$

where c is the speed of light. On the other hand, $E_{\text{KE(electron)}}$ can be represented as

$$E_{\text{KE(electron)}} = eV_t, \tag{2.5}$$

where e represents the charge. From Eq. 2.2, Eq. 2.3, Eq. 2.4 and Eq. 2.5, the following relationship can be determined:

$$\lambda_{\text{min}} = \frac{hc}{eV_t}. \tag{2.6}$$

In Fig. 2.2(b), the spectrum generated by Bremsstrahlung when the target is hit with electrons that have been accelerated with $V_t = 90$ kVp is presented. It can be seen that the

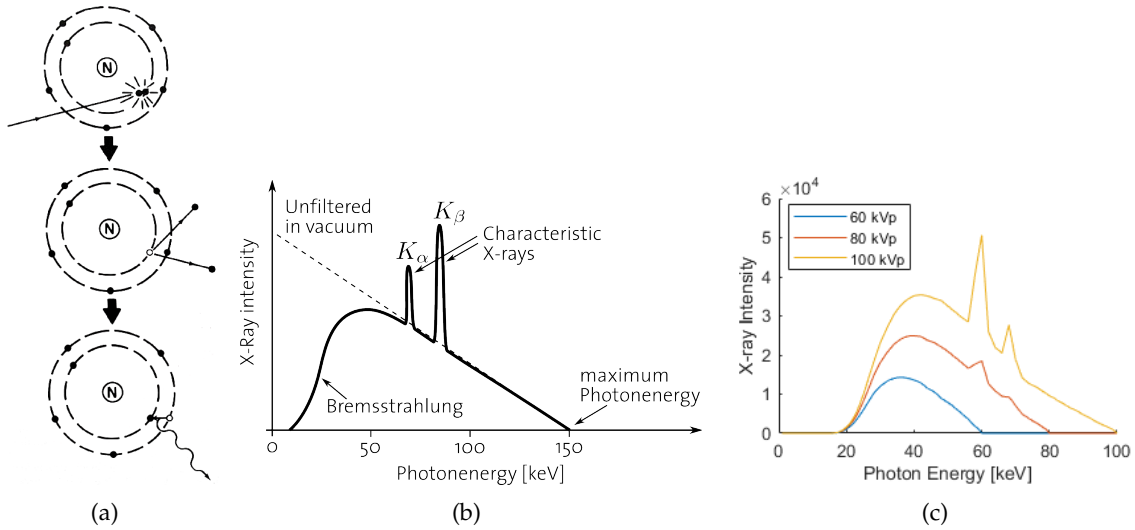
maximum energy of the emitted X-ray radiation is proportional to the maximum acceleration voltage of the electrons. However, the probability of electrons colliding head on with atomic nuclei emitting X-rays with 90 keV is small. This probability increases for lower values of keV. In theory, this probability increases linearly with a decrease in the energy of the radiated X-ray photon. This can be observed in the unfiltered Bremsstrahlung spectrum (Fig. 2.2(b)). The value is highest when the energy of the radiated X-ray photon is 0 keV. However, in practice, the increasing probability becomes stable at a point and then reduces again. This happens because the produced X-rays get filtered or absorbed, for example, by the cooling liquid surrounding the tube and by a window (made of beryllium or carbon, 300 - 500 microns thick) when it leaves the vacuum. In addition, electrons get self absorbed while hitting the target. In this case, the electrons, while reaching the target penetrate a few microns into the material. When X-rays are produced by these electrons a few microns within the target, the X-rays get absorbed by the dense target. In other words, X-rays produced deep inside the target cannot leave the X-ray tube and this also causes the Bremsstrahlung spectrum to decay towards lower keV energies. Moreover, due to anode heel effect, the diverging beams, which are parallel or near parallel to the inclined anode, get absorbed by the anode itself. Therefore the rays closer to the anode have less intensity than the rays close to the cathode end of the tube [11]. In Fig. 2.2(c), the Bremsstrahlung spectrum for three different values of V_t is presented. In all the cases, the maximum value of keV is proportional to the acceleration voltage V_t and the whole spectrum for each case is shifted accordingly. At the lower end of the spectrum, the self absorption process is similar for all the three cases. It is desired to have a cut-off at the lower energies as the X-ray photons at these energies cannot penetrate through the human body and do not contribute to the image formation. Instead, they only increase the dose absorbed by the patient.

On the other hand, in the case of X-ray generation by characteristic radiation (Fig. 2.3(a)), a high energy electron collides with an inner shell electron and displaces it, resulting in the creation of a hole. This hole is filled by a free electron or an electron from a higher energy level. This creates a difference of energy which is emitted by radiation. The resulting spectrum is discrete. This phenomenon depends on the material of the target and is related to its shell composition and can be represented using Moseley's law:

$$\frac{1}{\lambda} = R_H \left(Z - K\right)^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \tag{2.7}$$

where λ is the wavelength of the emitted X-rays, R_H is Rydberg's constant, Z is the atomic number of the material in the periodic table, K is the shielding constant and n_1 and n_2 represent the main quantum numbers of the shells involved in the transition, respectively. That is, the energies of the X-rays that would be released for a certain transition between two shells can be calculated using Moseley's law. K_α represents the transition between the inner K shell and the next L shell. Similarly, K_β represents the transition between the M and the K shells. The energy difference for K_α and K_β are called characteristic lines.

In Fig. 2.3(b) and Fig. 2.3(c), the full spectrum of a tungsten tube is presented. This includes the effect of Bremsstrahlung and characteristic radiation. The peaks in the spectrum are caused by characteristic radiation. In Fig. 2.3(c), it is to be noted that only when V_t is much greater than 60 kVp, characteristic radiation is present. When V_t is equal to or below 60 kVp there is no excitation of the characteristic lines and only Bremsstrahlung plays a role.



(a) (b) (c)

Figure 2.3.: Visualization of (a) the generation of X-rays by characteristic X-ray radiation [16], (b) the continuous Bremspectrum with characteristic lines (K_α, K_β) for tungsten for an acceleration voltage of V_t = 150 kVp and (c) the spectra for tungsten for different values of V_t.

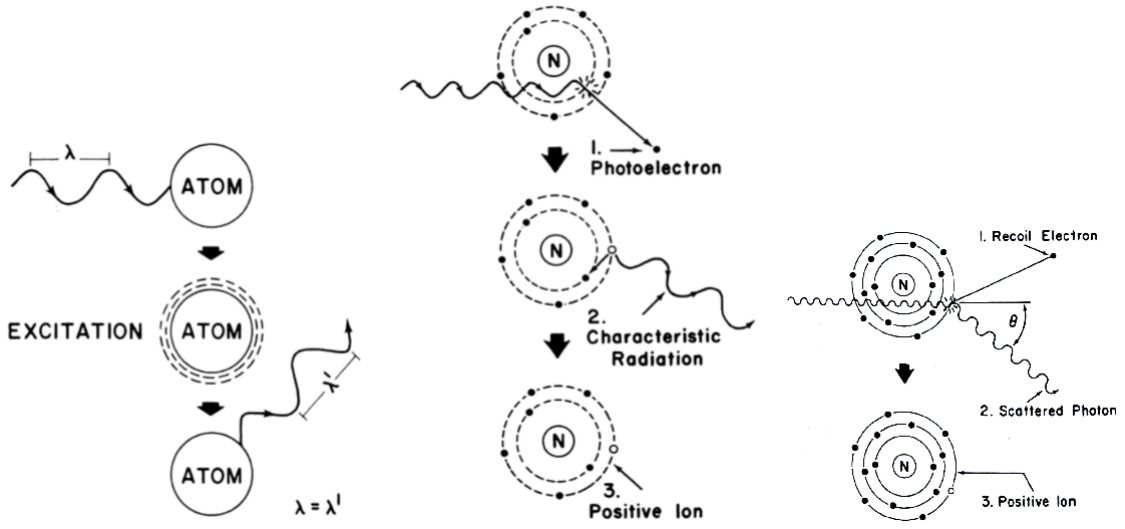
Except mammography, where soft tissue contrast is created using discrete spectra, diagnostic imaging of other tissues are based on continuous spectra [16].

2.3.2. Interaction Mechanisms of X-ray Photons in Matter

In medical applications such as radiography, mammography and CT, contrast in images is formed by the attenuation (gradual loss of intensity) of the incoming X-ray beam by an object with certain attenuating properties. For example, bones and metals result in a higher attenuation, whereas, soft tissue results in lesser attenuation. In the following paragraphs, the principles behind attenuation of X-rays are presented briefly. In addition, the characteristics of an X-ray beam used in conventional medical imaging are presented.

Attenuation can be caused by absorption, coherent and incoherent scattering, reflection and refraction of X-ray photons. However, for medical X-ray imaging, attenuation is caused mainly by absorption and scattering when X-ray photons interact with orbital electrons of atoms of the imaged object. In particular, the interaction happening in the range of X-ray energy used in diagnostics is through coherent scattering, the photoelectric effect and Compton scattering [16], where the effects depend on the density and the composition of the material as well as the incoming energy of the X-rays.

In coherent scattering, also referred to as classical scattering, radiation encounters the electrons of an atom and sets them into vibration at the frequency of the radiation. As shown in Fig. 2.4(a), these vibrating electrons emit radiation, while returning to their original state. This is the only type of interaction, where ionization does not occur. The percentage of radiation that undergo coherent scattering is small compared to the other types of interactions, typically less than 5%.



(a) (b) (c)

Figure 2.4.: Illustration of the mechanisms (a) coherent scattering, (b) photoelectric effect and (c) Compton scattering by which X-ray photons interact with matter. Most of the interactions are due to Compton scattering and the photoelectric effect. Only about 5% of the radiation undergo coherent scattering [16].

The photoelectric effect occurs when photons having energy more than the binding energy of an electron, kicks out such an electron from its orbit. The incident photon is completely absorbed in the process. The probability of this effect is the highest when the electron is tightly bound (as in the K-shell). The electron void on the K-shell of the atom is instantly filled, usually by an electron from the outer adjacent L-shell, occasionally from the M-shell and very rarely by other free electrons from the same or a different atom. As an electron drops into the K-shell, it emits the energy in the form of an X-ray photon. The amount of energy emitted depends on the element and the characteristic radiation is produced by the transition of electrons within the atom. Then an electron from the M-shell drops down to fill the void in L-shell, which also results in the emission of radiation [17].

The photoelectric effect reduces as the photon energy (E) increases. That is, to break the binding energy of 33.2 keV, a photon with 34 keV would be more successful compared to a photon with 100 keV. The photoelectric effect is also more likely to occur for atoms in which the electrons are tightly bound in the orbit. This characteristic is possessed by elements with higher atomic number (Z). For elements with low atomic numbers, most interactions occur at the K shell, whereas, for elements with high atomic numbers, interactions can take place even at the L and M shells, if the energy of the incident photon is insufficient for the K shell [16]. This process is shown in Fig. 2.4(b) and the occurrence of the photoelectric effect can be represented as follows:

$$\text{photoelectric effect} \propto \frac{Z^3}{E^3}. \tag{2.8}$$

For diagnostic imaging, the photoelectric effect is responsible for the tissue contrast in

radiographic images, as this effect or absorption differs for different elements. As photoelectric reactions also result in radiation exposure for patients, it is desirable to use high photon energies to reduce the absorbed dose, but it will yield X-ray images with lower contrast [37].

Almost all scatter radiation encountered in diagnostic radiology is from Compton scattering [16]. In this process, represented by Fig. 2.4(c), an incident photon with relatively high energy ejects an electron from the outer shell of an atom, but, during this process the photon also gets deflected. The reaction produces an ion pair, a positive atom and a negative electron, which is also referred to as a recoil electron. A part of the energy of the incident photon is transferred to the recoil electron as kinetic energy and the rest is retained by the deflected photon. When compared to the photoelectric effect, in Compton scattering, the amount of energy spent on freeing the photoelectron is significantly less. The energy retained by the deflected photon can be calculated using the initial energy of the incident photon and the angle of deflection of the recoil electron [16]. Compton scattering impairs contrast between soft tissue types and vessels. To overcome this problem, anti-scatter grids are mounted in front of X-ray detectors. Another way to increase contrast in specific structures, for example, blood vessels, is to use contrast agents like iodine and barium [38, 39].

The intensity of an X-ray beam is given by the product of the number of photons in the beam and the energy of the photons. For monochromatic radiation, containing a single energy level, attenuation depends on the material through which the beam passes and it can be described using the Beer-Lambert law

$$I = I_0 exp\left(-\int_{x0}^{x1} \mu(x)dx\right), \tag{2.9}$$

where I_0 is the initial intensity of the beam, I is the reduced intensity after attenuation, $x \in [x_0, x_1]$ denotes the path of the X-ray photons and $\mu(x)$ denotes the linear attenuation coefficient of the material. For objects with constant attenuation coefficient μ and thickness d, attenuation according to Beer-Lambert can be simplified to:

$$I = I_0 e^{-\mu d}. \tag{2.10}$$

For conventional medical X-ray imaging, only polychromatic beams are available. Unlike monochromatic beams, polychromatic beams contain a whole spectrum of various energies. Their mean energy is from one third to one half of the peak energy. As polychromatic radiation passes through a material, the transmitted beams undergo change in the number of photons, and the energy spectrum of the beam changes too. A large percentage of low energy photons are attenuated when the beam enters the material and this results in the increase in the mean energy of the remaining photons that continue to pass through the material [16]. This effect is called beam hardening. In order to prevent the lower energy photons from being rapidly absorbed in the tissue, they can be filtered out, using metal filters placed after the X-ray tube and before the patient. By this process, only those photons that are useful for imaging are retained.

2.3.3. X-ray Detection

Detectors serve as a means for measuring the intensity of an attenuated X-ray beam. Over the years, the fundamental design of the X-ray tube has basically remained the same, whereas the detectors have constantly undergone changes. Detectors can be broadly classified into analog and digital detectors. In the case of analog detectors, the absorbed energy can be used to directly form an image, for example, on a photographic plate. Whereas, in digital detectors, a charge generated based on the absorbed energy is first converted into a digital signal and then stored as a digital image. Analog film detectors are still used for static X-ray images while digital detectors are necessary to record dynamic X-ray image sequences [11].

Requirements of X-ray detection

In clinical X-ray imaging, X-ray energy, X-ray flux and spatial resolution play an important role and are described in the following paragraphs.

The X-ray energy used for clinical applications is in the range $30-130$ keV. The detection systems have to be most efficient in this range. The choice of the energy depends on the anatomy under observation and the procedure at hand. The lower end is usually used for mammography and the upper for CT. The intensity or the flux of X-ray photons (photons per second per surface area) is in the range $10^0 - 10^9$ ph/sec/mm^2 for clinical applications. For interventional as well as non-interventional radiology, the flux of the X-ray photons is reduced significantly due to attenuation by the patient. In other words, when a low X-ray dose level is used, the number of photons emitted by the X-ray tube is also low. As a consequence, only a few X-ray photons are available for image formation after the attenuation by the patient. The effect of attenuation of the incident photons by a patient with an average thickness of 20 cm has been evaluated in this thesis and is shown in Table 2.1. Here, the patient entrance as well as detector entrance dose rates associated with different X-ray energies and intensities used in interventional radiology are presented. The patient has been emulated with plexiglass of thickness 20 cm. The dose rates have been measured using RaySafe X2 X-ray Measurement System, Fluke Biomedical, Billdal, Sweden. In order to calculate the number of photons that contributed to image formation, first the mean energy associated with a spectrum has been computed using Spektr 3.0 [40]. Then, the relationship between mean X-ray energy and dose per mm^2 given in [17] has been used for estimating the number of X-ray photons present per mm^2 in the attenuated beam. Finally, by assuming a dimension of a sensing element on the detector of 0.15 mm $\times$ 0.15 mm, the number of associated photons was computed. It has been found that at very low dose levels, only a few hundred photons per second are available for image formation. The value will significantly reduce at higher frame rates.

The flux of the X-ray photons controls the dynamic range of the images. The dynamic range is defined as ratio of the maximum to minimum value detected within the linear range of the detector [13]. This is because there are limitations in the lower and upper end (minimum and maximum values). If the input signals are outside the detector's dynamic range, these values are clipped and the detector behaves non-linearly. Small values at the lower end are corrupted by electronic read-out noise, whereas, at the upper end, the values might get saturated, for example, due to limitations in the capacitors of the photo diodes

Table 2.1.: The patient entrance and detector entrance dose rates (μG/s) associated with different X-ray energies (kVp) and intensities, determined by tube current (mA), pulse width (ms) and Copper filtration (mm), as well as the number of X-ray photons available for image formation. The patient entrance dose has been measured by placing plexiglass (polymethyl methacrylate, PMMA) of thickness 20 cm to substitute the patient. The number of X-ray photons per detector element (0.15 mm $\times$ 0.15 mm) has been computed from the mean energy associated with the chosen spectrum based on the relationship between X-ray dose and mean X-ray energy presented in [17]

X-ray Energy	Tube Current	Pulse Width	Copper Filter	Dose Rate (μG/s)		No. Photons/s Per Detector Element
				Patient Entrance	Detector Entrance	
68	35	3.5	0.2	66.66	0.16	98
68	70	3.5	0.2	133.33	0.32	197
68	138	3.5	0.2	255.00	0.62	381
68	150	5.6	0.2	441.66	1.04	639
68	153	11.1	0.2	866.66	2.08	1278
65	395	12	0.2	2083.33	4.58	2721
65	399	22	0.2	4383.33	8.43	5007
70	425	23.3	0.0	16000.00	22.50	10581
75	400	24.8	0.0	18333.33	28.27	14184
83	360	25	0.0	21000.00	39.60	22453

for indirect conversion detectors. The analog to digital conversion contributes some more noise.

Energy resolution or the ability to discriminate between different energies is a good characteristic to have. This is because the interaction of the X-ray beam with the object to be imaged depends strongly on the X-ray energy [41]. Therefore, the energy of the attenuated beam can be used for resolving the different material of a heterogeneous object. This information cannot be exploited if it is lost in the detection. For example, one photon at 60 keV could produce, in principle, the same signal as two photons of 30 keV or three photons of 20 keV, while the transmission through the target of a photon of 60, 30, or 20 keV is very different [41]. Although energy resolution, ability to discriminate between different energies, is a good characteristic to have, it is not yet clinically available, at least not for typical X-ray angiography devices, and is still under research.

Spatial resolution is another important property in X-ray imaging as it is related to detecting details and hence to the quality of the acquired data [42]. In clinical applications, the required spatial resolution is in the range 50 μm for mammography to 0.5 mm for dental imaging and most applications in CT. More details on spatial resolution are given in Sec. 2.3.5.

Analog Image Detectors

Conventional X-ray imaging is based on analog detectors. They make use of photographic films – coated with silver bromide emulsion – which are exposed to X-rays and then developed. X-ray absorption leads to the splitting of silver bromide into silver and bromide ions. The development process forms oxidized silver that appears black on the photographic film. Though the film has an excellent spatial resolution of about 0.025 mm, it is not very sensitive to radiation as only 1% of the X-ray quanta contribute to the image [16]. By placing the film between two intensifier screens in film cassettes, the quantum efficiency can be increased and this allows a reduction of X-ray dose. The screens emit visible light by luminescence when exposed to radiation. This luminescence, in turn, exposes the photographic emulsion, which results in the generation of a static image. This type of imaging is still widely used, for example, for dental imaging because it is very compact compared to any digital detector.

Static Digital Imaging Detectors

In the beginning of the 1980s, computed radiography (CR), a digital technique, succeeded the above discussed analog technique. CR uses photostimulable phosphor as image receptor. The phosphor is composed of europium activated Barium fluorohalide, which is coated on an imaging plate. The phosphor temporarily stores a latent image on the imaging plate and this latent image is converted to a light image, using laser-stimulated luminescence. A photo-multiplier tube converts the light image into an analog electrical signal, which is amplified, converted to a digital signal and stored in the computer. Though the spatial resolution is only 0.1 mm, the resulting clinical images are often as useful as film systems. These CR systems provide the advantage of reusable CR cassettes, which contributes to the reduction of X-ray image acquisition cost.

The technique CR has been succeeded by digital radiography (DR) (using flat panel detectors) in the late 1990s. This technique makes use of the semiconductor flat panel detectors. Amorphous silicon has been a suitable choice of material for making photodiodes due to its superior real-time imaging capability [43]. As silicon is not sufficiently sensitive for detecting X-ray energies, it is coated with a layer of fluorescent caesium iodide (CsI) or gadolinium oxiSulfate (GOS), which act as an image conversion layer by absorbing X-ray photons and emitting visible light. GOS is less expensive than CsI and provides good light output but the emitted light will spread out resulting in the blurring of the emitted signal. On the other hand, CsI has less light spread since the crystal has a columnar structure which makes the light follow the columns of the CsI as it passes through the crystal. This reduced blurring leads to higher spatial resolution for CsI based systems compared with GOS systems. Due to the high quantum detection efficiency (60%), the resulting image is of superior quality compared to CR systems. Flat panel detectors yield an improved workflow at reduced operating costs. The main advantage of the DR systems is the instant availability of the image. There is no need for film development or storage cassette readers.

Dynamic Digital Imaging Detectors

In the early days of X-ray imaging, dynamic processes were visualized on fluorescent screens. Those were succeeded later on by image intensifiers. Nowadays flat-panel digital detectors have been gradually replacing the image intensifier systems [11]. The detected quantum efficiency of flat panel detectors is about 20% higher than that of conventional intensifiers and this leads to better image quality and dose reduction. In particular, the flat panel detectors offer distortion-free images, improved contrast, large dynamic range and high X-ray sensitivity [12]. Moreover, as these digital detectors weigh less when compared to conventional intensifiers, the X-ray imaging systems can be made more manoeuvrable.

There are two types of flat detector technologies. The first one makes use of indirect X-ray conversion processes, while the second one converts X-ray photons directly into an image. Both these technologies can be coupled with active pixel readout matrices made of hydrogenated amorphous silicon (a-Si:H) as they feature several favourable characteristics. First, they exhibit semiconductor properties such as doping, photo conductivity or junction formation, which makes it suitable for the manufacture of electrical components such as thin film transistors (TFTs) and photodiodes. Second, the plasma decomposition process permits large-area decomposition and this allows for the production of active matrices of several million pixels with sizes exceeding 40x40 cm^2. Third, as amorphous silicon has proven to be highly robust to radiation, it is suitable for applications in medical X-ray imaging [12].

The flat detector technology with the two-level indirect X-ray conversion process using cesium iodide (CsI) scintillators is the dominating method as it offers considerable advantages in radiography, angiography and fluoroscopy [44, 45, 46, 47, 13]. In this detection process, an X-ray photon is absorbed in the scintillator mainly by the photoelectric effect. This creates a large number of electron-hole pairs, which are trapped in luminescence centers. These pairs then recombine to produce scintillation photons in the visible range [46]. The light is converted into electrical charge in a pixelized matrix of photodiodes [13]. For flat detectors, the scintillator of choice is thallium-doped CsI (CsI:TI) as it provides several favourable properties for medical imaging. The high atomic numbers of Cs (55) and I (53), and the effective density of CsI (about 3.6 g/cm^3) results in very good X-ray absorption properties. In particular, it is well suited to enable high detective quantum efficiency (DQE) for different clinically relevant frequencies and energy spectra required for a wide range of imaging tasks [13].

A schematic representation of this detection process is presented in Fig. 2.5. The scintillator comprises CsI:TI shaped in needle-like structure with individual crystals measuring 5-10 μm in diameter and up to 600 μm or more in length. This structure facilitates the collection of the emitted light into the photo-diode below the scintillator [13]. In the acquisition and read out process, the inversely biased photodiodes are first charged to a given voltage. Then, during X-ray acquisition, the impinging optical photons, created during the photoabsorption and scintillation process, generate electron-hole pairs which gradually discharge the photodiodes [13]. The readout process is carried out by addressing all the TFTs line by line. Specifically, in this process, the charge required to recharge the respective photodiodes to the predefined bias voltage is computed. Low-noise application-specific integrated circuits are used to carry out charge-sensitive amplification. Multiplexing is performed to feed analog-to-digital converters with bit depths of 14 bits or 16 bits.

After this, the image data is available for preprocessing (e.g., flat-field correction, defective pixel correction), and clinical image post-processing (e.g., denoising or contrast enhancement) [13]. Recently, a new readout technology has been introduced by replacing the a-Si photodiode and the readout structure with a complementary metal oxide semiconductor (CMOS) sensor containing crystalline Si. The electrical properties of crystalline Si enable high performance analog or digital circuitry. This reduces readout electronic noise and makes on-pixel amplification possible (active pixel sensors). This, in turn, helps in moving towards the goal of quantum-noise-limited X-ray detectors for low-dose levels.

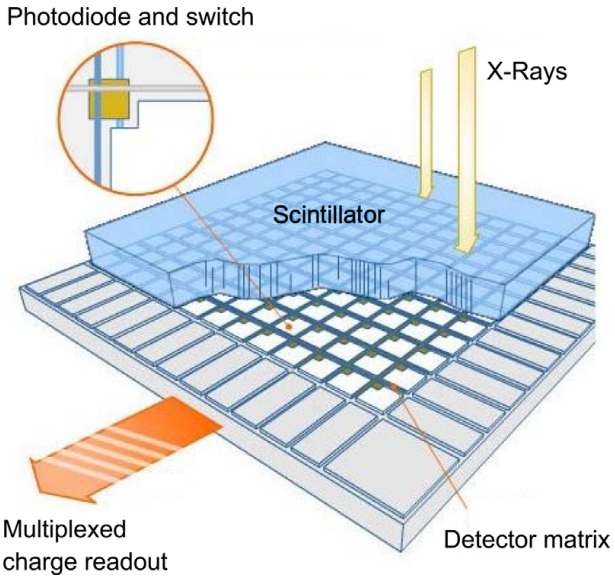


Figure 2.5.: Schematic view of an indirect conversion flat panel detector based on columnar cesium iodide and an active matrix of amorphous silicon photodiodes [13].

On the other hand, in the direct conversion flat detector technology, photoelectrons generated during the X-ray absorption process directly create electron-hole pairs which are collected by applying an electric field across the detector material [13]. This method makes use of amorphous selenium (a-Se) as the X-ray detection material. A-Se can be deposited directly onto an active matrix of a-Si, where each pixel consists of a charge collecting electrode and readout TFT. The low K-edge (12.6 keV) of a-Se makes it suitable for the relatively soft X-ray spectra used in mammography [13].

Apart from cost reduction, digital imaging provides various advantages such as the opportunity to process the image before and after visualization as well as data archival. By post-processing, the image quality can be improved. This, in turn, improves the diagnostic accuracy as well as the outcome of interventional procedures.

2.3.4. X-ray Image Formation

By modeling the behavior of an indirect-detection, flat-panel X-ray detector using a cascaded linear systems approach, the formation of fluoroscopic X-ray images from X-ray photons can be described as a succession of stages. Each stage represents a physical process having intrinsic gain or spreading properties [44]. Assuming that the flat-panel detec-

tor has a linear detector response, the average output signal $\bar{z}$ can be written as

$$\bar{z} = a_{pd}^2 \cdot \bar{x} \cdot \bar{g_1} \cdot \bar{g_2} \cdot \bar{g_3} + g_o, \tag{2.11}$$

where a_{pd}^2 represents the light sensitive area of the sampling aperture (photo-diode), $\bar{x}$ is the mean fluence of Poisson-distributed incident X-ray quanta, $\bar{g_1}$ is the mean fraction of X-rays that interact with the converting medium (scintillator) to produce light, $\bar{g_2}$ represents the mean quantum gain associated with the generation and emission of optical quanta in the X-ray converter, $\bar{g_3}$ denotes the mean gain associated with the coupling of optical quanta to the detector elements and g_o is the overall offset. As described in [44], when the detector is used for acquiring image sequences, for example, in fluoroscopic mode, the variance of noise associated with the exposure can be expressed as

$$\sigma_{\bar{z}}^2 = \left(\frac{1-t}{1+t}\right) a_{pd}^2 \cdot \bar{x} \cdot \bar{g_1} \cdot \bar{g_2} \cdot \bar{g_3} \left(1 + \bar{g_3}\left(\bar{g_2} + \epsilon_{g_2}\right) s\right) + \sigma_n^2. \tag{2.12}$$

In Eq. 2.12, $t \in [0, 1]$ denotes the fraction of the number of electrons trapped in the photo-diode after readout and the parameter $\epsilon_{g_2} = \frac{\sigma_{g_2}^2}{\bar{g_2}} - 1$ is the Poisson excess (where $\bar{g_2}$ is the mean gain, and $\sigma_{g_2}^2$ is its variance) [44]. Poisson excess noise is a result of variations in the number of secondary quanta detected per primary quanta absorption [48]. The parameter s in Eq. 2.12 refers to the sharpness factor of the system and can, for example, be calculated in the frequency domain as follows:

$$s = a_{pd}^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} T_1^2(u, v) \cdot T_2^2(u, v) \, \mathrm{d}u\mathrm{d}v. \tag{2.13}$$

Here, $T_1(u, v)$ and $T_2(u, v)$ represent the system modulation transfer functions (MTFs) describing the spreading of optical photons in the converting medium and the integration of quanta by the photo-diode, respectively. The variables u, v are the spatial-frequency coordinates. The sharpness factor $s \in [0, 1]$ accounts for the effects of image blur on the noise and it is one for an ideal converter (no spreading) [44]. Finally, the parameter σ_n^2 in Eq. 2.12 refers to the variance of the total additive noise due to different sources, such as intrinsic noise from the photo-diode pixel, voltage fluctuations on gate and bias lines, noise from the amplifier and digitization noise [44]. Equation 2.12 shows that the variance $\sigma_{\bar{z}}^2$ is proportional to the mean signal $\bar{z}$ with an offset due to additive noise σ_n^2.

By rewriting Eq. 2.12 with respect to $\bar{z}$ (given by Eq. 2.11) and taking samples at detector pixel positions, the following is obtained:

$$\sigma_{\bar{z}}^2[r, c] = \alpha[r, c]\left(\bar{z}[r, c] - g_o\right) + \sigma_n^2 \;\text{ with }\; \alpha[r, c] = \left(\frac{1-t}{1+t}\right)\left(1 + \bar{g_3}\left(\bar{g_2} + \epsilon_{g_2}\right) s\right), \tag{2.14}$$

where $[r, c]$ are spatial coordinates and $\alpha[r, c]$ is the pixel-dependent gain associated with the pixel at $[r, c]$. In general, the gain is spatially varying because it depends on the energy of the X-ray beam reaching the associated detector elements [49, 50]. Equation 2.14 represents the noise level function (NLF), relationship between noise variance and gray values, that can also be written as the variance of a Poisson variable with mean $\bar{z}[r, c]$ scaled by $\alpha[r, c] > 0$ and corrupted by additive Gaussian noise $\rho[r, c]$ as follows:

$$\sigma_z^2[r, c] = \alpha[r, c] \cdot \bar{z}[r, c] + \rho[r, c] \quad \text{with} \quad \rho[r, c] = \sigma_n^2 - \alpha[r, c] \cdot g_o. \tag{2.15}$$

From Eq. 2.15, it can be observed that $\alpha[r, c]$ and $\rho[r, c]$ are the slope and the y-intercept of the NLF, respectively. A different aspect of noise in digital X-ray detectors is given in the following subsection.

2.3.5. Characteristics of an X-ray Imaging System

The characteristics of an X-ray imaging system is determined mainly by the spatial resolution, effectiveness of the system in converting incident X-ray photons into signal, and the characteristics of noise associated with the acquired images. In the following paragraphs, these aspects are explained briefly.

Spatial resolution does not just depend on the nominal pixel size given by the manufacturer but on the overall image acquisition process. If the detector is modeled as a linear system, a point spread function (PSF) can be used for characterizing its properties, in particular, the spatial resolution. A PSF shows the response of the detector to a point-like X-ray beam [51]. The MTF of the detector is the Fourier transform of the PSF. It is a contrast function that defines how different spatial frequencies are actually transferred through the detector. It specifies the contrast in the output for different spatial frequencies at the input. In practice, the MTF is typically measured using line pairs with varying spatial frequencies. The spatial resolution limit of an imaging system is typically defined as 10% MTF, i.e., the point where the contrast has dropped to 10% of its original value.

Quantum efficiency (QE) measures how many of the incoming photons contribute to the actual image formation. In other words, it is the measure of the effectiveness of an imaging device to convert incident X-ray photons into the signal responsible for image formation. On the other hand, the DQE is the ratio of the detected signal-to-noise ratio (SNR) to the incoming SNR [17]:

$$\text{DQE} = \frac{\text{SNR}_{det}^2}{\text{SNR}_{in}^2}. \tag{2.16}$$

Both the signals are typically Poisson-distributed. The DQE gives the complete performance of the detector and also includes the influence of noise. As described in [52], the DQE associated with a spatial frequency f can also be expressed in terms of measurable quantities such as MTF, noise power spectrum (NPS), system gain α and mean incident fluence $\bar{x}$:

$$\text{DQE(f)} = \frac{\bar{x}\alpha^2\text{MTF}^2(f)}{NPS(f)}. \tag{2.17}$$

The NPS describes the spectral composition of the noise variance in an image as a function of spatial frequency.

As mentioned in Sec. 2.3.4, noise in digital X-ray detectors consists of different components such as quantum noise (primary and secondary quanta), excess Poisson noise, aliasing noise and additive electronic noise encountered during read-out. In addition, structure or fixed pattern noise occurs due to the spatially fixed variations in the gain across the detector. The effect of this noise is removed as a part of clinical imaging routine through flat-field image correction. According to the sampling theorem, aliasing occurs when an image (acquired at the detector entrance) contains frequencies higher than the Nyquist frequency

– which is defined as $f_{\text{NYQ}} = 1/(2\Delta y)$, where Δy is the pixel pitch in mm. Aliasing noise has a strong effect on direct conversion flat panel detectors. However, for the scintillator-based indirect conversion detectors its effect is almost negligible. This is because the blur of the scintillator acts as a low-pass filter which reduces the high frequencies above f_{NYQ}.

In practice it is non-trivial to measure the pre-sampling NPS [48], whereas, it is possible to compute the NPS associated with the acquired digital images from flat-field X-ray images. After correcting the fixed pattern noise in X-ray images, the NPS can be calculated from flat-field images as follows. First, the global mean associated with the flat-field images is subtracted. Then, the 2D-NPS is calculated from P non-overlapping regions of interest $y_p \in \mathbb{R}^{M' \times N'}$ extracted from the field images as shown below:

$$\text{NPS}[u, v] = \frac{\Delta r \cdot \Delta c}{P \cdot M' \cdot N'} \sum_{p=1}^{P} |\mathcal{F}\{y_p[r, c] - \mu_{y_p}\}|^2, \tag{2.18}$$

where u and v are spatial frequencies corresponding to r and c, $\mathcal{F}$ represents the Fourier transform, Δr and Δc are the vertical and horizontal pixel pitches, and μ_{y_p} is the mean value of the region y_p.

2.4. Clinical Applications and Challenges

X-ray imaging can be used to reveal internal structures of the body that is made up of various parts with differing densities, such as bones and soft tissue. This is the reason behind the wide use of X-ray imaging in diagnostics, therapy planning, real-time navigation guidance and procedure monitoring. Both static and dynamic X-ray imaging techniques have their own set of applications.

Static images are mainly used for the examination of the skeleton and for detecting disease processes in soft tissue. For example, X-rays are suitable for visualizing bony fractures and ingested foreign objects with high density. Not only diseases concerning bone pathology such as osteoarthritis [53], osteomyelitis [54] and osteosarcoma [55], can be diagnosed using radiographs, but also growth studies such as achondroplasia and scoliosis can be conducted [56]. X-rays of the chest are useful for identifying lung diseases such as pneumonia, lung cancer and pulmonary oedema [57]. Abdominal X-rays help to detect bowel obstruction, gallstones and kidney stones. Dental radiography allows the diagnosis of cavities in the teeth. X-ray examination of the breasts (mammography) is used to detect breast cancer. With the exception of mammography, X-ray dose minimization is often not of the highest priority in these contexts, as only a limited number of radiographs are taken per patient. Therefore, X-ray dose is often adjusted to yield high contrast to aid diagnosis. Another exception to this are paediatric examinations, as growing tissues are very sensitive to ionizing radiation.

Fluoroscopy is the acquisition of a sequence of X-ray images in real-time. This technique is mainly performed to view the movement of contrast agents (angiography) and also to guide medical instruments in the interior of the human body. A wide variety of applications such as abdominal aortic aneurysm repair, percutaneous coronary intervention [58], digital subtraction angiography (DSA), ablation, stenting, endovascular coiling for brain aneurysms, endovascular embolization for treating arteriovenous malformation in the brain rely on this technique [59].

Ideally the choice of the parameters used for imaging should be organ-specific. For medical diagnostic and interventional imaging, typical values for the X-ray energy E range from 25 to 150 KeV [11]. Even with a specific choice of parameters, it is difficult to discriminate different tissue types from blood vessels. This is solved by the injection of contrast agents, typically iodine [60]. However, iodinated contrast agent puts a burden on the patients' kidneys.

Although the per-image dose used in fluoroscopy is lower compared to radiographs, dose minimization is still important. This is due to the fact that a large number of images are acquired during interventional procedures. For example, complex cases such as percutaneous coronary interventions [58] may expose the patients and the staff, over a period of time, to high doses of radiation. Traditional methods for dose management include beam pre-filtration and pulsed fluoroscopy at 7.5 to 30 fps [11]. In order to achieve further radiation reduction, the acquisition dose can be reduced, but this degrades the image quality. In addition, organ motion increases the complexity of fluoroscopic X-ray imaging as this can result in significant blurring. Although the amount of blurring caused by motion can be reduced significantly using pulsed fluoroscopy, the effect cannot be completely removed. For example, to produce an X-ray image at the lowest possible dose and with an acceptable image quality, X-rays at a certain energy and a pulse duration need to be used. A reduction in the pulse duration would also reduce the image quality if it is not possible to increase the tube current. Therefore, it is necessary to restore the image quality associated with such acquisitions.

Sophisticated image processing techniques, in particular, image denoising, have been shown to be successful in lowering the X-ray dose level and at the same time produce images with the required quality [61]. However, utmost care has to be taken while processing the images, as artifacts can lead to wrong clinical decisions. For instance, artifacts in mammograms might be mistaken for cancer and will result in an unnecessary biopsy of a healthy woman. In Chapters 5 and 6, advanced X-ray image denoising techniques are presented along with a thorough evaluation with respect to denoising performance and freedom from artifacts.

2.5. Summary

X-rays are ionizing radiation that can penetrate objects and, hence, provide information about tissues and structures inside the body based on their X-ray absorption properties. They are generated in an evacuated X-ray tube enclosing two electrodes known as the cathode and the anode, respectively. For angiography, tungsten is usually used for making the cathode and the anode as it has a high melting point and can dissipate heat rapidly. There are three steps in the production of X-rays: (i) thermal emission of electrons in the cathode, (ii) electron acceleration in the vacuum and (iii) physical processes that generate X-rays in the target, namely Bremsstrahlung and characteristic radiation. In the case of Bremsstrahlung, the spectrum of the generated X-rays is continuous, whereas, for characteristic radiation, the spectrum is discrete. For diagnostic imaging (except mammography) and interventional imaging, Bremsstrahlung is used for generating X-rays. Flat detectors that utilize indirect X-ray conversion process are typically used for detecting and converting the transmitted X-rays into images. During the process of X-ray image formation, there

are two main sources of noise, namely quantum noise and electronic noise. The image formation process can be modeled using a succession of stages where each stage represents a physical process having intrinsic gain or spatial spreading. This facilitates the derivation of the relationship between signal and noise. In order to keep the exposure of the ionizing X-ray radiation to a minimum, the X-ray dose needs to be kept as low as reasonably achievable. However, this reduces the SNR. Hence, it is common practice to use denoising algorithms to improve the image quality. In Chapter 3, an overview of well established denoising approaches is presented. In order to achieve superior denoising performance, the characteristics of noise has to be taken into account. In Chapter 4, a data-driven method for extracting the noise characteristics associated with low-dose X-ray images is presented. Finally, the techniques developed during this research are presented in Chapters 5 and 6 .

3. Image Denoising

3.1. Introduction

As mentioned in Chapter 2, low-dose X-ray images are severely corrupted by noise. As a consequence, they have a low signal-to-noise ratio (SNR). In order to improve the image quality of such noise-corrupted low-dose X-ray images, especially the visibility of clinically relevant information, denoising techniques can be designed and applied by carefully taking into account the associated noise characteristics (mentioned in Sec. 2.3.4 and Sec. 2.3.5). Noise reduction has been an important requisite not only for medical images but also for optical images, e.g., digital photographs. Hence, this problem has been studied for a long time. Mathematically, the problem can be represented as

$$y = x + n, \tag{3.1}$$

where $y \in \mathbb{R}^{M \times N}$ is the observed noisy image, $x \in \mathbb{R}^{M \times N}$ is the unknown noise-free image and $n \in \mathbb{R}^{M \times N}$ represents the noise. The parameter n is modeled based on the characteristics of noise, e.g., a Gaussian distribution, a Poisson distribution or a mixture (as mentioned in Sec. 2.3.4). The purpose of denoising is to get a good estimate $\hat{x}$ of the noise-free image by reducing the noise n in the image y and hence enhancing the visibility of information present in the image, i.e., improving the SNR. Therefore, it is important to minimize loss of structures (in particular, edges and texture) during this process. Since denoising is an ill-posed inverse problem with a solution that is not unique, it is a challenging task to obtain a good noise-free estimate. In addition to preserving the essential information in the images, denoising should not introduce any artifacts. This is particularly important in the case of medical imaging as this could lead to a wrong medical decision.

In this chapter, a brief overview of the achievements including some current state-of-the-art approaches made in the area of denoising that are relevant to this thesis is presented. To this end, the denoising methods are broadly classified into analytical and learning-based approaches. A comprehensive overview of the methods can be found in [62, 63, 64]. The overview is followed by commonly used image quality metrics for evaluating the results of denoising.

3.2. Analytical Approaches

Prior to the advent of learning-based methods (introduced in Sec. 3.3), denoising approaches were designed by explicitly taking into account certain characteristics present in the images. In this thesis, the term analytical approaches has been chosen to refer to these methods. Well established denoising techniques include spatial domain filtering, variational approaches, patch-based and transform-based methods, and approaches that

involve sparse and low-rank modeling. In this section, these approaches are introduced and described in more detail.

3.2.1. Spatial Domain Filtering

Approaches that involve filtering in the spatial domain are widely used as they are simple to implement and also provide a good trade-off between performance and speed. The approaches can be classified into linear and non-linear filtering. The traditional filtering approaches used are linear and separable, i.e., a 2D filter can be replaced by two 1D filters. In these techniques, a noisy image y is convolved with a filter kernel g to obtain the filtered image $\hat{x}$ as per

$$\hat{x} = y * g. \tag{3.2}$$

Mean and Gaussian filters are commonly used examples of the filter kernel [65, 66].

Although linear filters were designed to remove noise and can achieve this quite efficiently, they often fail to preserve image textures [64]. For example, mean filtering, which has been adopted for Gaussian noise reduction, can over-smooth images that are corrupted by a high amount of noise due to the rather large support needed in this case. On the other hand, non-linear noise-reduction filters such as the median filter and weighted median filtering were introduced to deal with impulse noise [67, 68]. In median filtering, each pixel is replaced by the median value of its neighboring pixels. Bilateral filtering is a widely used non-linear and edge-preserving noise-reduction approach used for denoising images corrupted by Gaussian noise [69]. The intensity values of the pixels in an image are replaced by a weighted average of their spatial neighbors. The weights are assigned to the spatial neighbors based on their distance to the current pixel as well as the similarity in the intensity values of the corresponding pixels.

3.2.2. Variational Denoising Approaches

Variational denoising approaches are based on utilizing image priors (assumptions on the images) and minimizing an energy function $E(.)$ to estimate $\hat{x}$:

$$\hat{x} = \arg\min_x E(x). \tag{3.3}$$

The maximum a posteriori (MAP) probability estimation is the backbone of variational denoising methods. From a Bayesian perspective, the MAP probability estimate of x based on the observations y is given by

$$\begin{aligned} \hat{x} &= \arg\max_x p(x|y) \\ &= \arg\max_x \frac{p(y|x)p(x)}{p(y)}, \end{aligned} \tag{3.4}$$

where $p(y|x)$ is a likelihood function of x and $p(x)$ represents the image prior [64]. In the case of additive white Gaussian noise (AWGN), i.e., additive Gaussian noise with a flat noise power spectrum (NPS), Eq. 3.4 can be rewritten as

$$\hat{x} = \arg\min_x \left(\frac{1}{2}||y - x||_2^2 + \lambda R(x) \right), \tag{3.5}$$

where $||y - x||_2^2$ is the data term, $R(x)$ is the regularization term and λ is the regularization parameter [64]. In the case of variational methods, the regularization term $R(x)$ and the regularization parameter λ play an important role. In Tikhnonov regularization [70, 71], $R(x)$ is minimized with respect to the L2-norm. Unfortunately, this may result in the over-smoothing of image details such as edges and texture.

Anisotropic diffusion-based methods have been designed to preserve image details [72, 73]. In the case of anisotropic diffusion, smoothing is performed iteratively only within regions defined by edges. In other words, they apply an inhomogenous process that reduces the diffusivity at those locations which have a larger likelihood to be edges [73]. The Perona-Malik filter [72] is based on the equation

$$\partial_t y = \text{div}(g(|\nabla y|^2)\nabla y), \tag{3.6}$$

with ∇ being the gradient operator and $\text{div}(.)$ is the function that calculates the divergence. The Perona-Malik filter [72] uses diffusivities $g(.)$ such as

$$g(s^2) = \frac{1}{1 + s^2/\kappa^2} \qquad \text{with } (\lambda > 0), \tag{3.7}$$

where s is the function variable and κ is the gradient magnitude threshold parameter that controls the rate of the diffusion. That is, it serves as a soft threshold between the image gradients that are attributed to noise and those attributed to edges [74]. Even after carefully choosing diffusitives, anisotropic diffusion may still result in blurred edges [75, 76]. To address the issue of over-smoothing, total variation (TV)-based regularization has been proposed [64, 77].

The assumption made here is that noisy images are grainy throughout and that noise-free images have a set of connected objects, with smooth contours. The images are also assumed to be smooth inside the connected objects with variations only across the boundaries. Therefore, denoising is performed by minimizing discrete gradients - that correspond to noise - while making sure that the denoised image closely resembles that of the noisy image [78]. TV regularization is defined as follows:

$$R_{\text{TV}}(x) = ||\nabla x||_1. \tag{3.8}$$

Even though this approach can yield denoised images with sharp edges, it has three major drawbacks. First, textures tend to be over-smoothed. Second, flat areas are approximated by a piece-wise constant surface resulting in stair-casing effect. Finally, the results suffer from a loss of contrast [79, 80, 81]. In order to improve the performance of this regularization model, partial differential equations have been extensively studied [82, 83, 84].

3.2.3. Transform-based Approaches

Contrary to filtering in the spatial domain, transform-based approaches involve the transformation of a noisy image into another domain where the transformed coefficients are processed. Usually, the coefficients are thresholded with respect to the standard deviation of noise in the image. Then, an inverse transform is applied to the processed coefficients to obtain the filtered or processed image as shown below:

$$\hat{x} = T^{-1}(\gamma(T(y))), \tag{3.9}$$

where $T(.)$ and $T^{-1}(.)$ represent the forward and inverse transformation, respectively, and $\gamma(.)$ is the thresholding function. The Fourier transform, the discrete cosine transform (DCT) and the wavelet transform are some well-established transforms in the context of denoising [85, 65, 86, 87, 88]. In order to process the transform coefficients hard thresholding [89], soft thresholding [90], and empirical Wiener shrinkage [91] are commonly used approaches.

Linear hard thresholding retains only those transform coefficients that are greater than τ^{hard}. Donoho and Jonestone [89] suggest $\tau^{hard} = \sigma_n\sqrt{2\log(M \times N)}$ as a good choice for the hard threshold, where M and N represent the number of rows and columns in the image, respectively. Although hard thresholding is effective in reducing the noise associated with the transform coefficients, it introduces artifacts as the retained coefficients may already be corrupted by noise. The associated basis functions of these coefficients appear similar to salt and pepper noise in the denoised image.

In soft thresholding, the threshold value τ^{soft} for a coefficient depends on its statistical property across windowed data samples. For example, in the case of the DCT, each coefficient may be thresholded differently using MAP [92] along with the assumption that the corresponding DCT coefficients across the overlapping patches follow a Laplacian distribution [93]. Soft thresholding on DCT coefficients $y^{DCT}[r,c]$ using the threshold $\tau^{soft}[r,c]$ (in the location $[r,c]$) can be performed as follows:

$$\gamma(y^{DCT}[r,c], \tau^{soft}[r,c]) = \begin{cases} y^{DCT}[r,c] + \tau^{soft}[r,c], & \text{if } y^{DCT}[r,c] < -\tau^{soft}[r,c], \\ 0, & \text{if } -\tau^{soft}[r,c] \leq y^{DCT}[r,c] \leq \tau^{soft}[r,c], \\ y^{DCT}[r,c] - \tau^{soft}[r,c], & \text{if } y^{DCT}[r,c] > \tau^{soft}[r,c]. \end{cases} \tag{3.10}$$

The soft threshold $\tau^{soft}[r,c]$ for the coefficient $[r,c]$ can, for example, be computed from the standard deviation of the corresponding noise-free coefficient $\breve{\sigma}^{DCT}[r,c]$ and the standard deviation of the AWGN σ_n using the following equation:

$$\tau^{soft}[r,c] = \frac{\sigma_n^2\sqrt{2}}{\breve{\sigma}^{DCT}[r,c]}, \tag{3.11}$$

where $\breve{\sigma}^{DCT}[r,c]$ is computed from the standard deviation of the noisy coefficients $\sigma^{DCT}[r,c]$ by

$$\breve{\sigma}^{DCT}[r,c] = \sqrt{{\sigma^{DCT}}^2[r,c] - \sigma_n^2}. \tag{3.12}$$

The effect of basis functions associated with the DCT appearing in the denoised image occurs even with soft thresholding, but it is less compared to hard thresholding.

Instead of thresholding the coefficients, shrinking them can prevent the salt and pepper noise like artifacts caused by the appearance of DCT basis functions. Empirical Wiener shrinkage [94] is a suitable choice for this task, as it shrinks the coefficients using the SNR:

$$\gamma(y^{DCT}[r,c], \tau^{ws}[r,c]) = \begin{cases} y^{DCT}[r,c] \times \tau^{ws}[r,c], & \text{if } \tau^{ws}[r,c] > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3.13}$$

where the shrinkage factor $\tau^{ws}[r,c]$ is given by

$$\tau^{ws}[r,c] = \frac{\breve{\sigma}^{DCT}[r,c]^2}{\breve{\sigma}^{DCT}[r,c]^2 + {\sigma_n}^2}. \tag{3.14}$$

Another well established shrinking approach is the BayesShrink proposed by Chang et al. [87]. It facilitates the estimation of (sub-band) adaptive thresholds. From a theoretical point of view, BayesShrink involves the soft-threshold τ^s that minimizes the Bayes risk assuming a Generalized Gaussian prior. Based on the thresholded wavelet coefficients $\hat{Y}$ and the regular wavelet coefficients Y, the Bayes risk is defined as

$$\gamma(\tau^s) = E(\hat{Y} - Y)^2. \tag{3.15}$$

The optimal threshold is the argument minimizing this function. Although no analytical solution for τ^s could be derived, Chang et al. [87] found an approximation closely matching their numerically obtained results:

$$\tau^s = \frac{\sigma_n^2}{\sigma_Y} \tag{3.16}$$

Here, σ_n^2 is the noise variance and σ_Y is the standard deviation of the signal's wavelet coefficients. Both values can be calculated from the image data. This facilitates a data-driven, sub-band-adaptive thresholding approach. At high noise levels, the basis functions associated with the transform may still appear in the denoised image.

3.2.4. Patch-based and Transform-based Approaches that Utilize Non-local Self-similarity

In patch-wise denoising, overlapping patches of size $K \times K$ are denoised separately and inserted back into an image. The corresponding pixels of the overlapping denoised patches are then averaged, before insertion. Non-local means (NL-means) [95], block matching 3D (BM3D) [96] and video block matching 3D (V-BM3D) [97] are some methods that fall into this category.

The NL-means algorithm, takes advantage of the high degree of redundancy in images. That is, a small patch in an image is likely to be associated with many similar patches in the same image. The neighborhood of a pixel $y[r, c]$ can be a set of pixels $y[r', c']$ represented by their row locations r' and column locations c'. In the neighborhood, the patch around many $y[r', c']$ and the patch around $y[r, c]$ are similar. All the pixels in the neighborhood can be used to estimate a new value for $y[r, c]$. The NL-means algorithm estimates the new value $y_{NL}[r, c]$ by computing the average of all the pixels in its neighborhood. This can be represented by

$$y_{NL}[r, c] = \Sigma_{r'} \Sigma_{c'} w_{NL}[r, c, r', c'] y[r', c'], \tag{3.17}$$

where $[r', c'] \in \Omega_R([r, c])$ and the normalized weights $w_{NL}[r, c, r', c']$ depend on the similarity between the pixels in the K $\times$ K patch around $y[r, c]$ as well as $y[r', c']$. The similarity measure is computed by weighing the Euclidean distance with a Gaussian filter (w_g), with standard deviation greater than 0. The weights are defined as follows:

$$w^{NL}[r, c, r', c'] = e^{-\frac{\sum_{i=-K/2}^{K/2} \sum_{j=-K/2}^{K/2} (y[r+i, c+j] - y[r'+i, c'+j])^2 \times w_g[i,j]}{h^2}}, \tag{3.18}$$

where the parameter h controls the degree of filtering. These weights are within [0,1] and are normalized such that they sum up to 1 as shown below:

$$w_{NL}[r, c, r', c'] = \frac{1}{S} w^{NL}[r, c, r', c'], \text{ with } S = \Sigma_{r'} \Sigma_{c'} w^{NL}[r, c, r', c'], \tag{3.19}$$

where $[r', c'] \in \Omega_R([r, c])$. In order to find similar pixels, the NL-means algorithm not only compares the gray level in single pixels but also compares the gray level of all the pixels in the patches around respective pixels. This makes the computations of similar pixels robust and ultimately leads to good denoising.

BM3D [96] has been proposed as an improvement to NL-means. It combines patch-based and a transform-based denoising. The algorithm is divided into two major steps: computation of an initial estimate of the denoised image followed by the computation of a refined estimate. In the first step, for each overlapping reference patch $\boldsymbol{p} \in \mathbb{R}^{k \times k}$, similar patches are found and grouped into a set $\boldsymbol{P} \in \mathbb{R}^{k \times k \times n}$. Then collaborative filtering is performed on the sets. In this process, a 3D isometric linear transformation T_{3D} is applied on these sets [96]. Then, the transformed coefficients are thresholded using a hard thresholding function γ_{3D} to reduce noise and finally an inverse transform T_{3D}^{-1} is applied to return to the image domain. This process is represented by

$$\hat{\mathbf{P}} = T_{3D}^{-1}(\gamma_{3D}(T_{3D}(\mathbf{P}))), \tag{3.20}$$

where

$$\gamma_{3D}(s) = \begin{cases} 0, & \text{if } s < 0, \\ s, & \text{otherwise.} \end{cases} \tag{3.21}$$

In order to reduce the amount of computations, the 3D transform is performed by a 2D transform on each patch, and subsequently by a 1D transform along the third dimension of the 3D group. Collaborative filtering provides an estimate for each used patch. Thus a variable number of estimates for every pixel is obtained. The estimates for a particular pixel that is to be calculated are weighted, based on the characteristics of the patch it belongs to. Patches containing an edge will be given less importance, compared to homogeneous ones [96]. The denoised image obtained at the end of this step is only a basic estimate. The basic estimate is then used for associating similar patches. In the second step, Wiener filtering of the patches of the original image is performed, instead of hard thresholding. Experiments show that by following such an approach, more details are restored and the denoising performance is improved. Weights or the empirical Wiener coefficients are calculated based on the grouped similar patches. Subsequently, collaborative filtering is performed on the grouped noisy patches using the calculated weights. Similar to the first step, weighted aggregation is performed on the estimates to compute the final denoised image.

As mentioned previously, V-BM3D [97] is an extension to BM3D targeted at denoising video sequences. Contrary to BM3D, where similar patches are obtained within a single frame, in V-BM3D, patches are obtained within a single frame, as well as across frames [97]. Though these methods perform better than transform-based approaches applied on pixels, they may still suffer from artifacts caused due to thresholding of transform coefficients. Moreover, they are computationally expensive.

3.2.5. Sparse Modeling and Low-rank Approximation

The motivation behind sparse modeling is to find a representation for an unobserved high-dimensional signal from a limited number of observations, for example decoding a signal from its noisy encoding. The ground-truth solution (e.g., the noise free estimate) can be well approximated by a sparse vector where only a few variables are truly important. Even if the original problem specification does not yield a sparse solution, one can typically find a mapping to a new coordinate system or a dictionary that allows for a sparse representation [98]. Algorithms in the context of sparse modeling and low-rank approximation are often implemented using patch-based methods. In the denoising methods that make use of dictionary learning, a noisy patch is approximated from a sparse linear combination of the atoms or columns of a dictionary D, resulting in a general model:

$$\hat{\alpha} = \arg\min_{\boldsymbol{\alpha}} ||y - \boldsymbol{D\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1, \tag{3.22}$$

where α is a matrix containing vectors of sparse coefficients and λ is the regularization parameter [99, 100]. The sparse representation model or the dictionary $\boldsymbol{D}$ is computed from a dataset as well as the image itself using the K-singular value decomposition (K-SVD) algorithm by solving the joint optimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\hat{y}, \boldsymbol{D}, \boldsymbol{\alpha}} \gamma ||y - \hat{y}||_2^2 + \sum_i ||\boldsymbol{R_i}\hat{y} - \boldsymbol{D\alpha_i}||_2^2 + \sum_i \lambda_i ||\boldsymbol{\alpha_i}||_1, \tag{3.23}$$

where $\boldsymbol{R_i}$ is the matrix for extracting patch $\hat{y}_i$ from the estimated image $\hat{y}$ at location i, and λ_i and γ are regularization factors [101, 102]. Learned dictionaries are more flexible in representing image structures [103]. Hence learned dictionaries perform better than designed dictionaries. This approach has been further improved by coupling with the sparsity from the self-similarity properties of optical images [104, 105, 106].

On the other hand, low-rank-based approximation focuses on recovering the noise-free estimate of a patch based on a set of similar patches. In other words, the patch to be denoised $\boldsymbol{p} \in \mathbb{R}^{k \times k}$ as well as $n-1$ similar patches are first vectorized and stacked into a matrix $\boldsymbol{P} \in \mathbb{R}^{k^2 \times n}$. Then, to estimate the noise-free data $\hat{\boldsymbol{P}}$, a low-rank approximation is performed on the matrix. Low-rank estimation approaches for the reconstruction of a signal from noisy estimates can be grouped into two main categories: (i) methods based on low-rank factorization [107, 108, 109, 110, 111, 112], and (ii) methods based on nuclear norm minimization (NNM) [113, 114, 115, 116].

Methods belonging to the first category approximate a data matrix as a product of matrices of fixed rank. Since singular value decomposition (SVD) results in optimal energy compaction in the least square sense, it is a suitable choice for performing the decomposition and estimating the low-rank matrix [117]. Using SVD, $\boldsymbol{P}$ is first decomposed into a singular value matrix $\boldsymbol{S} \in \mathbb{R}^{k^2 \times n}$ with singular values s_i present in decreasing order in the diagonal elements (i.e., the largest singular value is present in the first singular vector s_0 and the smallest in the last s_n) and the associated singular vector matrices $\boldsymbol{U} \in \mathbb{R}^{k^2 \times k^2}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times n}$. Then, the low-rank matrix can, for example, be obtained by processing the singular values in the BayesShrink framework [109]. However, for this approach the rank of the matrix must be known a-priori. In addition, the processing of the singular values that are too low or too high might result in the loss of details or the preservation of noise.

Methods based on NNM aim to find the lowest rank approximation of the data matrix by solving

$$\hat{\boldsymbol{P}} = arg\min_{\boldsymbol{P'}} ||\boldsymbol{P} - \boldsymbol{P'}||_F^2 + \lambda ||\boldsymbol{P'}||_*, \tag{3.24}$$

where $||.||_F$ denotes the Frobenius norm and $||.||_*$ the nuclear norm. $||\boldsymbol{P'}||_*$ is given by $\sum_i ||s_i(\boldsymbol{P'})||_1$, where $s_i(\boldsymbol{P'})$ is the i-th singular value of $\boldsymbol{P'}$. The solution is a low-rank approximation obtained by taking only a few large singular values and the corresponding singular vectors. This can be represented by

$$\hat{\boldsymbol{P}} = \boldsymbol{U} f_\lambda(\boldsymbol{S}) \boldsymbol{V^T}, \tag{3.25}$$

where $\boldsymbol{P} = \boldsymbol{USV^T}$ is the SVD of $\boldsymbol{P}$ and $f_\lambda(\boldsymbol{S})$ is the singular value thresholding operator. The choice of the largest l singular values required for obtaining the low-rank approximation, is computed by solving the following equation [117]:

$$\sum_{i=l}^{n} s_i^2 > \sigma_{\eta'}^2 \times (k^2 \cdot n - 1) \geq \sum_{i=l+1}^{n} s_i^2, \tag{3.26}$$

where $\sigma_{\eta'}$ is the standard deviation of Gaussian noise associated with $\boldsymbol{P}$. The low-rank matrix $\hat{\boldsymbol{P}} \in \mathbb{R}^{k^2 \times n}$ can then be constructed from the l largest singular values as follows:

$$\hat{\boldsymbol{P}} = \sum_{i=1}^{l} s_i \mathbf{u_i} \mathbf{v_i^T}. \tag{3.27}$$

In the case of NNM, all the singular values are considered to be equally important. Hence, the same threshold is applied to each singular value. However, different singular values have different levels of importance. On this basis, a weighted approach known as weighted nuclear norm minimization (WNNM) was presented in [113, 114], where weights were adaptively assigned to different singular values and denoising was carried out by applying a soft threshold on the weighted singular values. In other words, given a weight vector $\boldsymbol{w}$, the weighted nuclear norm problem is based on finding an approximation $\hat{\boldsymbol{P}}$ of $\boldsymbol{P}$ that minimizes the following cost function:

$$\hat{\boldsymbol{P}} = \arg\min_{\boldsymbol{P'}} ||\boldsymbol{P} - \boldsymbol{P'}||_F^2 + ||\boldsymbol{P'}||_{\boldsymbol{w},*}, \tag{3.28}$$

where $||\boldsymbol{P'}||_{\boldsymbol{w},*} = \sum_i ||w_i s_i(\boldsymbol{P'})||_1$ is the weighted nuclear norm of $\boldsymbol{P'}$. The variable w_i is the weight assigned to the singular value $s_i(\boldsymbol{P'})$. Equation 3.28 has a unique global minimum when the weights satisfy $0 \leq w_1 \leq w_2 \leq \ldots \leq w_n$ [113]. The WNNM algorithm has been shown to achieve advanced denoising performance and is more robust to noise compared to other methods based on NNM [113]. These approaches perform better than patch- and transform-based approaches, however, they are computationally expensive.

3.3. Learning-based Approaches

Analytical denoising approaches, such as prior-based methods [113, 114, 96], usually rely on solving complex optimization problem to find an optimal solution for the denoising

problem. Therefore, these methods usually involve time-consuming and iterative inference. In addition, the models used require several manually chosen parameters that have to be tuned for getting the best denoising performance. To overcome the limitations of analytical approaches, discriminative learning methods have been developed to learn image prior models and eliminate the iterative optimization procedure during the application phase [118]. On the other hand, learning-based denoising has also been treated as a plain discriminative learning problem, i.e., separating noise from a noisy image by feed-forward convolutional neural networks (CNNs) [119]. CNNs have been considered to be a suitable choice due to the following reasons: (i) CNNs with deep architecture are flexible and effective in increasing the capacity and flexibility for exploiting image characteristics [120], and most importantly it is easy to modify the depth of a CNN, (ii) continuous advances in the design of network architectures and components, and (iii) CNNs are well suited for parallel computation on modern powerful GPUs. Furthermore, adversarial learning approaches that make use of CNNs have been designed to generate images with certain signal and noise characteristics [121]. A brief overview of these methods is presented in this section.

3.3.1. Discriminative Learning-based Methods

In this category, the restoration of the noise-free image x from its corrupted observation y (based on the noise corruption model presented in Eq. 3.1) is considered by combining an observation likelihood and an image prior invoking Bayes' rule [122], given as

$$\begin{aligned} p(x|y) &\propto p(y|x) \cdot p(x) \\ &\propto \mathcal{N}(y; Kx, \lambda) \cdot \prod_{i=1}^{N} \prod_{c \in \mathcal{C}} exp(-p_i(\boldsymbol{f}_i^T \boldsymbol{x}_{(c)})). \end{aligned} \tag{3.29}$$

In Eq. 3.29, the corruption process $\mathcal{N}(.)$ is modeled with a Gaussian likelihood, where $Kx \equiv k \otimes x$ denotes the convolution of x with a kernel k, and λ is related to the strength of the assumed additive Gaussian noise [122]. Regularization is provided through a Markov random field model [123] with N robust potential functions $p_i(.)$ that model the responses $\boldsymbol{f}_i^T \boldsymbol{x}_{(c)}$ of filter $\boldsymbol{f}_i$ over all regions $c \in \mathcal{C}$ of the image x [122]. The posterior distribution is given by $p(x|y) \propto exp(-E(x|y))$, where $E(x|y)$ can be expressed by the associated Gibbs energy

$$E(x|y) = \frac{\lambda}{2}||y - Kx||^2 + \sum_{i=1}^{N} \sum_{c \in \mathcal{C}} p_i(\boldsymbol{f}_i^T \boldsymbol{x}_{(c)}). \tag{3.30}$$

In order to minimize Eq. 3.30, independent auxiliary z_{ic} are introduced for all filter responses $\boldsymbol{f}_i^T \boldsymbol{x}_{(c)}$ to obtain an augmented energy function $E(x, z|y)$ such that

$$\arg\min_{x} E(x|y) = \arg\min_{x,z} E(x, z|y). \tag{3.31}$$

In order to solve for $\arg\min_{x,z} E(x, z|y)$, a block coordinate descent strategy is used. The strategy alternates between minimizing with respect to x and z. In addition, this strategy has a faster convergence compared to minimizing $\arg\min_x E(x|y)$ directly [122]. The auxiliary variables are introduced in such a way that $E(x|z, y)$ becomes a quadratic

function. In this case, minimizing $E(z|x, y)$ will amount to solving many independent univariate optimization problems. That is, each iteration of the algorithm uses a different quadratic relaxation $E(x, z|y)$ determined by z in order to solve the original objective function. This approach was introduced first in [124, 125] under the name half-quadratic regularization and can be further categorized into additive [124] and multiplicative [125] forms. A main computational difference between the two forms is that for the multiplicative form $E(x|z, y) = \mathbf{\Omega}(z, y)^{-1}\boldsymbol{\eta}(y)$, whereas, for the additive form $E(x|z, y) = \mathbf{\Omega}(y)^{-1}\boldsymbol{\eta}(z, y)$. The variable $\mathbf{\Omega} \in \mathbb{R}^{B\times B}$ is a sparse system matrix with B being the number of pixels and the variable $\boldsymbol{\eta} \in \mathbb{R}^B$ is a vector. This implies that the quadratic function can be minimized by solving a system of linear equations. In the case of the multiplicative form z influences only $\mathbf{\Omega}$, whereas, in the additive form, z influences only the right hand $\boldsymbol{\eta}$ of the equation system [122].

The parameters $\boldsymbol{\eta}$ and $\mathbf{\Omega}$ are determined from model parameters θ (learned from the observed images y and their noise-free counterparts x), the observed image y and the point spread function (PSF) k [122]. The model parameters θ_t at each iteration t of the algorithm are learned by minimizing the cost function:

$$J(\theta_t) = \sum_{s=1}^{S} l(\hat{x}_t^{(s)}, x^{(s)}), \tag{3.32}$$

with $\hat{x}_t^{(s)}$ being the result of processing $y^{(s)}$ and $x^{(s)}$ the associated noise-free ground truth. For performing the training, S pairs of noisy and noise-free data are used. In addition, the PSF k associated with the images is considered during the optimization in the case of blind processing [122]. Continuously differentiable loss functions such as the negative peak signal-to-noise ratio (PSNR) are used as a cost function (see Sec. 3.4.2). In order to minimize Eq. 3.32, a gradient-based method is relied on [122]. That is, the loss is differentiated with respect to model parameters and the parameters are updated based on the estimated loss.

Although discriminative learning-based methods have shown promising results toward bridging the gap between computational efficiency and denoising quality, their performance is inherently restricted to the specified forms of prior information (based on the analysis model), which are limited in capturing the full characteristics of image structures. Moreover, the parameters are learned by stage-wise training combined with joint fine-tuning across all the stages. Furthermore, similar to the analytical approaches specified in Sec. 3.2 these methods involve many handcrafted parameters. Another important drawback is that these approaches involve the training of a specific model for a certain noise level, and are hence limited in blind image denoising [118].

3.3.2. Denoising using CNNs

CNNs are an integral part of deep learning that are well known for processing data having a grid pattern, e.g., images. Their design is inspired by the organization of an animal's visual cortex with the goal of automatically and adaptively learning a spatial hierarchy of features, from low- to high-level patterns [126]. They are widely used for tasks such as data analysis [127], object detection and classification [128], segmentation [129], image enhancement [118] and image synthesis [130]. A CNN typically consists of three types of

layers namely convolutional, pooling and fully connected layers. The convolutional and pooling layers perform the task of feature extraction. The fully connected layers are used for mapping the extracted features into an output, e.g., for the task of classification. In order to train a CNN to perform the required task, the different parameters associated with the network, e.g., filter kernels or weights, need to be learned. This is performed by setting up a loss function or a cost function between the output of the network and the ground-truth. The parameters are updated to minimize the output of the loss function. To this end, the backpropagation algorithm is used [131]. It works by computing the gradient of the loss function with respect to each weight using the chain rule, i.e., the gradient is computed one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms [131]. The different components of a CNN relevant to this thesis are briefly described in the following paragraphs. A comprehensive overview of CNNs can be found in [132, 133, 134].

Basic Components of CNNs

As mentioned above, a CNN consists of convolutional, pooling and fully connected layers. A convolutional layer is composed of a stack of linear mathematical operations, such as convolutions. This layer receives an array of numbers (also known as a tensor) as input and provides the transformed input as output (also known as a feature map). The transformation is performed by convolving (actually correlating) the input with a filter kernel whose parameters can be learned to optimally perform the given task. A convolutional layer can have more than one filter kernel. Therefore, multiple kernels can be used to form an arbitrary number of feature maps, which represent different characteristics of the input tensors. The two key hyperparameters that define the convolution operation are the size and the number of kernels. The number of kernels is arbitrary and determines the depth of output feature maps [126]. Due to the filtering operation, the size of the output tensors will be smaller than that of the input depending on the size of the filter kernels. If the size of the feature maps should be same as the input, the input tensor can be padded before the convolution is performed. Typically, zero-padding is applied but other options are also available, e.g., mirroring. The size and the number of kernels have an influence on the receptive field, the region in the input space that a particular feature of the CNN is exposed to. Since every unit or neuron in a convolutional layer represents the response of a filter applied to the previous layer, deeper networks will result in a larger receptive field.

The output of the convolutions are then passed through a nonlinear activation function. Previously, smooth nonlinear functions such as sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{3.33}$$

or the hyperbolic tangent

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}, \tag{3.34}$$

have been used as they are mathematical representations of a biological neuron behavior. The sigmoid and the hyperbolic tangent functions are bounded and do not result in the activations to explode. But they may cause the gradients (used in backpropagation) to vanish. In addition, one needs to normalize the data carefully. The rectified linear unit

(ReLU) is considered a better alternative and shows superior convergence performance [119]. The ReLU is defined by

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{otherwise.} \end{cases} \tag{3.35}$$

The ReLU is computationally more efficient and is significantly less affected by the vanishing gradient problem. Moreover, it is not as sensitive as the above activation functions with respect to network intitialization and data normalization. However, when too many activations get below zero, most of the units in the network with ReLU will simply output zero. This will prohibit learning. This problem can be handled to some extent by the Leaky ReLU

$$f(x) = \begin{cases} 0.01x, & \text{if } x < 0 \\ x, & \text{otherwise.} \end{cases} \tag{3.36}$$

Here, the output is no longer zero when the input is less than 0 but a small number. Parametric rectified linear unit (PReLU) takes this idea further by making the coefficient of leakage into a parameter a that is learned along with the other neural-network parameters. It is defined by

$$f(x) = \begin{cases} ax, & \text{if } x < 0, \\ x, & \text{otherwise.} \end{cases} \tag{3.37}$$

The exponential linear unit (ELU), defined by

$$f(x) = \begin{cases} x, & \text{if } x > 0, \\ a(e^x - 1), & \text{otherwise,} \end{cases} \tag{3.38}$$

the sigmoid linear unit (SiLU), defined by

$$f(x) = \frac{x}{1 + e^{-x}}, \tag{3.39}$$

and the Gaussian Error Linear Unit (GELU), defined by

$$f(x) = x\phi(x), \tag{3.40}$$

are some nonlinear variants of the ReLU [135, 136, 137]. In Eq.3.40, $\phi(x)$ is the cumulative distribution function of the standard normal distribution [135].

A pooling layer performs a downsampling operation, i.e., it reduces the in-plane dimensionality of the feature maps and, as a consequence, reduces the number of subsequent learnable parameters. However, it is to be noted that pooling layers do not contain any learnable parameters. But there are hyperparameters such as filter size, stride, and padding that can be tuned. Max pooling is currently the most popular form of pooling operation, which extracts the maximum value of a region and discards all the other values. For example, max pooling with a filter of size 2×2 and a stride of 2 downsamples the in-plane dimension of feature maps by a factor of 2. The height and the width of the feature maps change, whereas, the depth remains unaltered. Global average pooling is another pooling operation. Here, the average of the region is taken while downsampling instead

of the maximum value. Both, max pooling as well as averaging introduce translation invariance to small shifts and distortions.

For a regression task, such as denoising, it is desirable for the spatial dimensions of the input and the output of the network to be the same, if not, there will be a loss in information. Hence, downsampled feature maps need to be upsampled. There are several methods for performing upsampling such as the nearest neighbor interpolation, bi-linear interpolation and bi-cubic interpolation. Max unpooling is another option. In max unpooling, first, the index of the maximum value is saved for every max-pooling layer. The saved index is then used during the upsampling step where the input pixel is mapped to the saved index and the other indices are filled with zeros. It is to be noted that all the above-mentioned techniques are predefined and do not depend on data, and this makes them task-specific. In addition, they do not learn from data and hence they are not a generalized technique. To enable the network to perform optimal upsampling, it needs to be learnable. Hence, transposed convolutions have been introduced [138]. Here filters are learned to perform deconvolution and upsample the input feature map to a desired output feature map.

In order to make the training of the networks faster (i.e., speed up the convergence) and more stable, batch normalization (BN) has been introduced [139, 140, 141]. It normalizes the output of the hidden layers of the current batch using its mean μ_B and variance σ_B^2 such that the output has standard normal distribution. In BN, first the input is normalized by

$$Z_{\text{norm}}^{(i)} = \frac{Z^{(i)} - \mu_B}{\sqrt{\sigma_B^2 - \epsilon}}, \tag{3.41}$$

where $Z^{(i)}$ is the input, $Z_{\text{norm}}^{(i)}$ is the normalized value and ϵ is a small positive infinitesimal quantity. Then, a linear transformation is applied to $Z_{\text{norm}}^{(i)}$ with two trainable parameters γ_B and β_B. The transformed value $Z_{\text{BN}}^{(i)}$ is given by

$$Z_{\text{BN}}^{(i)} = \gamma_B \times Z_{\text{norm}}^{(i)} + \beta_B. \tag{3.42}$$

This step with the learnable parameters allows the model to choose the optimum distribution for each hidden layer as γ_B allows for the standard deviation to be adjusted and β_B aids in controlling the bias. BN can be applied right before or right after the nonlinear activation function.

The learning strategies relevant to this thesis can be broadly classified into supervised learning, self-supervised learning and unsupervised learning. In supervised learning, a CNN is trained to find a mapping between the input and the target data. In the context of denoising, noisy images are the input and the associated noise-free images are the ground-truth. In self-supervised and unsupervised learning, a-priori information regarding the characteristics of the images is used for training the CNN to perform denoising [142, 143, 144, 145]. For example, autoencoders [142] take the input data, compress it into a code (latent variable), and then try to recreate the input data from that summarized code. The recreated data is denoised as the network cannot replicate the stochastic noise in the input [142]. Unsupervised approaches also involve the learning of the mapping of the input based on the characteristics (e.g., noise properties) associated with the target images [121].

The learning strategy and the network architecture have been designed based on the scenario, i.e., the task and the availability of data for the training phase. A brief overview of the strategies used in different scenarios is presented in the following sections:

Scenario: Spatially Aligned Pairs of Noisy and Noise-free Images are available

When pairs of spatially aligned noisy and noise-free data are available, supervised approaches are preferred. The goal here is to train a CNN to find a mapping between the noisy and noise-free ground-truth images. For images of static objects, noise-free images can, for example, be computed by averaging several noisy acquisitions. However, in practice, an adequate amount of noise-free images and real noisy images is often not available. Hence, in such cases, training data is generated by adding white noise to the available data [146, 118]. As loss functions, mean squared error and mean absolute error are commonly used.

In one of the most established learning-based denoising approaches [118], a deep CNN architecture has been proposed for performing the task of denoising. Rather than directly yielding the denoised image, the approach is designed to predict the residual image, i.e., the difference between the noisy observation and the latent noise-free image. The loss function is given by

$$l(\theta) = \frac{1}{2P} \sum_{i=1}^{P} \left\| \left(D(y^{(i)}, \theta) - (x^{(i)} - y^{(i)}) \right) \right\|_F^2, \tag{3.43}$$

where D represents the network, $D(y^{(i)}, \theta)$ represents the output of processing the sample $y^{(i)}$ (with ground truth $x^{(i)}$) using D and $||.||_F$ denotes the Frobenius norm. As shown in Fig. 3.1, the network has three types of layers: (i) convolution followed by activation using ReLU (Conv+ReLU), (ii) Convolution followed by BN and activation using ReLU (Conv+BN+ReLU) and (iii) convolution. Type (i) is used as the first layer of a network with D_L layers. 64 filters of size 3×3×c are applied to generate 64 feature maps, where c represents the number of channels. Type (ii) is used for layers 2 to (D_L - 1) where 64 filters of size 3 × 3 are used. Finally, type (iii) for the last layer with c filters of size 3 × 3 are involved to reconstruct the output. The approach implicitly removes the latent noise-free image with the operations in the hidden layers and only yields the noise image. Furthermore, batch normalization is introduced to stabilize and enhance the training performance. It has been shown in [118] that residual learning and batch normalization can benefit from each other, and their integration is effective in speeding up the training and boosting the denoising performance.

Wang et al. [147], Bae et al. [148] and Jifara et al. [149] also presented residual learning methods along with deeper CNNs for image denoising. However, as the approaches relied on deeper architectures with several layers, they result in the prior states or layers having little influence on the subsequent ones. Tai et al. [150] have proposed a very deep persistent memory network that introduces a memory block, consisting of a recursive unit and a gate unit, to explicitly mine persistent memory through an adaptive learning process. The recursive unit learns multi-level representations of the current state under different receptive fields. The representations and the outputs from the previous memory blocks

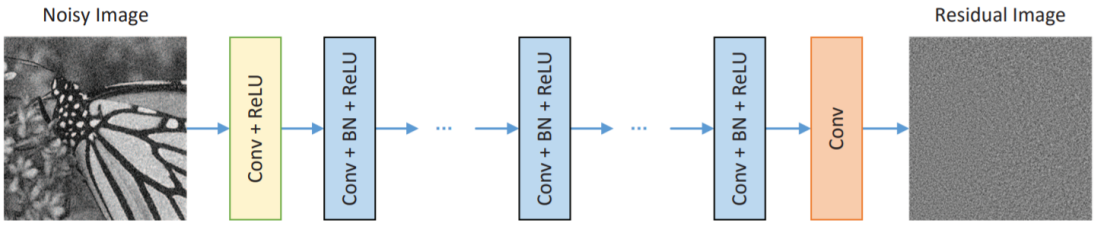


Figure 3.1.: Network architecture of DnCNN [118] including an example of a noisy image used as input to the network and the resulting residual image obtained as output. The denoised image is obtained by subtracting the residual image from the noisy image. DnCNN consists of convolutional filters, rectified linear unit (ReLU) activations and batch normalization (BN).

are concatenated and adaptively controlled by the gate unit, which decides how much of the previous states should be reserved and how much of the current state should be stored.

In order to cope with the limited availability of training data in certain domains, several methods have been attempted to extract more useful information by taking into consideration prior knowledge of the data [151, 152, 153]. For example, Yuan et al. [154] combined a deep CNN, residual learning and multi-scale knowledge to remove the noise from hyperspectral data. However, these CNNs usually involve increased computational costs and memory requirements. To address these issues, Gholizadeh et al. [155] proposed the use of dilated convolutions [156] to enlarge the receptive field and reduce the depth of network. Lian et al. [157] presented a residual network via multi-scale cross-path concatenation to perform denoising. This demonstrates that improving the network architecture can result in superior denoising.

A modification to the network architecture usually involves the following changes: (i) fusing the features from multiple inputs of a CNN, (ii) modifying the depth or the width of the CNN for enlarging the receptive field size [158, 159, 160], (iii) adding auxiliary plugins, (iv) introducing skip connections or cascade operations to provide complementary information to different layers [161, 162, 163, 157, 164] and (v) changing the loss function. Category (i) includes providing different parts of one sample as multiple inputs for different networks [165], different perspectives on the one sample as input, such as multiple scales [166, 167, 168]. Category (iii) involves different activation functions, dilated convolutions and pooling operations as well as their placement [169, 170]. Finally, different loss functions are defined according to the characteristics of the images to extract more robust features [171]. Perceptually motivated image quality metrics, such as structural similarity index measure (SSIM) and its different variants, have also been used [172, 173, 174]. Chen et al. [175] proposed joint Euclidean and perceptual loss functions to extract more edge information.

Scenario: Spatially Aligned Pairs of Noisy and Noise-free Images are not available

When noise-free images are not available or when it is not possible to generate spatially aligned pairs of noisy and noise-free images, a change in the training strategy is required. Lehtinen et al. [176] have shown that a denoising network can be trained by presenting

the network with spatially aligned noisy images alone. As network architecture they used a U-Net [129]. The architecture has an encoder (contraction path) and a decoder part (expansion path) as shown in Fig. 3.2. In the encoder, downsampling blocks are repeatedly applied one after another, where each block comprises convolutions using 48 3×3 filters, ReLU and 2× 2 max pooling. In the decoder, upsampling blocks are repeatedly applied. As input to these blocks, complementary information is provided from the associated encoding blocks through skip connections. In these blocks, convolutional layers, activation using ReLU and transpose convolutions (for upsampling) are used. As shown in Fig. 3.2, the final block has convolutional layers followed by activation through ReLU. This combination is repeated twice. Then, convolutional filters are applied to the output and as activation, a leaky ReLU with $\alpha = 0.1$ is used to get the denoised image. The performance of the denoising network is similar – sometimes even better – compared to the case where noise-free images are used in the training phase [176]. In this case, neither an explicit statistical likelihood model of the corruption nor an image prior is required.

Similar approaches that require neither noisy pairs nor noise-free image targets have been presented in [145, 144]. They come under the category of self-supervised approaches and make two simple statistical assumptions: (i) the signal is not pixel-wise independent and (ii) the noise is conditionally pixel-wise independent given the signal. In the approaches, both parts of the training sample, the input and the target, are derived from a single noisy training image. However, if the same patch is extracted as input and its center pixel is used as target, the network will learn the identity by directly mapping the value at the center pixel to the output. To prevent this, the authors have presented a blindspot network. The receptive field of this network has a blind-spot in its center. Therefore, the prediction for a pixel is affected by all input pixels in a square neighborhood except by the input pixel at its very location (center pixel). The blindspot network has a slightly less information available for the predictions. Due to this reason, its accuracy is expected to be slightly impaired compared to a normal network. As only one pixel out of the entire receptive field is removed, the approach is expected to perform reasonably well, but not as effective as a traditional training approach. Such self-supervised denoising approaches have been further developed in [177, 143].

For training denoising networks in an unsupervised manner, generative adversarial networks (GANs) have been employed [178, 179]. The idea of GANs is to have a generator, which is trained to generate the desired image from noisy input, and a discriminator, which is trained to discriminate between the noise-free image and the generated image. The generator and discriminator are adversarial. They are trained together so that after training, the generator would be good at generating images that look authentic [130, 180]. This training strategy does not require spatially aligned pairs of noisy and noise-free images. However, the network needs to be presented with noisy and noise-free images so that it learns the characteristics associated with them. The Wasserstein distance is an example of the loss function that is used as a discrepancy measure between distributions and a perceptual loss that computes the difference between images in an established feature space [181]. GANs have also been used to generate noise and noisy images from noise-free images. [175]. The training pairs can then be used for performing supervised training. Kang et al. [121] have proposed a cycle-consistent adversarial denoising network to learn the mapping between two domains, low-dose and high-dose computed tomography images. The approach involves cycle consistency loss [182] that is typically used for

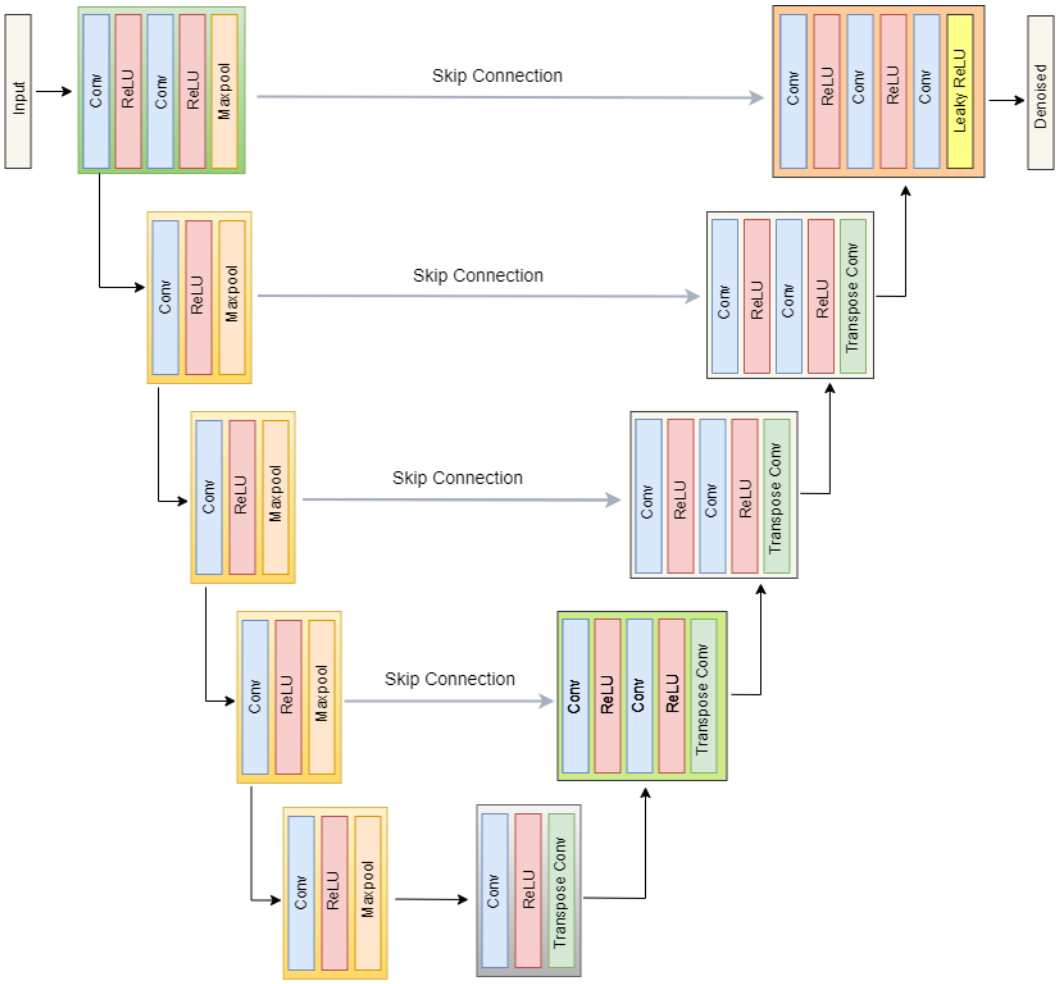


Figure 3.2.: Network architecture of U-Net used in [176]. It is a convolutional neural network that has a contracting path and an expansive path giving it the u-shaped architecture. It consists of downsampling or convolutional filters (conv) and upsampling or transpose-convolutional filters, rectified linear unit (ReLU) activations and maxpooling operations. Filters of size 3×3 and 2×2 are used. Skip connections or cascade operations (copy) are used to provide complementary information to different layers.

unpaired image-to-image translation.

The components of such a training strategy are presented in Fig. 3.3. In this approach, mapping functions between two domains low-dose X-ray images Y_{LD} and high-dose X-ray images Y_{HD} are learned given training samples $y_{LD} \in Y_{LD}$ and $y_{HD} \in Y_{HD}$. The data distributions are denoted by $y_{LD} \sim p_{data}(y_{LD})$ and $y_{HD} \sim p_{data}(y_{HD})$. The model includes two mapping functions or generators: (i) G_{HD} for mapping data in Y_{LD} to Y_{HD} and (ii) G_{LD} for mapping data in Y_{HD} to Y_{LD}. In addition, two adversarial discriminators are used: (i) D_{HD} for distinguishing between $G_{HD}(y^{LD}, \theta)$ and y^{HD} and (ii) D_{LD} for distinguishing between $G_{LD}(y^{HD}, \theta)$ and y^{LD}. The objective contains two types of terms: adversarial

losses for matching the distribution of generated images to the data distribution in the target domain given by

$$\begin{aligned} L_{ADV}(G_{HD}, D_{HD}, Y_{LD}, Y_{HD}) &= \mathbb{E}_{y_{HD}\sim p_{data}(y_{HD})}\left[log(D_{HD}(y_{HD}, \theta)\right], \\ L_{ADV}(G_{LD}, D_{LD}, Y_{HD}, Y_{LD}) &= \mathbb{E}_{y_{LD}\sim p_{data}(y_{LD})}\left[log(D_{LD}(y_{LD}, \theta)\right], \end{aligned} \tag{3.44}$$

and cycle consistency losses to prevent the learned mappings G_{HD} and G_{LD} from contradicting each other defined by

$$\begin{aligned} L_{CYC}(G_{HD}, G_{LD}) = \mathbb{E}_{y_{HD}\sim p_{data}(y_{HD})}&\left[\left\|\Big(G_{HD}(G_{LD}(y_{HD}, \theta), \theta) - y_{HD}\Big)\right\|_1\right] \\ + \mathbb{E}_{y_{LD}\sim p_{data}(y_{LD})}&\left[\left\|\Big(G_{LD}(G_{HD}(y_{LD}, \theta), \theta) - y_{LD}\Big)\right\|_1\right]. \end{aligned} \tag{3.45}$$

The idea is to enforce the intuition that these mappings should be reverses of each other and that both mappings should be bijections. This loss reduces the space of possible mapping functions learned by the network by enforcing forward and backward consistency. However, it is to be noted that for this strategy, high-dose X-ray images would be needed. Such images may be either not available or, if available, would be a limiting factor, since the results would at the most be as good as the image quality of the high-dose images used during training.

3.4. Image Quality Metrics

In order to measure the quality of the denoised image, approaches such as the PSNR, the SSIM and the contrast-to-noise ratio (CNR) are commonly used. For these metrics, a higher value indicates better performance. In this section, these approaches are briefly described.

3.4.1. Structural Similarity Index (SSIM)

The SSIM measures the similarity between an image of perfect quality, noise-free ground truth x in this case, and another image, e.g., a denoised image $\hat{x}$. The SSIM metric is assessed based on a luminance term l_t, contrast term c_t, and structure term s_t. It is computed as:

$$\begin{aligned} \text{SSIM}\,(x, \hat{x}) &= [l_t(x, \hat{x})]^{\alpha_t} * [c_t(x, \hat{x})]^{\beta_t} * [s_t(x, \hat{x})]^{\gamma_t} \\ &= \left(\frac{2\mu_{\hat{x}}\mu_x + c_1}{\mu_{\hat{x}}^2 + \mu_x^2 + c_1}\right)^{\alpha_t} \left(\frac{2\sigma_x\sigma_{\hat{x}} + c_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2}\right)^{\beta_t} \left(\frac{\sigma_{x\hat{x}} + c_3}{\sigma_x\sigma_{\hat{x}} + c_3}\right)^{\gamma_t} \end{aligned} \tag{3.46}$$

where μ_x, $\mu_{\hat{x}}$, σ_x, $\sigma_{\hat{x}}$ and $\sigma_{x\hat{x}}$ are the local means, standard deviations, and cross-covariance for images x and $\hat{x}$, respectively. c_1, c_2 and c_3 are regularization constants for the luminance, contrast and structural terms, respectively, and have been set to

$$\begin{aligned} c_1 &= (.01 * D_R)^2 \\ c_2 &= (.03 * D_R)^2 \\ c_3 &= c_2/2 \end{aligned} \tag{3.47}$$

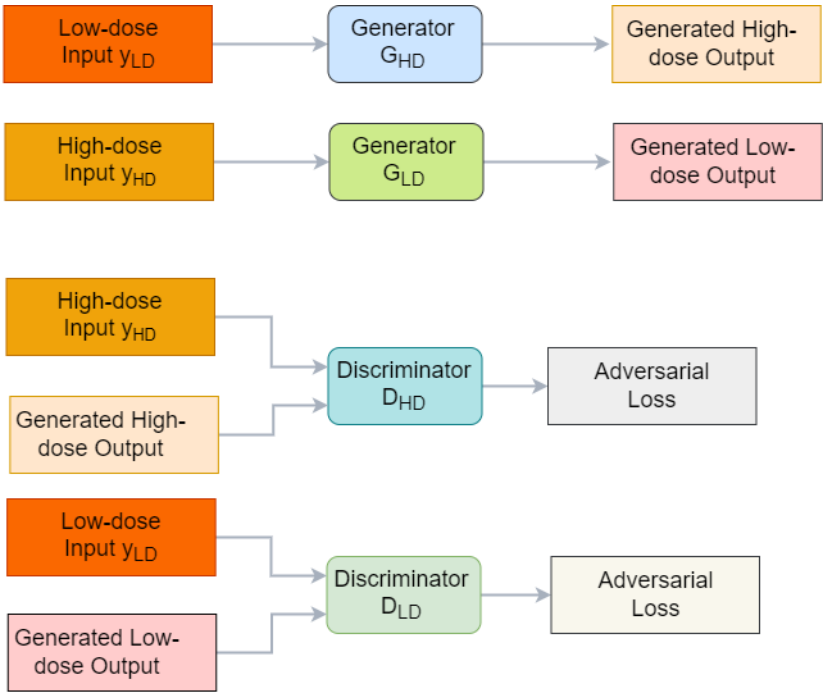


Figure 3.3.: Components of generative adversarial networks. There are two generators G_{HD}, for mapping low-dose input to high-dose, and G_{LD}, for mapping high-dose input to low-dose. There are two generators D_{HD}, for discriminating high-dose input from the high-dose output generated by G_{HD}, and and D_{LD}, for discriminating low-dose input from the low-dose output generated by G_{HD}.

where D_R is the dynamic range of the image x. The exponents for luminance, contrast and structural terms, α_t, β_t and γ_t, respectively, in Eq. (3.46) are set to 1 by default.

3.4.2. Peak Signal-to-noise Ratio (PSNR)

PSNR is the ratio of the power of a signal to the power of noise that affects the signal's quality, where the noise is computed by comparing an image with perfect quality x, to another image $\hat{x}$. It is computed as

$$\text{PSNR}\left(\hat{x}, x\right) = 10log_{10}\left(\frac{D_R}{\Sigma_r\Sigma_c\left(\hat{x}[r,c] - x[r,c]\right)^2}\right), \tag{3.48}$$

where D_R is the dynamic range of x. For positive pixel values, the maximum value of x is also used, i.e., for image pixels with a resolution of 10 bits, the maximum value is 1024.

3.4.3. Contrast-to-noise Ratio (CNR)

CNR is the ratio of signal intensity differences between two regions to image noise. It is defined as:

$$\text{CNR} = \frac{|\mu_{\hat{x}_{i,j}} - \mu_{\hat{x}_{i',j'}}|}{\frac{\sigma_{\hat{x}_{i,j}} + \sigma_{\hat{x}_{i',j'}}}{2}} \tag{3.49}$$

where $\hat{x}_{i,j}$ and $\hat{x}_{i',j'}$ represent neighboring flat patches of the same frame with a noticeable variation in their mean gray values, $\mu_{\hat{x}_{i,j}}$ and $\mu_{\hat{x}_{i',j'}}$ represent the corresponding patch means, and $\sigma_{\hat{x}_{i,j}}$ and $\sigma_{\hat{x}_{i',j'}}$ represent the corresponding patch standard deviations. Multiple CNR values are determined for each frame and averaged to ensure reliability.

3.5. Summary

Denoising is a viable strategy to improve the image quality associated with noisy low-dose X-ray images. Denoising algorithms need to be designed by carefully taking into account the associated noise characteristics. Existing denoising approaches can be broadly categorized into analytical and learning-based. Well established techniques such as simple spatial domain filtering, variational approaches, patch- and transform-based methods and approaches that involve sparse and low-rank modeling belong to the category of analytical approaches. These approaches usually rely on solving complex optimization problems to find an optimal solution. Hence, they are time-consuming during the inference phase. In addition, they typically involve multiple parameters that need to be tuned to get the best results. Learning-based methods, on the other hand, do not suffer from such issues. In particular, CNNs are well suited for parallel computation on modern GPUs. Moreover, network architectures have become more flexible and efficient in performing the required tasks. Finally, commonly used metrics for assessing the image quality in denoised images, such as SSIM, PSNR and CNR, have been presented. In chapters 5 and 6, novel analytical and learning-based denoising strategies have been proposed and thoroughly evaluated. The techniques make use of the X-ray imaging model and the noise estimation approach presented in Chapters 2 and 4. For evaluating the proposed denoising approaches, some of the techniques presented in this chapter, such as [96, 97, 113, 176, 145], have been considered as benchmark.

Part II.

Contributions

4. Estimation of X-ray Imaging and Noise Model

4.1. Introduction

In X-ray images, noise has various sources, e.g., Poisson noise, electronic noise, and quantization noise. Depending on the imaging situation, different detector gains and beam energies may be selected. This may also change the noise characteristics present in an X-ray image. Therefore, it is important to take the characteristics of noise into account accurately (and adaptively) while designing denoising algorithms. In this chapter, a data-driven method for estimating the noise characteristics of X-ray images is presented. The proposed method is published in [1].

4.2. Data-driven Estimation of Noise Variance Stabilization Parameters for Low-dose X-ray Images

4.2.1. Paper Summary

Denoising of X-ray images corrupted by signal-dependent mixed noise can be performed after stabilizing the noise variance to a known constant as this facilitates the application of denoising algorithms designed for additive Gaussian noise. The generalized Anscombe transform (GAT) is a well performing noise variance stabilization (NVS) transform [183]. However, it requires knowledge of the overall system gain and the variance of electronic noise. They are, unfortunately, difficult to predict from the X-ray tube settings in clinical practice due to the presence of the patient changing the X-ray beam properties.

Common approaches to obtain the system gain and the additive Gaussian noise component involve extensive calibrations or they are based on data-driven noise estimation algorithms. The data-driven techniques proposed in [184, 185] estimate the signal-dependent noise in the frequency domain to obtain a noise level function (NLF) describing the relationship between noise variance and intensity levels. In [186, 187], a noise model, for raw data of digital sensors, comprising two mutually independent parts, a signal-dependent part and a signal-independent part, has been presented. In addition, a method for estimating them is presented. Even though the methods proposed in [184, 185, 187] yield good results in the case of optical images (digital photographs), they can be improved further when applied to X-ray images.

The contribution of the following article [1] is a computationally efficient method for estimating X-ray imaging parameters from X-ray images. The approach extends the NLF estimation proposed in [185]. It utilizes the energy compaction property of the discrete cosine transform (DCT) and is based on the following two properties of medical images:

First, low-frequency DCT coefficients predominantly contain image information (signal). Second, high-frequency DCT coefficients capture edges and noise. According to these properties, high-frequency DCT coefficients associated with flat patches, i.e. patches having a little texture, can be used to estimate noise. Based on this and the linear x-ray image formation model, the NLF associated with an X-ray image is estimated. Subsequently, the imaging parameters are computed from the NLF by a weighted model fitting scheme that minimizes the influence of outliers arising due to the absence of sufficient flat patches for some gray values.

The method has been experimentally validated with respect to beam hardening as well as denoising performance for different dose and scatter levels using X-ray images of an anthropomorphic phantom. The results have revealed that the method has robustly estimated the parameters for performing the GAT with an average error of only 4.2%. Performance gains of 5% for peak-signal-to-noise ratio and 4% for structural similarity index can be obtained, when the proposed method is used for performing the GAT followed by the application of denoising compared to existing methods designed for extracting the NLF.

Author's Final Manuscript

Data-driven Estimation of Noise Variance Stabilization Parameters for Low-dose X-ray Images

Sai Gokul Hariharan[1,2], Norbert Strobel[3], Christian Kaethner[2], Markus Kowarschik[1,2], Rebecca Fahrig[2,4], Nassir Navab[1,5]

[1]Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2]Siemens Healthineers AG, Advanced Therapies, Forchheim, Germany
[3]Institute of Medical Engineering Schweinfurt, University of Applied Sciences, Würzburg-Schweinfurt, Schweinfurt, Germany
[4]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[5]Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America

Abstract.
Purpose: Denoising X-ray images corrupted by signal-dependent mixed noise is usually approached either by considering noise statistics directly or by using noise variance stabilization (NVS) techniques. An advantage of the latter is that the noise variance can be stabilized to a known constant throughout the image, facilitating the application of denoising algorithms designed for the removal of additive Gaussian noise. A well-performing NVS is the generalized Anscombe transform (GAT). To calculate the GAT, the system gain as well as the variance of electronic noise are required. Unfortunately, these parameters are difficult to predict from the X-ray tube settings in clinical practice, because the system gain observed at the detector depends on the beam hardening caused by the patient.
Material and Methods: We propose a data-driven method for estimating the parameters required to carry out an NVS using the GAT. It utilizes the energy compaction property of the discrete cosine transform to obtain the NVS parameters using a robust regression approach relying on a linear Poisson-Gaussian model. The method has been experimentally validated with respect to beam hardening as well as denoising performance for different dose and scatter levels.
Results: Across a range of low-dose X-ray settings, the proposed robust regression approach has estimated both system gain and electronic noise level with an average error of only 4.2%. When used to perform a GAT followed by the denoising of low-dose X-ray images, performance gains of 5% for peak-signal-to-noise ratio and 4% for structural similarity index can be obtained.
Conclusion: The parameters needed to calculate the GAT can be estimated efficiently and robustly using a data-driven approach. The improved parameter estimation method facilitates a more accurate GAT-based NVS and, hence, better denoising of low-dose X-ray images when algorithms designed for additive Gaussian noise are applied.

1. Introduction

X-ray fluoroscopy is an important imaging technique for navigating endovascular tools during an intervention. Although X-ray imaging has distinct advantages, it may also expose the patients as well as the clinical staff to a non-negligible amount of ionizing radiation. To keep potential consequences to a minimum, the X-ray dose has to be kept as low as reasonably achievable (ALARA principle). Unfortunately, lowering the dose also reduces the number of photons that are available for image generation. This results in a reduced signal-to-noise-ratio (SNR). Depending on the imaging system and the clinical situation, X-ray images can be corrupted by different amounts and types of noise, in particular, signal-dependent quantum noise and signal-independent electronic noise. A possible solution to retain image quality in the presence of noise is to apply denoising techniques that are designed by carefully taking into account the associated noise characteristics. This has been done either by directly considering the statistics of the particular noise model [1, 2] or by applying noise variance stabilization (NVS) techniques [3] together with algorithms designed for the removal of additive Gaussian noise [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. An interesting example combining NVS with the popular denoising approach namely block matching 3D (BM3D) [14], is the iterative algorithm designed for low SNR scenarios proposed by Azzari et al. [15].

If the signal-independent noise component of X-ray images can be safely ignored, a simple square root transform or the conventional Anscombe transform can be applied to perform an NVS [16]. However, at very low dose levels, neglecting signal independent noise and applying traditional NVS techniques may no longer be sufficient. A possible solution to this is the generalized Anscombe transform (GAT) [3] for mixed noise scenarios [7, 8]. To apply the GAT, the system gain and the additive Gaussian noise component need to be known. These parameters are, unfortunately, difficult to predict from the (known) tube settings due to beam hardening caused by the patient's body.

Common approaches to obtain the system gain and the additive Gaussian noise component involve extensive calibrations or they are based on data-driven noise estimation algorithms. The data driven algorithms can be broadly classified into four groups: (i) filtering of noisy images to separate the noise from the actual signal [17, 18, 19, 20, 21, 22] and an exclusion of high gradients to prevent their influence on the noise estimation [23], (ii) wavelet-based approaches that assume that the noise and the signal can be separated into different sub-bands [24], (iii) techniques such as a principal component analysis or a singular value decomposition operating on homogeneous patches to obtain noise estimates [25, 26, 27, 28, 29] and (iv) frequency domain techniques including the discrete cosine transform (DCT) for estimating the noise [30, 31, 32].

Ponomarenko et al. [31] have proposed a method for estimating the signal-dependent noise in the frequency domain to obtain a noise level function (NLF) that gives the association between noise variance and intensity levels. The method is based on iteratively identifying patches with low texture (flat patches) from low-frequency DCT

coefficients and subsequently estimating the noise level from associated high-frequency DCT coefficients. This approach has been modified by Colom et al. [32]. They have replaced the iterative noise estimation by selecting a fixed number of patches whose low-frequency DCT coefficients have low energy, consequently selecting flat patches effectively.

Similarly, Foi et al. [24] and Pyatykh et al. [33] have presented a noise model for raw data of digital sensors and an algorithm for estimating noise from a single noisy image. They have assumed that the noise is composed of two mutually independent parts, a signal-dependent part and a signal-independent part. Pyatykh et al. [33] have proposed an optimization procedure based on principal component analysis to find the parameters for performing an NVS that will transform the noisy image into an image with signal-independent noise.

Even though the methods proposed in [31, 32, 33] yield good results in the case of optical images, they can be improved further when applied to X-ray images. This can be done by utilizing the fact that a flat-panel X-ray detector's response is designed to be strictly linear with respect to the incoming dose [34, 35, 36].

The main contribution of this paper is a feasible and practical approach for estimating the parameters needed to compute the GAT-based NVS in the context of X-ray imaging. The approach extends the NLF estimation proposed in [32]. While estimating the NLF, it is essential to avoid patches containing texture since the texture in such patches may be mistaken to be noise. If this is not possible, at least their impact needs to be reduced as much as possible. As a solution to this problem, we propose the use of weights when fitting a linear model. They ensure that noise is estimated predominantly from flat patches. This prevents patches with texture from influencing the noise estimation process hence making the overall NVS more stable. In order to apply the GAT, the pixel-wise system gain and the additive Gaussian noise component need to be known. Since estimating the gain for each pixel is not feasible in practice, e.g., due to beam hardening and patient motion, we have experimentally investigated whether it would be sufficient and acceptable to use an average system gain in the case of low-dose X-ray imaging instead of pixel-wise system gain. A thorough evaluation using X-ray images from two different imaging systems has shown that the proposed GAT-based approach works robustly and outperforms the currently used NVS techniques. Although the focus of this paper is not on denoising, we have used the proposed approach as a pre-processing step for the well-known BM3D denoising technique to illustrate its potential for performance gains. The positive results that we have obtained with BM3D suggest that other denoising techniques designed for the removal of additive Gaussian noise, e.g., the ones mentioned in [37], could benefit from the proposed GAT-based NVS approach as well.

2. Material and Methods

In this section, a computationally efficient method for estimating X-ray imaging parameters from X-ray images based on a linear imaging model, is proposed. The estimated parameters can be used to perform an NVS using the GAT [3].

2.1. X-ray Imaging System Model

The behavior of an indirect-detection flat-panel X-ray detector can be modelled using a cascaded linear systems approach [34]. According to this model, the formation of fluoroscopic X-ray images from X-ray photons can be described in a series of stages where each stage represents a physical process having intrinsic gain or spreading properties [34]. Assuming that the flat-panel detector has a linear detector response, the average output signal $\bar{z}$ can be written as

$$\bar{z} = \bar{g} \cdot \bar{x} + g_o, \tag{1}$$

where $\bar{g}$ is the overall system gain, $\bar{x}$ is the mean fluence of Poisson-distributed incident X-ray quanta and g_o is the overall offset. When the detector is used for acquiring image sequences, the variance of noise associated with the exposure at different detector pixel positions can be expressed as

$$\sigma^2_{\bar{z}}[r,c] = \alpha[r,c]\left(\bar{z}[r,c] - g_o\right) + \sigma^2_n, \tag{2}$$

where $\alpha[r,c]$ is the pixel-dependent gain associated with the pixel located at $[r,c]$ [34, 38]. The parameter σ^2_n refers to the variance of the total additive noise due to different sources, such as intrinsic noise from the photo-diode pixel, voltage fluctuations on gate and bias lines, noise from the amplifier and digitization noise [34]. For simplicity, we have assumed that the electronic noise is independent of location but could be extended to position-dependent electronic noise. In general, the gain is spatially varying because it depends on the energy of the X-ray beam reaching the associated detector elements [39, 35]. Equation (2) represents the NLF. It can be seen as the variance of a Poisson variable with mean $\bar{z}[r,c]$ scaled by $\alpha[r,c] > 0$ and corrupted by additive Gaussian noise $\rho[r,c]$ and also written as follows:

$$\sigma^2_z[r,c] = \alpha[r,c] \cdot \bar{z}[r,c] + \rho[r,c] \quad \text{with} \quad \rho[r,c] = \sigma^2_n - \alpha[r,c] \cdot g_o. \tag{3}$$

In (3), $\alpha[r,c]$ and $\rho[r,c]$ are the slope and the y-intercept of the NLF, respectively.

The imaging model presented in (3) has been experimentally verified using an image quality phantom comprising a Copper step wedge placed on a 40 mm Aluminum block (for more details see Appendix A). The results of the experiment have shown that the gain $\alpha[r,c]$ depends on the spectrum of the X-ray photons collected at the detector [40]. This is expected, since the sensitivity of the X-ray detector is energy-dependent [35].

2.2. Noise Variance Stabilization and Denoising

As mentioned above, the GAT can be used as an NVS transform to stabilize mixed Poisson-Gaussian noise to a constant [10, 3]. The forward transform is given by

$$z'[r,c] = f(z[r,c], \alpha[r,c], \rho[r,c]) = \frac{2}{\alpha[r,c]}\sqrt{\alpha[r,c]\cdot z[r,c] + \frac{3}{8}\alpha[r,c]^2 + \rho[r,c]}. \quad (4)$$

After denoising an image in the GAT domain, either an algebraic or an unbiased inverse GAT [10] can be used to return to the original image domain.

It can be inferred from (3) and (4) that the pixel-wise gain $\alpha[r,c]$ and the intercept $\rho[r,c]$ need to be known to perform the forward GAT. However, in a practical scenario, e.g., during cardiac catheterization, it is difficult to compute pixel-wise parameters accurately, e.g., due to the presence of signal-dependent noise. Therefore, we have investigated based on realistic imaging conditions, if it is feasible to use an average gain instead of pixel-wise gains (for more details see Appendix B). We have found that the benefit of using an accurate $\alpha[r,c]$ is marginal compared to using an average estimate $\bar{\alpha}$, derived by averaging all the pixel-wise estimates $\alpha[r,c]$. Since accurate estimates of $\alpha[r,c]$ are difficult to compute in practice, using a somewhat less accurate gain for each pixel instead can be considered acceptable.

2.3. Method for Estimating the NVS Parameters

The proposed method involves the following steps: First, patches of an X-ray image are grouped into bins based on their mean gray values. Second, a bin-wise estimation of the noise variance, taken from high frequencies, and texture variance based on low frequencies, is performed from the patches within the bin. Third, bin-wise weights are computed from the texture variance and the number of elements associated with the respective bins. Finally, based on the weights, a model is fit to the noise variances resulting in the parameters needed to carry out the GAT. The weights reduce the influence of unreliable bins on the model fitting process. Unreliable bins are those containing either an inadequate number of patches or patches with texture. The different parts of the method are described in detail in the subsequent sections and the general workflow is illustrated in figure 1.

2.3.1. Estimation of the NLF As a first step, the noisy image $\boldsymbol{z} = \{z[r,c]\} \in \mathbb{R}^{M\times N}$ is decomposed into overlapping square patches $\boldsymbol{z_{i,j}}$, of width W, extracted at every possible pixel location i, j. The patches are then transformed using the DCT. In the DCT domain, $\zeta_{i,j}[u,v]$ represents the value of the DCT coefficient at row u and column v in the matrix $\boldsymbol{\zeta_{i,j}}$ associated with the DCT-transformed image patch $\boldsymbol{z_{i,j}}$. Next, the patches are grouped into $b = 1, \ldots, B$ bins based on the mean of the patches μ as found in the zero-frequency DCT coefficient. For every bin b, an estimate of the noise variance σ_H^2 is computed. This technique is described in more detail in the following paragraphs.

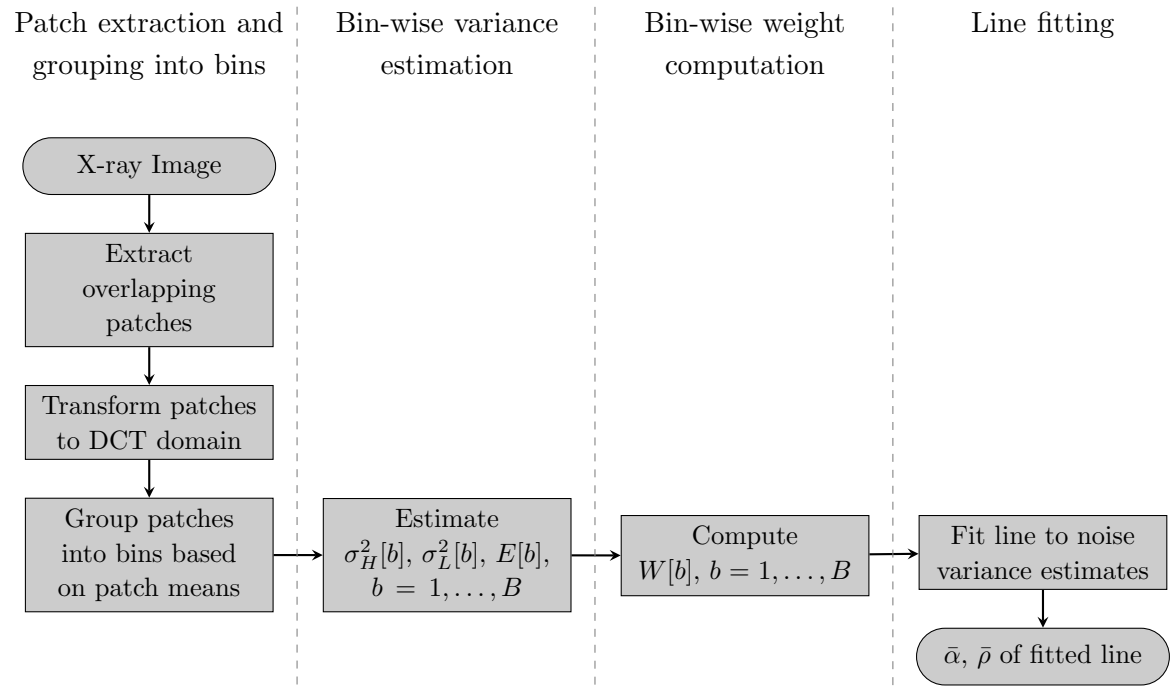


Figure 1: Visualization of the schematic workflow for the proposed parameter estimation approach to get the slope $\bar{\alpha}$ and the intercept $\bar{\rho}$ of the fitted line. For $b = 1, \ldots, B$ bins, $\sigma_H^2[b]$ is the variance due to noise, $\sigma_L^2[b]$ is the variance due to texture, $E[b]$ is the number of elements associated with bin b, and $W[b]$ is the estimated weight.

Due to the energy compaction property of the DCT, the DC (at zero frequency) coefficient and its low-frequency neighbors usually capture most of the energy as they are associated with low-frequency information and this phenomenon is typically dominant in natural and medical images. The higher-frequency DCT coefficients mostly represent noise and remaining edge information. To get the noise variance estimates, high-frequency DCT coefficients of patches with little texture (flat patches) are preferred and we give more importance to bins with a high number of flat patches using weights. Such flat patches are characterized by low texture variance, i.e., a low energy across low-frequency DCT coefficients associated with a patch.

High-frequency coefficients can be differentiated from low-frequency coefficients by choosing transform coefficients based on their position $[u, v]$ in the DCT domain. This can be achieved using the parameter τ as per

$$\delta[u,v] = \begin{cases} 1, & \text{if } (u+v < \tau) \cap (u+v \neq 0)\,, \\ 0, & \text{if } (u+v \geq \tau) \cap (u+v \neq 0)\,, \\ \text{-1}, & \text{if } (u=v=0)\,. \end{cases} \tag{5}$$

If $\delta[u, v] = 0$, the position has been taken as a high-frequency coefficient, and if $\delta[u, v] = 1$, it has been taken as a low-frequency coefficient. The DC coefficient has not been taken into consideration. To keep the math consistent and compact, $\delta[0, 0]$ has

been assigned the value -1.

The empirical low-frequency patch variances $\nu_{i,j}$ have been derived from the low-frequency coefficients of $\boldsymbol{\zeta_{i,j}}$ as follows:

$$\nu_{i,j} = \frac{1}{\theta} \sum_{u=0}^{W-1} \sum_{v=0}^{W-1} \zeta_{i,j}[u,v]^2 \times \max\left\{\delta[u,v], 0\right\} \text{ with } \theta = \sum_{u=0}^{W-1} \sum_{v=0}^{W-1} \max\left\{\delta[u,v], 0\right\}. \quad (6)$$

Since the low-frequency DCT coefficients (except the DC coefficient) of flat patches are small, the patches with a low value for $\nu_{i,j}$ are likely to feature low texture. To identify low-texture patches, we have proceeded as follows: For each (gray value) bin b with mean $\mu[b]$, we have sorted the DCT transformed patches $\boldsymbol{\zeta_{i,j}}$ in ascending order with respect to $\nu_{i,j}$. The sorted patches can be viewed as vectors $\boldsymbol{\zeta_{b,p}}$, where p denotes the index of a patch in bin b. Colom et al. [32] have suggested that using the first 210 sorted patches is sufficient to get a robust estimate for the noise variance $\sigma_H^2[b]$ associated with a bin b. They have computed the estimates of the noise variance from each of the high-frequency DCT coefficients and subsequently computed the median of the estimates to obtain the final noise variance [32]. However, for X-ray images, we have found that using the median of the estimates underestimates the amount of noise. To arrive at a more accurate result, we have modified their method and computed the bin-wise final noise variance estimate as the mean of all high-frequency coefficients $\sigma_H^2[b]$ as follows:

$$\sigma_H^2[b] = \frac{1}{\gamma} \sum_{p=0}^{P-1} \sum_{u=0}^{W-1} \sum_{v=0}^{W-1} \zeta_{b,p}[u,v]^2 (1 - |\delta[u,v]|) \text{ with } \gamma = P \sum_{u=0}^{W-1} \sum_{v=0}^{W-1} (1 - |\delta[u,v]|), \quad (7)$$

where γ is a normalization factor and P is the number of patches used for computing the noise variance for each bin b. Similarly, we have computed the bin-wise texture variance $\sigma_L^2[b]$ as per

$$\sigma_L^2[b] = \frac{1}{\psi} \sum_{p=0}^{P-1} \sum_{u=0}^{W-1} \sum_{v=0}^{W-1} \zeta_{b,p}[u,v]^2 \times \max\left\{\delta[u,v], 0\right\} \text{ with } \psi = P \times \theta. \quad (8)$$

2.3.2. Estimation of Imaging Parameters for NVS After calculating the variances, each bin b has the associated mean $\mu[b]$, an estimate for the noise variance $\sigma_H^2[b]$, an estimate for texture variance $\sigma_L^2[b]$ and the number of elements in the bin $E[b]$. The pairs ($\mu[b]$, $\sigma_H^2[b]$) can be used to estimate the NLF [32]. According to our noise model presented in (2), fitting a line with slope $\bar{\alpha}$ and intercept $\bar{\rho}$ to the estimated noise variances is a logical choice. Furthermore, to reduce the influence of outliers during the fitting process, weights can be used. This line fitting process can be represented as

$$\bar{\alpha}, \bar{\rho} = \arg\min_{\alpha,\rho} \sum_b W[b] \left(\sigma_H^2[b] - \alpha \times \mu[b] - \rho\right), \quad (9)$$

where the weights for the data elements have been set to $W[b] = E[b]/\sigma_L^2[b]$. The use of weights is necessary as certain (gray value) bins may not have enough patches or

the patches may not be flat enough to give a good estimate. Thanks to the weights, such bins are given lesser importance. On the other hand, if a bin has many elements and a low variance associated with texture, it is considered more reliable and a higher weight is assigned. Thus, the higher the weight, the higher is the weighted distance. To minimize the weighted distance, the point has to be close to the fitted curve.

In order to solve the optimization problem presented in (9) and compute the model parameters, we have made use of a weighted line fitting approach based on the random sample consensus algorithm [41]. According to this approach, the data elements ($\mu[b]$, $\sigma_H^2[b]$) vote for a model depending on their weighted distance (standard deviation) from the model. Hence $\bar{\alpha}$ and $\bar{\rho}$ are obtained from the point pairs with minimum weighted distance, i.e., from the most reliable point pairs.

2.4. Material

To evaluate the proposed method, we have acquired data from two X-ray systems with different detector technologies (Siemens Healthineers AG, Erlangen, Germany) denoted as System-1 and System-2, respectively. System-1 is a general-purpose angiography system, whereas System-2 has been optimized for low-dose X-ray applications. The acquired data sets comprise raw X-ray images (after detector correction) of a thorax phantom with a thickness of 20 cm. The X-ray images have been taken at 25%, 50% and 100% of the standard low-dose level (SD). To generate a realistic amount of scatter, plexiglass (polymethyl methacrylate, PMMA) of different thickness ranging from 2 cm to 16 cm has been used.

The proposed method follows Ponomarenko et al. [31, 32], but performs a robust weighted estimate of the NLF based on an X-ray imaging system model ("weighted Ponomarenko method", WPM). For comparison, the unweighted linear fit approach ("unweighted Ponomarenko method", UPM) has also been presented. In addition, the results obtained using the method by Pyatykh et al. [33] has been used as a benchmark to compare our results.

As described in section 2.3, the proposed method requires the image to be divided into patches, which are then grouped into bins associated with gray values $\mu[b]$. Since larger patches will increase the robustness of their mean gray value estimates in homogeneous regions, we have set the width of the square patches to $W = 30$ pixels corresponding to a detector area of 4.8 mm $\times$ 4.8 mm. In order to take into account the influence of noise on the mean gray values of the patches – while allowing for small deviations introduced due to the random nature of noise – we have set the bin width to $N = 2$ gray values. Then, by analyzing the NLFs and the fitted noise models obtained for images acquired at different dose levels (25% SD, 50% SD and 100% SD) and different scatter levels (2 cm, 4 cm, 8 cm and 16 cm PMMA) from two different detectors, we have ascertained that choosing at most $P = 210$ patches from each of the B bins – as suggested by Colom et al. [32] – is also suitable to estimate robustly and efficiently the required parameters in the case of low-dose X-ray images. To distinguish low-frequency

DCT coefficients from high-frequency DCT coefficients (as described in (5)), we have empirically determined $\tau = 7$ as a good choice.

To evaluate the different NVS approaches, the ground-truth (GT) estimates of the imaging parameters are required. The GT estimates can be computed using pairs of noisy and noise-free images. To this end, we have acquired 500 raw X-ray images of a static object (thorax phantom) and averaged them pixel-wise in the temporal direction to remove signal-dependent as well as signal independent noise. Subsequently, we have computed the average noise variance associated with every gray value from the corresponding pixels in the GT and the noisy images as follows: First, we have identified the pixel positions in the GT associated with a particular gray value. Then, we have computed the noise variance from the gray values at these pixel positions in the noisy image. Based on the noise model presented in (3), a line has been fitted to the noise estimates to obtain the parameters $\bar{\alpha}$ and $\bar{\rho}$ for the NVS. As an example, we show in figure 2 X-ray images of the thorax phantom acquired at 25% SD, the corresponding GT and the corresponding NLF as well as the fitted model. In fact, our experiments have shown that the imaging model presented in (3) is accurate for low-dose X-ray images.

To provide a practical example showing the merits of our approach, we have presented the results of denoising using the well-established BM3D [14] after performing an NVS based on the UPM, the WPM and the method by Pyatykh et al. [33]. We have also compared the results with those of Azzari et al. [15], an iterative denoising method that combines NVS and BM3D – as we consider the method as a benchmark. The denoised images have been evaluated with respect to peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

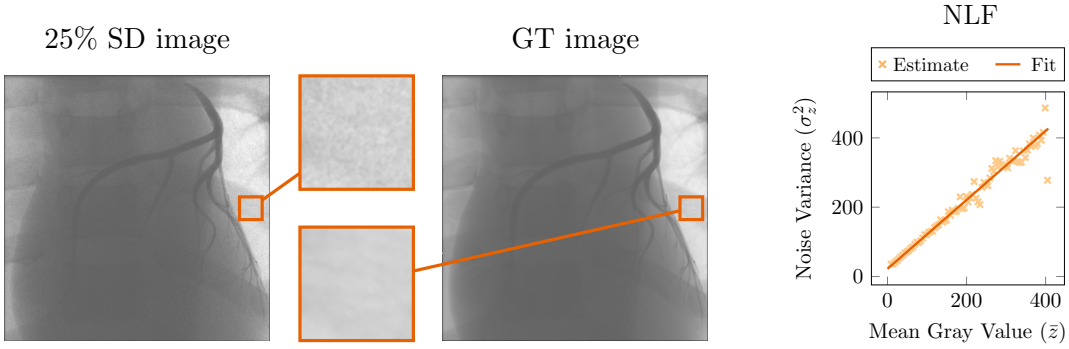


Figure 2: Visualization of an X-ray image of a thorax phantom acquired at 25% standard dose (SD), the corresponding ground truth (GT) image (average of 500 spatially aligned instances of the low-dose image that have been subsequently acquired), and the noise level function (NLF) associated with the 25% SD X-ray image. The mean gray values $\bar{z}$ have been sampled from the GT image and the noise variances have been computed by comparing the corresponding pixels in the GT image with those of the 25% SD image.

3. Results

In figure 3, we show the noise variance estimates σ_H^2, the texture variances σ_L^2 and the number of elements E for each bin as well as the weights W assigned to the different bins corresponding to different mean gray values μ. These components are used for estimating the parameters of the underlying X-ray imaging system model. It can be observed that the weights are high when the corresponding texture variances are low and the number of elements present in the associated bin is high. Figure 4 is a visualization of the fitted lines for the noise variance estimates obtained using the method by Pyatykh et al. [33], the UPM and the WPM together with the GT. It can be seen that the method by Pyatykh et al. [33] has resulted in inaccurate estimates for $\bar{\alpha}$ and accurate estimates for $\bar{\rho}$. The UPM has yielded better estimates but has been affected by outliers. In comparison, the WPM, that involves weights, has resulted in accurate estimates even in the presence of outliers.

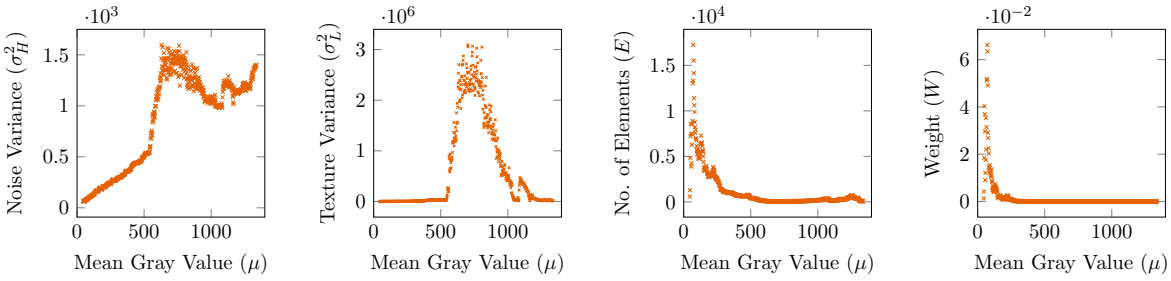


Figure 3: Illustration of the noise variance estimates σ_H^2, the texture variances σ_L^2 and the number of elements E for each bin associated with different mean gray values μ. The weights W computed from σ_L^2 and E are also presented.

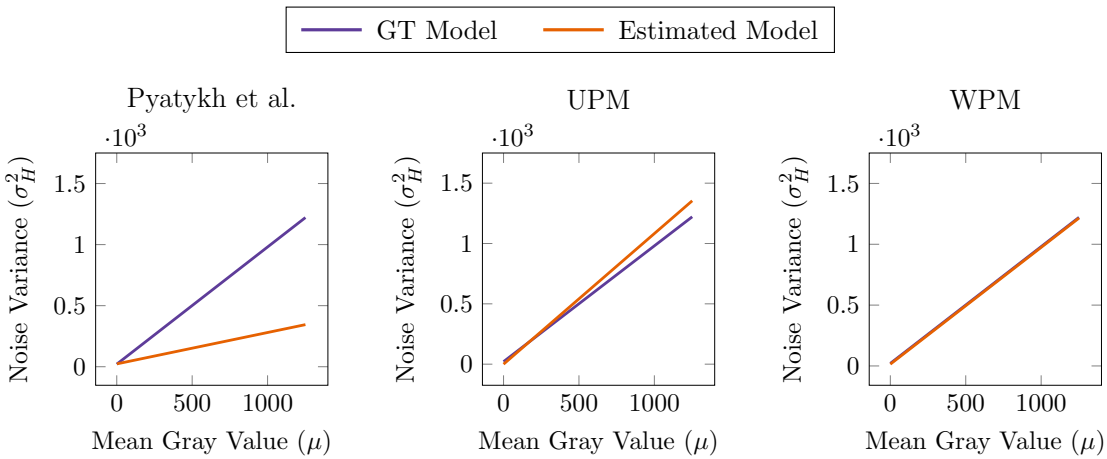


Figure 4: Visualization of the fitted models for the ground truth (GT) and the estimated noise level functions based on the method by Pyatykh et al. [33], the unweighted Ponomarenko method (UPM) and the weighted Ponomarenko method (WPM).

To highlight these findings, we display in figure 5 the GT imaging parameters as well as the estimated imaging model parameters computed from 50 images using the method by Pyatykh et al. [33], the UPM and the WPM for different dose and scatter levels in the cases of System-1 and System-2. From collimated regions of the X-ray images, we have estimated the system offset g_o to be 0. Therefore, the slope and the intercept of the fitted curve are given by $\bar{\alpha}$ and σ_n^2, respectively. It can be seen that the method by Pyatykh et al. [33] has consistently underestimated $\bar{\alpha}$ and the UPM has yielded an inaccurate estimation of σ_n in a few cases. The WPM, however, has provided an accurate estimation of both $\bar{\alpha}$ and σ_n throughout the investigated cases.

A quantitative analysis of the results has shown that the WPM has accurately estimated $\bar{\alpha}$ with an average error of $2.98\% \pm 3.93\%$ for System-1 and $5.45\% \pm 2.30\%$ for System-2. The method by Pyatykh et al. [33] (average error of $127.24\% \pm 21.74\%$ for System-1 and $282.11\% \pm 36.97\%$ for System-2) and the UPM (average error of $7.35\% \pm 8.03\%$ for System-1 and $8.43\% \pm 7.95\%$ for System-2), however, have yielded less reliable estimates. For the estimation of σ_n, the WPM has provided the most accurate results (average error of $3.97\% \pm 3.39\%$ for System-1 and $4.46\% \pm 5.74\%$ for System-2), followed by the method by Pyatykh et al. [33] (average error of $4.18\% \pm 17.12\%$ for System-1 and $6.58\% \pm 4.54\%$ for System-2) and the UPM (average error of $11.84\% \pm 18.06\%$ for System-1 and $28.56\% \pm 41.96\%$ for System-2).

In figure 6, we show denoising results for X-ray images representing an average patient (thorax phantom placed on 2 cm PMMA and acquired at 100% SD) obtained using the method proposed by Azzari et al. [15] and BM3D [14]. Prior to applying BM3D, an NVS has been applied based on the parameter estimates for $\bar{\alpha}$ and $\bar{\rho}$ computed as per the method by Pyatykh et al. [33], the UPM and the WPM. Overall, the results obtained using BM3D with the WPM and BM3D with the UPM have been visually superior compared to BM3D with the method by Pyatykh et al. [33] and the method by Azzari et al. [15]. The method proposed by Azzari et al. [15] has resulted in noise that is most visible in the case of System-1 and somewhat lesser noise in the case of System-2. In the case of BM3D with the method by Pyatykh et al. [33], it can be seen that there is significantly higher noise compared to BM3D with the UPM and BM3D with the WPM. By performing a quantitative analysis of the results with respect to PSNR and SSIM, we have found that, among the compared approaches, the highest improvement over the input has been achieved using BM3D where the NVS is based on the WPM (25.75% for PSNR and 30.77% for SSIM) followed by BM3D with the UPM (25.22% for PSNR and 30.33% for SSIM), BM3D with the method by Pyatykh et al. [33] (20.37% for PSNR and 27.39% for SSIM) and the method by Azzari et al. [15] (19.46% for PSNR and 24.33% for SSIM). Finally, in the case of BM3D where NVS parameters have been derived using WPM, there are only minor differences compared to the results obtained for BM3D with the GT NVS parameters (0.18% for PSNR and 0.07% for SSIM). These differences are higher for BM3D with the UPM (0.56% for PSNR and 0.26% for SSIM), BM3D with the method by Pyatykh et al. [33] (4.46% for PSNR and 1.68% for SSIM) and the method by Azzari et al. [15] (4.87% for PSNR and 3.49% for SSIM). For more

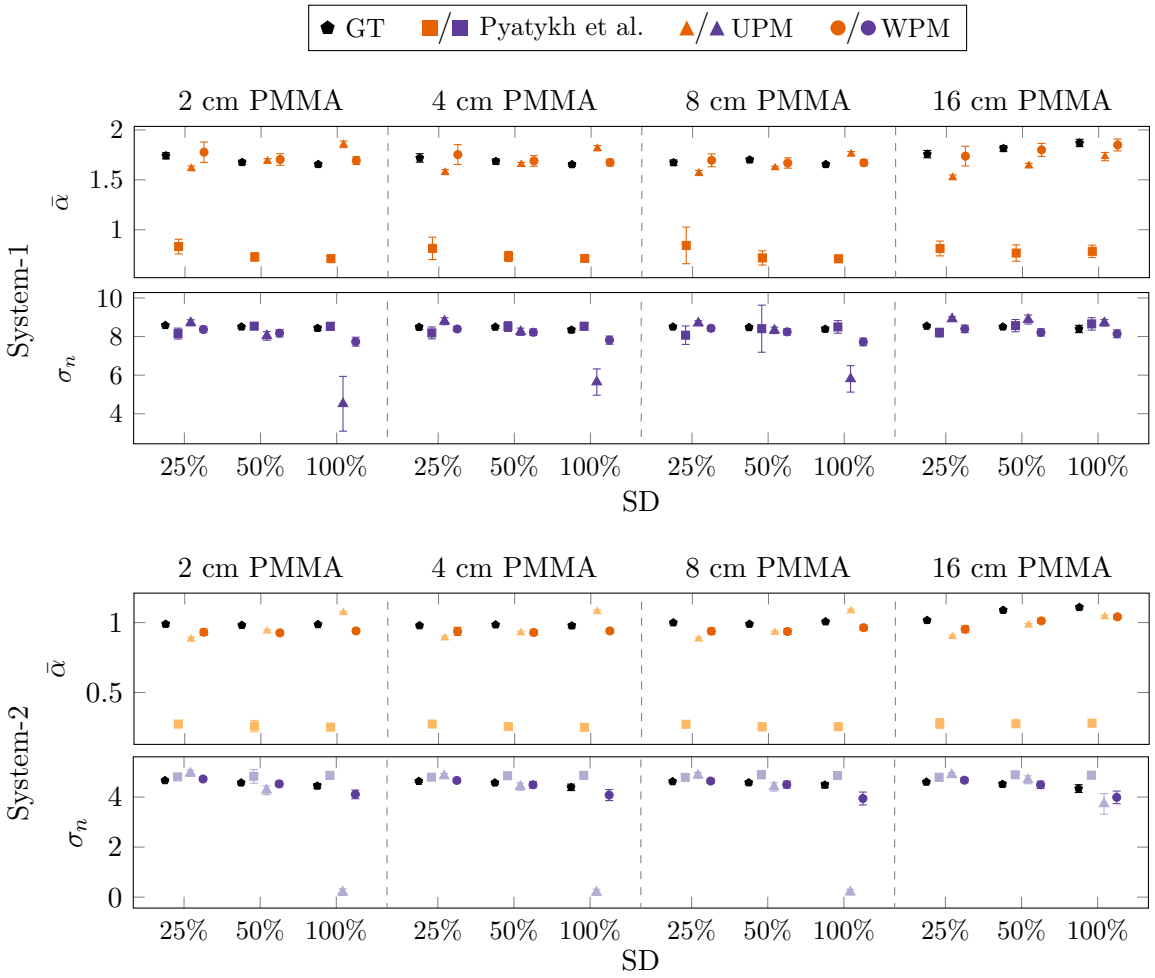


Figure 5: Visualization of the ground truth (GT) as well as the estimated imaging model parameters $\bar{\alpha}$ and σ_n based on the method by Pyatykh et al. [33], the unweighted Ponomarenko method (UPM) and the weighted Ponomarenko method (WPM) for X-ray images acquired using System-1 (a general-purpose angiography system) and System-2 (an angiography system optimized for low-dose X-ray applications). The X-ray images have been acquired at different dose levels (25% standard dose (SD), 50% SD and 100% SD). Additional scatter has been generated using 2 cm, 4 cm, 8 cm and 16 cm of polymethyl methacrylate (PMMA). The plots show the mean and the standard deviation of the estimates computed from 50 measurements.

details regarding the quantitative analysis, see Appendix C.

4. Discussion and Conclusion

Although denoising algorithms designed specifically for signal-dependent noise have been proposed [1, 2], the removal of such noise using algorithms developed for additive Gaussian noise, often involves an NVS [10, 9, 6] in practice. To apply a GAT-based NVS and stabilize the noise variance to a known constant, the underlying imaging

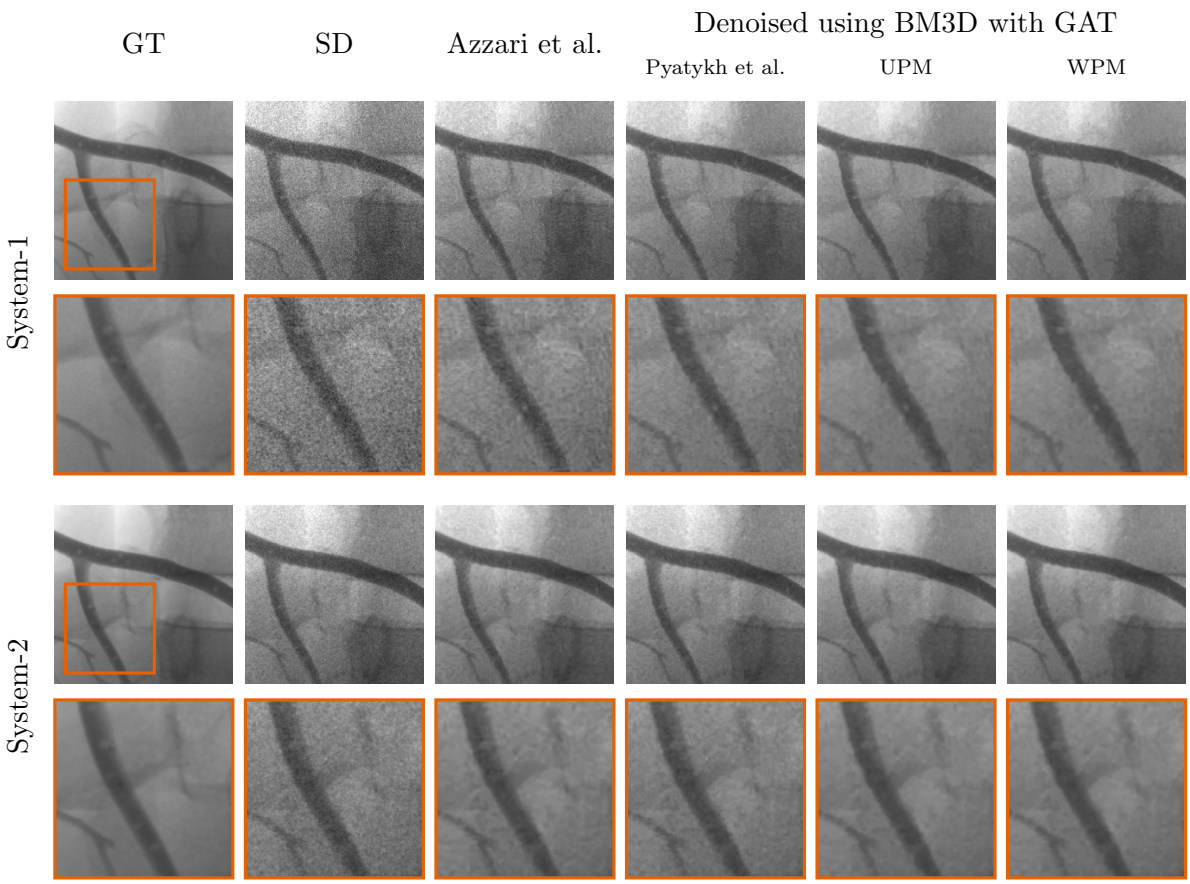


Figure 6: Visualization of X-ray images of the anthropomorphic thorax phantom placed on 2 cm polymethyl methacrylate acquired using System-1 (a general-purpose angiography system) and System-2 (an angiography system optimized for low-dose X-ray applications). Ground truth (GT) images (average of 500 spatially aligned instances of a standard-dose (SD) image that have been subsequently acquired) are followed by SD X-ray images together with the results of Azzari et al. [15] and block matching 3D (BM3D) [14]. Denoising using BM3D has been performed after applying the generalized Anscombe transform (GAT) on the noisy images based on the parameters obtained using the method by Pyatykh et al. [33], the unweighted Ponomarenko method (UPM) and the weighted Ponomarenko method (WPM).

parameters, i.e., the system gain and the electronic noise, are required. Unfortunately, it is difficult to derive the system gain directly from the system settings, because the X-ray spectrum received at the detector changes depending on the patient and the X-ray imaging settings. It is also not feasible to estimate accurately pixel-wise gains $\alpha[r, c]$ during a clinical intervention as it is in general not possible to acquire identical images that differ only in dose to estimate pixel-wise NLFs – due to patient motion and ethical reasons. Fortunately, in the context of image denoising, we could show that it is still possible to achieve an improved performance by resorting to an average gain $\bar{\alpha}$ associated with a single X-ray image when performing an NVS.

To sum up, we have developed a data-driven method by modifying and enhancing the approaches presented in [31, 32]. In particular, we have made use of a linear X-ray image acquisition model to calculate the imaging parameters directly from the acquired images. This is particularly relevant when the X-ray imaging parameters change, e.g., due to varying view directions during an X-ray guided minimally invasive catheterization procedure. The proposed approach utilizes the following two properties of natural and medical images: First, low-frequency DCT coefficients predominantly contain image information. Second, high-frequency DCT coefficients capture edges and noise. Hence, high-frequency DCT coefficients associated with flat patches, i.e., patches having little texture, can be used to estimate noise. Based on this and the linear X-ray image formation model, we can estimate an NLF by fitting a line ("unweighted Ponomarenko method", UPM). To obtain even more accurate results, we have developed a weighted model fitting scheme ("weighted Ponomarenko method", WPM) that minimizes the influence of outliers. Once the imaging parameters have been estimated from the NLF, they can be used to perform an NVS using the GAT.

Using the WPM, we have been able to estimate the system gain and the electronic noise with an average error of 2.98% and 3.97% in the case of a general purpose angiography system (System-1) and 5.45% and 4.46% in the case of a dedicated low-dose angiography system (System-2). These results are based on averaging associated deviations from the GT across all dose levels and PMMA thicknesses, respectively. Since the impact of quantum noise is usually much higher than that of electronic noise, an accurate estimation of the system gain is generally more important for performing an NVS. On the contrary, when UPM is used, the combined average error in the cases of System-1 and System-2 has been significantly higher (7.89% for system gain and 20.20% for electronic noise) as the fitted models are impacted by outliers. The method proposed by Pyatykh et al. [33] has also resulted in a higher average error in the case of System-1 and System-2 (204.67% for system gain and 5.38% for electronic noise). A possible explanation for the high average error is that the method does not exploit the model associated with an X-ray imaging system.

Comparing the results of denoising after applying an NVS based on the different parameter estimation methods, we have found that an accurate estimation of the NVS parameters, especially $\bar{\alpha}$, has a significant impact on the denoising performance. Specifically, the WPM that has resulted in the most accurate estimation of the NVS parameters has also yielded the best denoising results. Based on these results, we conclude that the use of an accurate NVS before applying denoising algorithms designed for the removal of additive Gaussian noise with a known variance is a very viable strategy due to its efficiency and robustness. Finally, it would be interesting to see whether advanced methods such as [42, 43] could also benefit from the proposed approach when used in the context of X-ray imaging. However, this needs more in-depth investigations and could be considered for future research.

Acknowledgement and Disclaimer: This work was supported by Siemens Healthineers AG. The concepts and results presented in this paper are based on research and are not commercially available.

References

[1] Florian Luisier, Cédric Vonesch, Thierry Blu, and Michael Unser. Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Processing*, 90(2):415–427, 2010.

[2] Florian Luisier, Thierry Blu, and Michael Unser. Image denoising in mixed Poisson–Gaussian noise. *IEEE Trans Image Process*, 20(3):696–708, 2011.

[3] Jean-Luc Starck, Fionn D Murtagh, and Albert Bijaoui. *Image processing and data analysis: the multiscale approach*. Cambridge University Press, 1998.

[4] Lucas R Borges, Lucio Azzari, Predrag R Bakic, Andrew DA Maidment, Marcelo AC Vieira, and Alessandro Foi. Restoration of low-dose digital breast tomosynthesis. *Meas Sci Technol*, 29(6):064003, 2018.

[5] Wenzhao Zhao, Qiegen Liu, Yisong Lv, and Binjie Qin. Texture variation adaptive image denoising with nonlocal PCA. *IEEE Trans Image Process*, 28(11):5537–5551, 2019.

[6] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017.

[7] Sai Gokul Hariharan, Norbert Strobel, Christian Kaethner, Markus Kowarschik, Stefanie Demirci, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. A photon recycling approach to the denoising of ultra-low dose x-ray sequences. *Int J Comput Assist Radiol Surg*, 13(6):847–854, 2018.

[8] Sai Gokul Hariharan, Christian Kaethner, Norbert Strobel, Markus Kowarschik, Julie DiNitto, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. Preliminary results of DSA denoising based on a weighted low-rank approach using an advanced neurovascular replication system. *Int J Comput Assist Radiol Surg*, pages 1–10, 2019.

[9] Sai Gokul Hariharan, Christian Kaethner, Norbert Strobel, Markus Kowarschik, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. Learning-based x-ray image denoising utilizing model-based image simulations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 549–557. Springer, 2019.

[10] Markku Makitalo and Alessandro Foi. Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *IEEE Trans Image Process*, 22(1):91–103, 2013.

[11] Minghui Zhang, Fengqin Zhang, Qiegen Liu, and Shanshan Wang. VST-Net: Variance-stabilizing transformation inspired network for Poisson denoising. *J Vis Commun Image Represent*, 62:12–22, 2019.

[12] Soumia Sid Ahmed, Zoubeida Messali, Larbi Boubchir, Ahmed Bouridane, Sergio Marco, and Cédric Messaoudi. Iterative Bayesian denoising based on variance stabilization using Contourlet Transform with Sharp Frequency Localization: application to EFTEM images. *BMC Biomedical Engineering*, 1(1):1–12, 2019.

[13] S Hosseinian and H Arefi. Assessment of restoration methods of x-ray images with emphasis on medical photogrammetric usage. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41, 2016.

[14] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans Image Process*, 16(8):2080–2095, 2007.

[15] Lucio Azzari and Alessandro Foi. Variance stabilization for noisy+ estimate combination in iterative Poisson denoising. *IEEE Signal Process Lett*, 23(8):1086–1090, 2016.

[16] VN Prudhvi Raj and T Venkateswarlu. Denoising of magnetic resonance and x-ray images using

variance stabilization and patch based algorithms. *The International Journal of Multimedia & Its Applications*, 4(6):53, 2012.

[17] Soeren I. Olsen. Estimation of noise in images: An evaluation. *Graphical Models and Image Processing*, 55(4):319–323, 1993.

[18] Dong-Hyuk Shin, Rae-Hong Park, Seungjoon Yang, and Jae-Han Jung. Block-based noise estimation using adaptive Gaussian filtering. *IEEE Transactions on Consumer Electronics*, 51(1):218–226, 2005.

[19] Sergey Klavdievich Abramov, Vladimir Vasilyevich Lukin, Benoit Vozel, Kacem Chehdi, and Jaakko T Astola. Segmentation-based method for blind evaluation of noise variance in images. *J Appl Remote Sens*, 2(1):023533, 2008.

[20] Marc Liévin, Franck Luthon, and Erwin Keeve. Entropic estimation of noise for medical volume restoration. In *Object recognition supported by user interaction for service robots*, volume 3, pages 871–874, 2002.

[21] Marc Hensel, Bernd Lundt, Thomas Pralow, and Rolf-Rainer Grigat. Robust and fast estimation of signal-dependent noise in medical x-ray image sequences. In *Bildverarbeitung für die Medizin 2006*, pages 46–50. Springer, 2006.

[22] Marc Hensel, Thomas Pralow, and Rolf-Rainer Grigat. Modeling and real-time estimation of signal-dependent noise in quantum-limited imaging. In *International Conference on Signal Processing, Robotics and Automation*. IEEE, 2007.

[23] Stephen M Kengyelics, Amber J Gislason-Lee, Claire Keeble, Derek R Magee, and Andrew G Davies. Noise estimation in cardiac x-ray imaging: a machine vision approach. *Biomedical Physics & Engineering Express*, 2(6):065014, 2016.

[24] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans Image Process*, 17(10):1737–1754, 2008.

[25] Emir Turajlic, Namir Škaljo, and Alen Begović. A block-based noise level estimation from X-ray images in SVD domain. In *International Conference on Systems, Signals and Image Processing*, pages 1–5. IEEE, 2017.

[26] Stanislav Pyatykh, Jürgen Hesser, and Lei Zheng. Image noise level estimation by principal component analysis. *IEEE Trans Image Process*, 22(2):687–699, 2013.

[27] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Single-image noise level estimation for blind denoising. *IEEE Trans Image Process*, 22(12):5226–5237, 2013.

[28] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *International Conference on Computer Vision*, pages 477–485. IEEE, 2015.

[29] Wenzhao Zhao, Yisong Lv, Qiegen Liu, and Binjie Qin. Detail-preserving image denoising via adaptive clustering and progressive PCA thresholding. *IEEE Access*, 6:6303–6315, 2017.

[30] Mikhail Uss, Benoit Vozel, Vladimir Lukin, Igor Baryshev, and Kacem Chehdi. Image noise-informative map for noise standard deviation estimation. In *International Conference on Acoustics, Speech and Signal Processing*, pages 961–964. IEEE, 2011.

[31] Nikolay N Ponomarenko, Vladimir V Lukin, MS Zriakhov, Arto Kaarna, and Jaakko Astola. An automatic approach to lossy compression of AVIRIS images. In *International Geoscience and Remote Sensing Symposium*, pages 472–475. IEEE, 2007.

[32] Miguel Colom and Antoni Buades. Analysis and Extension of the Ponomarenko et al. Method, Estimating a Noise Curve from a Single Image. *Image Processing On Line*, 3:173–197, 2013.

[33] Stanislav Pyatykh and Jürgen Hesser. Image sensor noise parameter estimation by variance stabilization and normality assessment. *IEEE Trans Image Process*, 23(9):3990–3998, 2014.

[34] JH Siewerdsen, LE Antonuk, Y El-Mohri, J Yorkston, W Huang, JM Boudry, and IA Cunningham. Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology. *Med Phys*, 24(1):71–89, 1997.

[35] JH Siewerdsen, LE Antonuk, Y El-Mohri, J Yorkston, W Huang, and IA Cunningham. Signal,

noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology. *Med Phys*, 25(5):614–628, 1998.

[36] Carey E Floyd Jr, Richard J Warp, James T Dobbins III, Harrell G Chotas, Alan H Baydush, Rene Vargas-Voracek, and Carl E Ravin. Imaging characteristics of an amorphous silicon flat-panel detector for digital chest radiography. *Radiology*, 218(3):683–688, 2001.

[37] Bhawna Goyal, Ayush Dogra, Sunil Agrawal, BS Sohi, and Apoorav Sharma. Image denoising review: From classical to state-of-the-art approaches. *Information Fusion*, 55:220–244, 2020.

[38] Sai Gokul Hariharan, Norbert Strobel, Christian Kaethner, Markus Kowarschik, Rebecca Fahrig, and Nassir Navab. An analytical approach for the simulation of realistic low-dose fluoroscopic images. *Int J Comput Assist Radiol Surg*, 14(4):601–610, 2019.

[39] LE Antonuk, Y El-Mohri, JH Siewerdsen, J Yorkston, W Huang, VE Scarpine, and RA Street. Empirical investigation of the signal performance of a high-resolution, indirect detection, active matrix flat-panel imager (AMFPI) for fluoroscopic and radiographic operation. *Med Phys*, 24(1):51–70, 1997.

[40] LE Antonuk, JH Siewerdsen, J Yorkston, and W Huang. Radiation response of amorphous silicon imaging arrays at diagnostic energies. *IEEE Trans Nucl Sci*, 41(4):1500–1505, 1994.

[41] Konstantinos G Derpanis. Overview of the RANSAC Algorithm. *Image Rochester NY*, 4(1):2–3, 2010.

[42] Suman Kumar Maji, Ramesh Kumar Thakur, and Hussein M Yahia. Structure-Preserving Denoising of SAR Images Using Multifractal Feature Analysis. *IEEE Geoscience and Remote Sensing Letters*, 2020.

[43] Lin Chen, Xue Jiang, Zhou Li, Xingzhao Liu, and Zhixin Zhou. Feature-Enhanced Speckle Reduction via Low-Rank and Space-Angle Continuity for Circular SAR Target Recognition. *IEEE Trans Geosci Remote Sens*, 2020.

Appendix A. Experimental Verification of the X-ray Imaging System Model

To verify the imaging model presented in (3), an image quality phantom comprising a Copper step wedge placed on a block of 40 mm Aluminum has been used (see figure A1). The phantom has been designed to represent an object with a water-equivalent thickness between 200 mm and 300 mm depending on the tube voltage. The Copper wedge has 6 steps, indicated as regions of interests (ROIs) 1 - 6 in figure A1, where the thickness of the steps are 0.3 mm, 0.5 mm, 0.8 mm, 1.1 mm, 1.5 mm and 2.0 mm, respectively. This phantom has been used to investigate the effect of beam hardening otherwise caused by different anatomical structures, such as bones and organs.

An NLF according to (3) associated with each ROI, i.e., each step of the phantom, has been computed for three different tube voltages: 80 kVp, 100 kVp and 120 kVp. The data points (mean gray values and associated noise variances) required for computing the NLF (for a specific tube voltage) have been obtained from the X-ray images of the phantom acquired at five different dose levels obtained by varying the mAs settings. The dose levels used for the acquisitions ranged from 35 nGy to 2100 nGy. These values have been measured by placing a dosimeter (RaySafe X1, Fluke Corporation, Glenwood, Ill, USA) on the detector's surface. In figure A1, the NLFs associated with the the steps of 0.3 mm and 2.0 mm thickness of the Copper wedge (highlighted as ROIs 1 and 6 in figure A1) have been presented for the tube voltages 80 kVp, 100 kVp and 120 kVp, respectively. It can be observed that the NLFs associated with the ROIs follow a linear

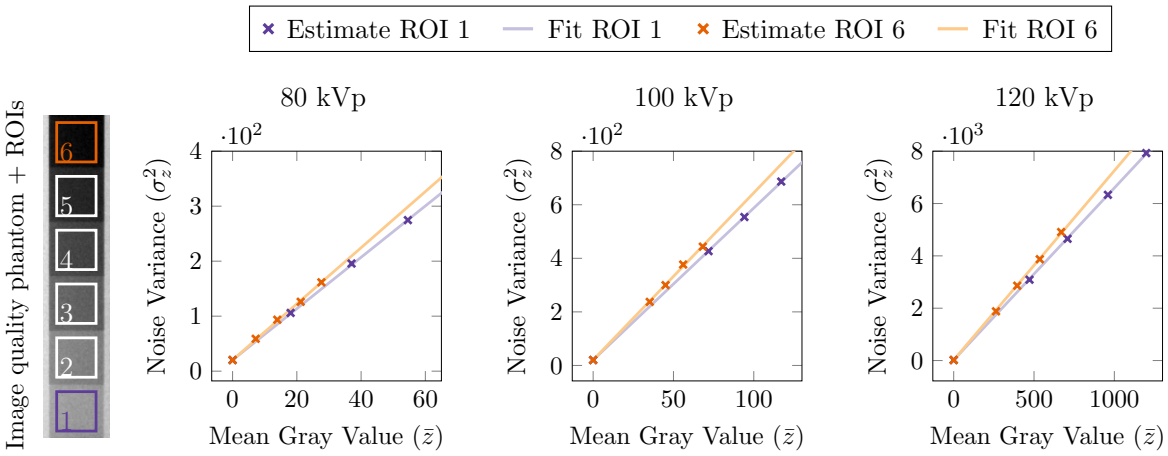


Figure A1: Visualization of an X-ray image of an image quality phantom comprising a Copper step wedge with six steps (0.3 mm, 0.5 mm, 0.8 mm, 1.1 mm, 1.5 mm and 2.0 mm, respectively) placed on a block of 40 mm of Aluminum and the noise level functions (NLFs) associated with ROIs 1 and 6. The NLFs have been measured by varying tube currents for the tube voltages 80 kVp, 100 kVp, and 120 kVp, respectively.

relationship according to the model presented in (3). In table A1, the values for $\alpha[r, c]$ associated with all the steps of the phantom are listed. The computed values confirm that $\alpha[r, c]$ depends on the tube voltage and the object imaged, i.e., it depends on the spectrum of the X-ray photons received at the detector [40]. They also indicate that a "harder" beam is associated with a higher $\alpha[r, c]$.

Table A1: System gain α associated with different thickness of Copper (placed on block of Aluminum) and different tube voltages.

ROI	**Object**	α **estimated for**		
		80 kVp	**100 kVp**	**120 kVp**
1	40 mm Al + 0.3 mm Cu	4.669	5.674	6.574
2	40 mm Al + 0.5 mm Cu	4.706	5.733	6.654
3	40 mm Al + 0.8 mm Cu	4.775	5.824	6.811
4	40 mm Al + 1.1 mm Cu	4.966	6.038	7.004
5	40 mm Al + 1.5 mm Cu	5.073	6.186	7.154
6	40 mm Al + 2.0 mm Cu	5.105	6.243	7.234

Appendix B. Impact of Pixel-dependent Gains on NVS

To study the impact of the pixel-dependent gains $\alpha[r, c]$ on the NVS based on GAT experimentally, we have utilized the ROIs highlighted in figure A1. To find the

differences in the standard deviations after NVS, the GAT has been performed using (i) an accurate estimate of $\alpha[r, c]$ and (ii) an average estimate $\bar{\alpha}$. The average estimate $\bar{\alpha}$ has been computed by averaging the estimates of $\alpha[r, c]$ obtained for a specific tube voltage (shown in table A1). The standard deviation of the noise variance stabilized ROIs associated with different tube settings, 80 kVp, 100 kVp and 120 kVp, are presented in table B1. It can be observed that the ROIs measured in thicker Copper steps of the phantom (associated with a higher $\alpha[r, c]$) have a marginally higher standard deviation, when $\bar{\alpha}$ is used for performing the GAT. On the contrary, regions measured in thinner steps appear to have a lower standard deviation when $\bar{\alpha}$ is used. Overall, the standard deviation of noise obtained in the case of $\bar{\alpha}$ has deviated from that obtained for specific $\alpha[r, c]$ by only 1.59%.

Table B1: Standard deviations of the noise σ associated with noise variance stabilized ROIs obtained based on accurate α and average estimates $\bar{\alpha}$ of the system gain for the tube voltages 80 kVp, 100 kVP and 120 kVP.

80 kVP

ROI	**15 mAs**			**30 mAs**			**60 mAs**			**90 mAs**		
	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$
1	1.005	0.987	0.019	0.994	0.973	0.020	1.022	1.001	0.021	1.053	1.031	0.022
2	0.990	0.976	0.015	0.990	0.974	0.016	1.010	0.992	0.017	1.071	1.052	0.019
3	1.024	1.015	0.009	1.010	1.000	0.010	1.021	1.011	0.010	1.021	1.010	0.011
4	1.022	1.028	0.006	0.993	1.001	0.007	1.017	1.026	0.008	1.040	1.048	0.008
5	1.033	1.047	0.014	1.019	1.034	0.016	0.994	1.011	0.017	1.007	1.025	0.018
6	1.016	1.032	0.015	1.009	1.027	0.018	0.997	1.017	0.020	0.996	1.016	0.020

100 kVP

ROI	**15 mAs**			**20 mAs**			**25 mAs**			**30 mAs**		
	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$
1	1.001	0.978	0.023	0.976	0.954	0.023	1.028	1.004	0.024	1.026	1.002	0.024
2	1.026	1.007	0.018	1.016	0.997	0.018	1.029	1.010	0.019	1.019	1.000	0.018
3	1.009	0.999	0.010	1.034	1.023	0.011	1.019	1.008	0.011	1.043	1.032	0.011
4	1.006	1.013	0.007	1.016	1.023	0.007	1.015	1.023	0.007	1.023	1.030	0.007
5	1.015	1.034	0.019	0.989	1.008	0.019	1.000	1.019	0.019	0.979	0.998	0.019
6	0.977	0.999	0.023	0.973	0.996	0.023	0.995	1.019	0.024	1.001	1.024	0.024

120 kVP

ROI	**15 mAs**			**16 mAs**			**17 mAs**			**18 mAs**		
	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$	σ_α	$\sigma_{\bar{\alpha}}$	$\lvert\sigma_\alpha - \sigma_{\bar{\alpha}}\rvert$
1	1.024	0.999	0.025	1.085	1.059	0.026	1.098	1.072	0.027	1.105	1.078	0.027
2	1.019	1.001	0.019	1.078	1.058	0.020	1.129	1.108	0.021	1.177	1.155	0.022
3	1.024	1.017	0.007	1.107	1.099	0.008	1.155	1.147	0.008	1.167	1.159	0.008
4	1.049	1.056	0.007	1.105	1.113	0.008	1.135	1.143	0.008	1.146	1.154	0.008
5	1.030	1.048	0.018	1.124	1.144	0.020	1.197	1.218	0.021	1.226	1.248	0.022
6	1.062	1.087	0.025	1.107	1.133	0.026	1.093	1.119	0.026	1.126	1.152	0.027

Appendix C. Influence of NVS on Denoising

We have studied the influence of NVS (based on GAT) obtained using the method by Pyatykh et al. [33], the UPM and the WPM on the denoising performance (table C1 and table C2). For this, we have first applied an NVS on the phantom images – acquired using System-1 and System-2 – based on the parameters ($\bar{\alpha}$ and $\bar{\rho}$) obtained using the above mentioned methods as well as computed GT imaging parameters. We have subsequently denoised the transformed images using BM3D [14]. Then, we have applied an unbiased inverse GAT to return to the original image domain [10]. We have also compared the results with those of the method by Azzari et al. [15]. In table C1 and table C2, we present the quantitative analysis of the results in the cases of System-1 and System-2, respectively, with respect to PSNR and SSIM. It can be seen that when the parameters obtained using the WPM are considered for performing the NVS, the quality metrics are most similar to those obtained using GT imaging parameters, i.e., the PSNR values in the cases of System-1 and System-2 differ on an average by 0.17% and 0.19%, respectively. The SSIM values in the cases of System-1 and System-2 differ on an average by 0.09% and 0.05%, respectively. On the other hand, when the UPM is used, there is a slightly higher deviation in the PSNR (0.75% for System-1 and 0.37% for System-2) and SSIM (0.43% for System-1 and 0.10% for System-2). When the method by Pyatykh et al. [33] is used, the deviation is even higher with respect to PSNR (3.62% for System-1 and 5.31% for System-2) and SSIM (1.75% for System-1 and 1.61% for System-2). Finally, in the application of the method by Azzari et al. [15], the deviation is similarly high with respect to PSNR (8.80% for System-1 and 0.94% for System-2) and SSIM (6.78% for System-1 and 0.29% for System-2).

By comparing the improvement in the image quality – with respect to the input – obtained using the above mentioned methods, we have found that in the case of System-1, the method by Azzari et al. [15] has resulted in a low improvement (19.94% for PSNR and 37.86% for SSIM), whereas, the other methods, BM3D with the method by Pyatykh et al. [33] (27.01% for PSNR and 47.40% for SSIM), BM3D with the UPM (30.77% for PSNR and 49.56% for SSIM) and BM3D with the WPM (31.57% for PSNR and 50.15% for SSIM), have yielded a significantly higher improvement. In the case of System-2, BM3D with the method by Pyatykh et al. [33] has yielded a small improvement over the input (13.74% for PSNR and 9.39% for SSIM), whereas the application of the other methods, the method by Azzari et al. [15] (18.97% for PSNR and 10.88% for SSIM), BM3D with the UPM (19.68% for PSNR and 11.11% for SSIM) and BM3D with the WPM (19.94% for PSNR and 11.17% for SSIM), have improved the quality by a higher factor.

Table C1: Comparison of the impact of noise variance stabilization (NVS) on denoising using block matching 3D (BM3D) [14] with respect to peak signal-to-noise-ratio (PSNR) and structural similarity index (SSIM) in the case of System-1 (a general-purpose angiography system). The NVS performed using the generalized Anscombe transform (GAT) is based on the parameters estimated using Pyatykh et al. [33], the unweighted Ponomarenko method (UPM) and the weighted Ponomarenko method (WPM) in addition to the ground truth (GT) parameters. For the comparison, X-ray images of a thorax phantom acquired at different dose levels and scatter levels (generated using polymethyl methacrylate (PMMA) of different thickness) have been used. The results of Azzari et al. [15] are also presented. The values closest to the GT are highlighted.

PMMA	**Dose**	**PSNR**					
		Input	**Azzari et al.**	**Denoised using BM3D with GAT**			
				Pyatykh et al.	**UPM**	**WPM**	**GT**
2 cm	25%	27.231	32.448	36.170	37.416	**37.454**	37.485
	50%	30.915	37.232	38.716	40.234	**40.265**	40.293
	100%	34.148	40.549	40.729	42.232	**42.272**	42.292
4 cm	25%	26.760	31.968	35.382	34.910	**36.438**	36.640
	50%	30.383	36.587	38.189	39.505	**39.566**	39.561
	100%	33.610	40.065	40.388	41.764	**41.797**	41.824
8 cm	25%	26.105	31.210	34.758	35.439	**35.866**	35.991
	50%	29.598	35.304	36.389	37.120	**37.374**	37.439
	100%	33.186	39.589	39.891	41.129	**41.275**	41.270
16 cm	25%	24.839	30.019	34.339	35.726	**35.737**	35.853
	50%	27.883	34.179	36.080	37.677	**37.700**	37.782
	100%	28.351	34.110	35.387	36.645	**36.704**	36.747

PMMA	**Dose**	**SSIM**					
		Input	**Azzari et al.**	**Denoised using BM3D with GAT**			
				Pyatykh et al.	**UPM**	**WPM**	**GT**
2 cm	25%	0.551	0.836	0.928	0.946	**0.947**	0.947
	50%	0.770	0.940	0.957	**0.970**	**0.970**	0.970
	100%	0.883	0.972	0.973	**0.981**	**0.981**	0.981
4 cm	25%	0.549	0.834	0.922	0.914	**0.939**	0.942
	50%	0.763	0.938	0.957	0.968	**0.969**	0.969
	100%	0.879	0.971	0.973	0.980	**0.981**	0.981
8 cm	25%	0.485	0.800	0.908	0.921	**0.929**	0.930
	50%	0.744	0.930	0.948	0.958	**0.961**	0.961
	100%	0.871	0.969	0.971	0.978	**0.979**	0.979
16 cm	25%	0.375	0.739	0.893	**0.921**	**0.921**	0.923
	50%	0.608	0.889	0.927	**0.949**	**0.949**	0.950
	100%	0.638	0.897	0.925	0.946	**0.947**	0.948

Table C2: Comparison of the impact of noise variance stabilization (NVS) on denoising using block matching 3D (BM3D) [14] with respect to peak signal-to-noise-ratio (PSNR) and structural similarity index (SSIM) in the case of System-2 (an angiography system optimized for low-dose X-ray applications). The NVS performed using the generalized Anscombe transform (GAT) is based on the parameters estimated using Pyatykh et al. [33], the unweighted Ponomarenko method (UPM) and the weighted Ponomarenko method (WPM) in addition to the ground truth (GT) parameters. For the comparison, X-ray images of a thorax phantom acquired at different dose levels and scatter levels (generated using polymethyl methacrylate (PMMA) of different thickness) have been used. The results of Azzari et al. [15] are also presented. The values closest to the GT are highlighted.

PMMA	**Dose**	**PSNR**					
		Input	**Azzari et al.**	**Denoised using BM3D with GAT**			
				Pyatykh et al.	**UPM**	**WPM**	**GT**
2 cm	25%	32.337	39.118	37.879	39.664	**39.700**	39.780
	50%	35.620	42.152	40.176	42.332	**42.346**	42.406
	100%	38.684	44.849	42.560	44.886	**44.955**	44.960
4 cm	25%	31.860	38.681	37.257	39.090	**39.151**	39.287
	50%	35.239	41.959	40.050	42.139	**42.217**	42.286
	100%	37.961	43.028	41.318	43.039	**43.053**	43.112
8 cm	25%	31.299	38.054	36.668	38.234	**38.520**	38.639
	50%	34.881	41.411	39.435	41.560	**41.769**	41.710
	100%	37.855	43.923	41.913	43.928	**43.985**	44.060
16 cm	25%	29.888	36.858	35.486	37.457	**37.475**	37.572
	50%	32.235	39.011	36.896	39.279	**39.325**	39.479
	100%	35.043	41.351	39.038	41.537	**41.687**	41.662

PMMA	**Dose**	**SSIM**					
		Input	**Azzari et al.**	**Denoised using BM3D with GAT**			
				Pyatykh et al.	**UPM**	**WPM**	**GT**
2 cm	25%	0.824	0.961	0.948	**0.966**	**0.966**	0.967
	50%	0.911	0.980	0.968	**0.981**	**0.981**	0.981
	100%	0.955	0.989	0.981	**0.989**	**0.989**	0.989
4 cm	25%	0.825	0.962	0.947	0.965	**0.966**	0.967
	50%	0.914	0.981	0.971	**0.982**	**0.982**	0.982
	100%	0.955	0.988	0.980	**0.988**	**0.988**	0.988
8 cm	25%	0.804	0.957	0.940	0.958	**0.961**	0.962
	50%	0.904	0.979	0.966	0.979	**0.980**	0.980
	100%	0.951	0.988	0.981	**0.988**	**0.988**	0.988
16 cm	25%	0.760	0.947	0.928	**0.954**	**0.954**	0.955
	50%	0.854	0.968	0.948	**0.970**	**0.970**	0.971
	100%	0.920	0.981	0.967	**0.982**	**0.982**	0.982

4.3. Summary

The use of an accurate NVS transform before applying denoising algorithms designed for the removal of additive Gaussian noise with a known variance is a viable strategy due to its efficiency and robustness. In this chapter, a data-driven method for estimating the imaging parameters required for performing an NVS using the generalized Anscombe transform has been presented. Compared to previous approaches, significant performance gains in the outcome of denoising has been achieved when using the proposed method prior to the application of a well established denoising algorithm designed for optical images. In the following chapters, novel denoising algorithms tailored for X-ray images, in order to enable low-dose X-ray imaging, are presented. The model-based X-ray imaging parameter estimation approach presented in this chapter is used for performing an NVS prior to the application of the denoising approaches.

5. Analytical Approaches for X-ray Image Denoising

5.1. Introduction

Reducing ionizing radiation is essential to minimize the consequences of exposure to radiation for patients and clinical staff. To reduce the X-ray dose and at the same time retain the required image quality, spatio-temporal denoising can be a suitable approach for certain applications. In this chapter, two spatio-temporal denoising approaches targeted at denoising of ultra-low dose X-ray fluoroscopic and digital subtraction angiography (DSA) sequences are presented.

5.2. A photon recycling approach to the denoising of ultra-low dose X-ray sequences

5.2.1. Paper Summary

To keep the X-ray dose as low as reasonably achievable while navigating surgical instruments under X-ray guidance, a patch-based spatio-temporal denoising approach is proposed in the following paper [2]. The method is inspired partly by patch-based denoising algorithms designed originally for optical images (such as [117, 188, 189]), and suitably modified and applied to ultra-low dose X-ray sequences. The motivation behind the proposed approach is to use accurate motion estimation to match patches from previous frames and subsequently use the associated pixels to improve the image quality.

The approach utilizes a sophisticated noise model to exploit the non-local self-similarity in the temporal direction while matching patches as well as computing a low-rank approximation of the temporally aligned patches. In addition, it has been shown that (i) by carefully taking into account the noise characteristics, robust patch matching can be achieved even in a high noise environment, and (ii) by carefully using noise characteristics while performing spatio-temporal denoising of ultra-low dose X-ray sequences, the necessity of a guide image, that is normally required to obtain good denoising performance, can be eliminated. This reduces the computational cost significantly.

An evaluation of the algorithm using 500 clinical images acquired at 50% of the standard low dose level has revealed that an average improvement in the contrast-to-noise ratio (CNR) by a factor of around 3.5 can be achieved. A qualitative evaluation by image quality experts has suggested that the proposed strategy of recycling photons from previous frames results in denoised images that comply with the required image quality criteria. The qualitative analysis by experts has also confirmed that the denoised ultra-low dose X-ray images obtained using the proposed method are more realistic compared to

existing state-of-the-art spatio-temporal denoising techniques with respect to appearance. However, at ultra low-dose levels, it is possible that systems designed for higher exposure levels no longer work as expected. For example, the beam energy may become inconsistent. In such a situation, patch matching may no longer yield suitable results causing sub-optimal denoising. Another drawback is that the denoised images may suffer from mild blurring around instrument edges. This is caused by imperfect matches due to the 3D motion of objects followed by low-rank approximation. In such a case, the low-rank approximation will result in the averaging of pixels that have not been well registered.

A photon recycling approach to the denoising of ultra-low dose X-ray sequences

Sai Gokul Hariharan[1,2] · **Norbert Strobel**[2,3] · **Christian Kaethner**[2] · **Markus Kowarschik**[1,2] · **Stefanie Demirci**[1] · **Shadi Albarqouni**[1] · **Rebecca Fahrig**[2,4] · **Nassir Navab**[1,5]

Abstract

Purpose. Clinical procedures that make use of fluoroscopy may expose patients as well as the clinical staff (throughout their career) to non-negligible doses of radiation. The potential consequences of such exposures fall under two categories, namely stochastic (mostly cancer) and deterministic risks (skin injury). According to the "as low as reasonably achievable" principle, the radiation dose can be lowered only if the necessary image quality can be maintained.

Methods. Our work improves upon the existing patch-based denoising algorithms by utilizing a more sophisticated noise model to exploit non-local self-similarity better and this in turn, improves the performance of low rank approximation. The novelty of the proposed approach lies in its properly designed and parameterized noise model and the elimination of initial estimates. This reduces the computational cost significantly.

Results. The algorithm has been evaluated on 500 clinical images (7 patients, 20 sequences, 3 clinical sites), taken at ultra-low dose levels, i.e., 50% of the standard low dose level, during electrophysiology procedures. An average improvement in the contrast-to-noise ratio (CNR) by a factor of around 3.5 has been found. This is associated with an image quality achieved at around 12 (square of 3.5) times the ultra-low dose level. Qualitative evaluation by X-ray image quality experts suggests that the method produces denoised images that comply with the required image quality criteria.

Conclusion. The results are consistent with the number of patches used, and

[1]Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2]Siemens Healthcare GmbH, Advanced Therapies, Forchheim, Germany
[3]Fakultät für Elektrotechnik, Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt, Schweinfurt, Germany
[4]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[5]Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America

they demonstrate that it is possible to use motion estimation techniques and "recycle" photons from previous frames to improve the image quality of the current frame. Our results are comparable in terms of CNR to Video Block Matching 3D (VBM3D) - a state-of-the-art denoising method. But qualitative analysis by experts confirms that the denoised ultra-low dose X-ray images obtained using our method are more realistic with respect to appearance.

1 Introduction

X-ray guided interventional procedures have become increasingly important in medical practice over the years. As new procedures, such as transcatheter aortic valve implantation procedures, are introduced and validated, they either tend to replace the equivalent surgical procedures or may enable new (hybrid) surgical interventions. Unfortunately, the medical benefits are accompanied by health risks due to radiation exposure. The stochastic risk is present for both the patient and the staff, and there may also be a deterministic risk of skin injury to the patient in some special cases [12]. This is why a reduction in X-ray dose has been an active and important area of research for several decades by now. Unfortunately, X-ray dose reduction typically results in a sacrifice of image quality. For example, if the dose is lowered by a factor of four, the signal-to-noise-ratio (SNR) will drop by a factor of two. In order to preserve the clinically required image quality, sophisticated image sequence processing techniques can be applied. Fortunately, we have ever-increasing computational resources at our disposal which we can employ to enhance image quality.

The most obvious approach to reduce X-ray exposure during fluoroscopy procedures involving successive frames of X-ray images is to either reduce the frame rate or the X-ray exposure per frame. Since lowering the frame rate may introduce jerkiness (abrupt motion), temporal interpolation methods have been applied to smoothen the appearance of the resulting X-ray sequences [10]. Due to the requirement of continuous visualization of flowing contrast media, Parker *et al.* [20] have investigated changing the X-ray intensity on a per-frame basis by varying it between two extreme levels (0.5mAs and 10mAs) while keeping the tube voltage constant. They have shown that a weighted combination of multiple X-ray images obtained at different X-ray intensities could improve the signal to noise to dose ratio.

Noise reduction has been an important requisite not only for X-ray images but also for optical still images and medical images. A prominent transform based denoising method that involves sophisticated thresholding of undecimated wavelet coefficients has been proposed by Luisier *et al.* [18]. Similarly, multiscale approaches [17] as well as diffusion filters [23] have been shown to produce good denoised images. Denoising methods have also been attempted to take advantage of the self-similar structures present in most images. Nonlocal-means [4, 8], nonlocal-median based approaches [6, 11] and block matching 3D (BM3D) [9] are well-known methods of this kind. A related approach involves redundant representations of patches taken from patch dictionaries [15]. The

use of patches for denoising have been exploited further by Albarqouni *et al.* [1], who have proposed a patch-based regularized graph spectral filtering technique. Learning based approaches, such as cascade of shrinkage fields [21] and trainable nonlinear reaction diffusion [7] have also been shown to outperform the well engineered BM3D for optical color still images. Deep learning has also facilitated the design of blind denoising methods [25]. Cerciello *et al.* [5] have compared the performance of some common denoising algorithms, designed for optical images - corrupted by independent additive, white noise or by signal-dependent noise - on X-ray fluoroscopy data and have shown that applying the version of BM3D designed for signal-dependent noise is the most effective approach.

While the denoising of single image frames is important, additional gains can be obtained by the joint processing of successive X-ray images taken during fluoroscopic procedures. Amiot *et al.* [2] have proposed the use of recursive temporal filtering in a multi-scales space and dual-tree complex wavelet domain. Wagner *et al.* [22] have developed an isotropic filtering technique that involves dedicated directional kernels. Lee *et al.* [16] have designed a 3D non-local means filter based on stochastic information that incorporates motion information. VBM3D (Video Block Matching 3D) [14], a video denoising method designed for optical video sequences, also provides excellent performance even for X-ray images [2]. It groups patches taken from the same frame and across frames and subsequently performs collaborative filtering. While Cerciello *et al.* [5] suggest the use of BM3D designed for signal-dependent noise, Boracchi *et al.* [3] emphasize the necessity of applying a noise variance stabilization transform before processing raw images using VBM3D.

Though VBM3D provides favorable denoising performance at today's range of X-ray dose levels, its performance may drop at ultra-low dose levels. While it is still possible to obtain a good contrast-to-noise ratio (CNR) under these circumstances, structures and regions may be oversmoothed [2]. Nevertheless, BM3D and VBM3D are excellent denoising methods [5] and a more recently proposed method [2], specifically designed for X-ray imaging, outperforms them only by a small degree for artificially introduced noise.

In order to obtain good image quality in ultra-low dose environments at an acceptable computational cost, we have exploited non-local self-similarity in the temporal direction by matching associated patches followed by computing a low-rank approximation of the temporally aligned patches in the SVD domain. The method is inspired partly by the algorithm proposed in [13], which has been designed originally for optical images, and leveraged to ultra-low dose X-ray sequences. Our goal is not only to remove noise, but also to produce images that are free from artefacts such that they create the impression of higher dose images. Our main contribution is the design of a dedicated algorithm for the denoising of ultra-low dose X-ray (fluoroscopy) sequences by recycling photons received in the past. In particular, we show that i) by carefully taking noise characteristics into account, we can achieve robust patch matching even in a high noise environment, and ii) by carefully using noise characteristics while performing spatio-temporal denoising of ultra-low dose X-ray sequences,

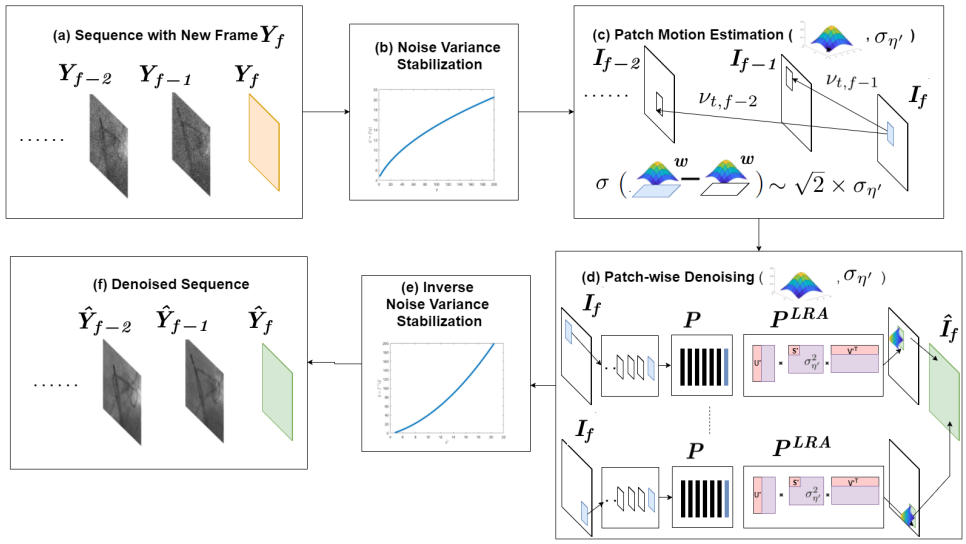


Fig. 1 Workflow of the spatio-temporal denoising algorithm.

we can eliminate the necessity of a guide image that is normally required to obtain good denoising performance.

2 Method

In X-ray images, noise has various sources (e.g., Poisson noise, electronic noise, quantization noise) and, depending on the imaging situation, different detector gains may be selected. To tackle this issue, we propose a denoising method that is partly related to the two-step denoising strategy designed for mixed noise environments [3]. The outline of the algorithm is presented in Fig. 1. As a first step, we stabilize the noise variance (signal-dependent mixed noise) across the input images (Fig. 1(a)) using the generalized Anscombe transform [19] (Fig. 1(b)). In the second step, we use weighted patch matching to estimate patch motion and align similar image regions in the current (reference) frame and the previous frames (Fig. 1(c)). We then apply patch-wise denoising to the spatial and temporal patches aligned along the temporal direction using a noise level based low rank approximation (Fig. 1(d)). This is followed by a weighted aggregation of the patches to obtain the denoised image. Finally, an inverse noise variance stabilization transform is applied to return to the X-ray image domain (Fig. 1(e)). The core of the algorithm is described below.

2.1 Noise Model

The formation of images from X-ray quanta can be assumed to follow a linear model [24]. The mean gray value $\bar{\boldsymbol{Y}}[r,c]$ at a particular location $[r,c]$ has a linear relationship with respect to the incident mean X-ray air kerma $\bar{\boldsymbol{X}}$ at that

location. That is, the measured mean quanta $\bar{\boldsymbol{X}}[r,c]$ is scaled by the detector gain α and shifted by the overall system offset g. This can be represented using the following equation:

$$\bar{\boldsymbol{Y}}[r,c] = \alpha \cdot \bar{\boldsymbol{X}}[r,c] + g. \tag{1}$$

The quantum nature of X-ray photons can be modelled using a Poisson distribution. The presence of electronic noise, e.g., due to read out noise and dark noise, can be modelled using a Gaussian distribution with zero mean. The formation of an image can thus be represented as:

$$\begin{aligned} \boldsymbol{Y}[r,c] &= \alpha \cdot \big(\boldsymbol{X}[r,c] + \boldsymbol{\eta_a}[r,c]\big) + p + \boldsymbol{\eta_q}[r,c] \\ &= \alpha \cdot \boldsymbol{X}[r,c] + g + \boldsymbol{\eta}[r,c], \end{aligned} \tag{2}$$

where $\boldsymbol{X}$, $\boldsymbol{\eta_a}$, $\boldsymbol{\eta_q}$ and $\boldsymbol{Y}$ represent the X-ray quanta, electronic noise due to analog sensor read out, electronic noise due to quantization (analog to digital conversion) and the noise-corrupted detector pixel value, respectively. The overall additive noise can be represented as $\boldsymbol{\eta}$. By taking into account the system parameters, the noise variance of the mixed signal-dependent noise can be stabilized to a constant $\sigma^2_{\eta'}$ using the generalized Anscombe transform [19], i.e., after transformation, the noise can be modelled using a Gaussian distribution with variance $\sigma^2_{\eta'}$. The following equation describes this process:

$$\boldsymbol{I}[r,c] = \begin{cases} \frac{2}{\alpha}\sqrt{\alpha \cdot \boldsymbol{Y}[r,c] + \frac{3}{8}\alpha^2 + \sigma^2_\eta - \alpha \cdot g} & \text{if } \boldsymbol{Y}[r,c] > -\frac{3}{8}\alpha - \frac{\sigma^2_\eta}{\alpha} + g \\ 0 & \text{if } \boldsymbol{Y}[r,c] \leq -\frac{3}{8}\alpha - \frac{\sigma^2_\eta}{\alpha} + g. \end{cases} \tag{3}$$

where $\boldsymbol{I} \in \mathbb{R}^{N \times M}$ represents the image after noise variance stabilization to $\sigma^2_{\eta'}$. In order to represent a frame in a sequence (after noise variance stabilization), we have introduced a subscript f to $\boldsymbol{I}$ ($\boldsymbol{I_f} \in \mathbb{R}^{N \times M}$).

2.2 Patch Motion Estimation

Criteria. To avoid matching of noise patterns at high noise levels, we include the characteristics of noise during patch matching. For a perfect match in noise variance stabilized images, the standard deviation of the difference between the matched and the reference patch, σ_d, should be close to $\sqrt{2} \times \sigma_{\eta'}$. If σ_d is much greater or much smaller than $\sqrt{2} \times \sigma_{\eta'}$, we can conclude that it is a poor match. Due to the random nature of noise, we introduce a threshold τ to compensate for the randomness, i.e, in the case of a good match the following condition is satisfied:

$$\left|\sigma_d - \sqrt{2} \times \sigma_{\eta'}\right| \leq \tau. \tag{4}$$

Strategy. As patch-wise denoising is applied, for each vectorized patch $q_{(\vec{t,f})} \in \mathbb{R}^{k^2 \times 1}$ (t represents the index of a patch of frame f and k represents the patch width), a set of similar patches is required. For simplicity, we use vectorized versions of patches. Usually, X-ray images present a stationary background (e.g., spine, rib cage), possibly some moving organs (lung, heart), and instruments navigated by a physician. Using the characteristics of noise, moving and non-moving parts across the frames of X-ray images can be identified. As a first step, we compare patches at the same locations in the pairs of images (reference frame $\boldsymbol{I_f}$ and previous frames $\boldsymbol{I_{f'}}$) using Eq. 4 to obtain stationary regions. For the non-stationary regions, we use a multi-scale search strategy using a weighted sum of the squared distance to obtain a set of temporally matched patches. The weight vector $\vec{w} \in \mathbb{R}^{k^2 \times 1}$ is Gaussian distributed to ensure that the pixels close to the patch center get more importance during patch matching. Patches in $\boldsymbol{I_{f'}}$ that are within the radius of r_t pixels around the current patch center are checked in this process. Eq. 5 describes the weighted patch matching process for patch t between frame f and f':

$$\nu_{t,f'} = \underset{\nu_{t,f'}}{\arg\min} \sum_{i=1}^{k^2} \left\{ \left(q_{(\vec{t,f})}[i] - q_{(t+\vec{\nu_{t,f'}},f')}[i] \right)^2 \cdot \vec{w}[i] \right\}, \tag{5}$$

where $\nu_{t,f'}$ is the displacement between the indices that matches patch $q_{(\vec{t,f})}$ with $q_{(t+\vec{\nu_{t,f'}},f')}$ and i is the pixel index of the patch.

Stack Building. Finally, for each patch $q_{(\vec{t,f})}$ in the reference image, a set of temporally matched patches is obtained using Eq. 5. To reduce the complexity introduced due to notations, we remove the subscripts t and f from here onwards as it can be generalized over all the patches, i.e, $q_{(\vec{t,f})}$ will be simplified to $\vec{q}$. In order to compensate for unmatchable deformable motion of instruments in some frames, we include spatial patches within the radius of at most r_s pixels from the reference patches in addition to the matched temporal patches. The reference and the matched patches are then vectorized and stacked into a matrix $\boldsymbol{P} \in \mathbb{R}^{k^2 \times m}$ (general notation $\boldsymbol{P_{t,f}}$) such that $\vec{p_m} \in \mathbb{R}^{k^2 \times 1}$ represents the vectorized reference patch and $p_{1..\vec{m-1}}$ represent the vectorized matched patches.

From the m vectorized patches, we then select only the most suitable n patches with the least deviation from expected standard deviation for a difference patch as per Eq. 4. This helps in removing those spatial and temporal patches that are not sufficiently similar to the reference patch. The deviations or the errors, $e \in \vec{\mathbb{R}}^{1 \times m}$, for all the vectors in $\boldsymbol{P}$ are computed by comparing $\vec{p_m}$ and the other vectors $\vec{p_{m'}}$ with respect to the amount of noise by:

$$e_{m'} = \left| \sigma(\vec{p_m} - \vec{p_{m'}}) - \sqrt{2} \times \sigma_{\eta'} \right|. \tag{6}$$

$\boldsymbol{P}$ is then constructed in such a way that $\vec{p_n}$ is the vectorized reference patch.

2.3 Denoising of Temporally Aligned Patches

In the ideal case, where the patch matching has been perfect, the noisy matrix $\boldsymbol{P}$ is a combination of a matrix that has a low rank and a noise matrix of standard deviation $\sigma_{\eta'}$. The optimal energy compaction of singular value decomposition in the least square sense makes it a suitable choice for estimating the low rank matrix [13]. From the singular value decomposition of $\boldsymbol{P}$ - into a singular value matrix $S \in \mathbb{R}^{k^2 \times n}$ with singular values s_i present in non-increasing order in the diagonal elements and the associated singular vector matrices $U \in \mathbb{R}^{k^2 \times k^2}$ and $V \in \mathbb{R}^{n \times n}$ - the low rank approximation can be obtained by taking only a few large singular values and the corresponding singular vectors. The choice of the largest l singular values required for obtaining the low rank approximation, is computed by solving the following equation [13]:

$$\sum_{i=l}^{n} s_i^2 > \sigma_{\eta'}^2 \times (k^2 \cdot n - 1) \geq \sum_{i=l+1}^{n} s_i^2. \tag{7}$$

The low rank matrix $\boldsymbol{P^{LRA}} \in \mathbb{R}^{k^2 \times n}$ can then be constructed from the l largest singular values as follows:

$$\boldsymbol{P^{LRA}} = \sum_{i=1}^{l} s_i \vec{u_i} \vec{v_i^T}. \tag{8}$$

We then set $p_n^{\vec{LRA}}$ as the vectorized denoised patch $\hat{}_{(t,f)}\vec{q}$. The denoised patches $\hat{}_{(t,f)}^{\vec{LRA}}q$ are finally combined using weighed aggregation to reconstruct the denoised image $\hat{\boldsymbol{Y}}_{\boldsymbol{f}}$.

3 Experiments and Results

We have evaluated the proposed denoising method on 20 clinical sequences (3 sites, 7 patients, 500 images) acquired at 50% of the standard low dose setting (ultra-low dose). 7 sequences (137 images) from site 1, 11 sequences (292 images) from site 2 and 2 sequences (71 images) from site 3 have been used. All X-ray sequences have been acquired using using Artis zee systems (Siemens Healthineers, Erlangen, Germany). Site 1 and site 3 have a 26×30 cm^2 crystalline Silicon flat panel detector, and site 2 has a 20×20 cm^2 amorphous Silicon flat panel detector. The ultra-low dose clinical sequences have been acquired during cardiac and renal electrophysiology (EP) procedures. The image sequences have been acquired either at a matrix size of 1024×1024 or 960×960 and at 3 frames per second. We have compared our method to the well established VBM3D as recent methods have not significantly outperformed VBM3D for images that have been corrupted by real (non-synthetic) noise [2]. Since VBM3D has been designed for additive Gaussian noise, we

have performed VBM3D in the generalized Anscombe transform domain after stabilizing the noise variance [19]. To perform quantitative evaluation, we have computed an average CNR from 2 pairs of manually selected regions within and outside instruments. We have also obtained qualitative evaluations from twelve X-ray image quality experts.

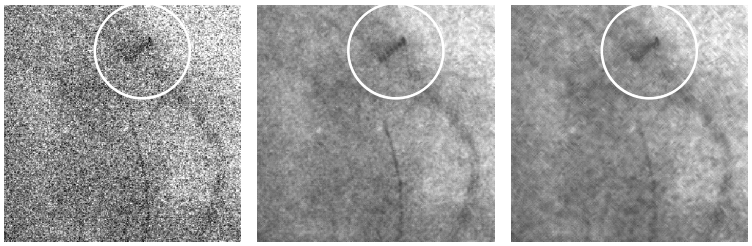
3.1 Parameter Setting

We have evaluated the algorithm for the patch width $k = 9$ and for different choices of aligned spatio-temporal patches in the stack ($n = 3, 5..\ 20$). We have chosen to use $k = 9$, as we have found this setting to preserve structures while still yielding good denoising. Since temporal coherence drops over time, we have restricted the temporal history to at the most 11 frames, i.e., to somewhat less than four seconds. For temporal patch motion estimation, r_t is set to 150 pixels to compensate for the comparatively low frame rate. The threshold τ for checking whether pixels have undergone motion is set to $0.05 \times \sigma_{\eta'}$. To reduce the computational load, only those spatial patches within $r_s = 2$ pixels around the reference patch and $r_s = 1$ around the matched patches have been considered as they are expected to share similar characteristics especially along the edges. From the set of all these matched patches, we have chosen n patches to perform spatio-temporal denoising. We have used patches only from the past (at the most 11) frames for VBM3D as it is not possible to use future frames during live X-ray imaging. Using the available VBM3D software [14], we have selected the normal balanced profile and computed the denoised sequence.

3.2 Results

In order to highlight the importance of using noise characteristics while matching patches, we have shown in Fig. 2 the results of denoising using low rank approximation when patches have been matched with and without taking the characteristics of noise into account. In the selected region of interest (ROI), the instrument is stationary, whereas the background is not stationary due to breathing motion. For both the cases (Fig. 2(b) and Fig. 2(c)), the value of n has been set to 9. In the highlighted region of Fig. 2(c), structures have been severely blurred and noise patterns can also be observed, whereas in Fig. 2(b) such artefacts are not present. In fact, it appears as a higher dose image.

The quantitative analysis of the proposed denoising method in terms of CNR for the clinical sequences has been summarized in Table 1. For the sequences from clinical sites 1 and 3, the proposed method outperforms VBM3D even for the choice of $n = 7$ by a factor of 1.2 and 1.03, respectively. For the sequences from clinical site 2 alone, VBM3D performs better the proposed method by a factor of 1.36 for the nominal case ($n = 9$). But, on an average, the proposed method outperforms VBM3D for the nominal case.



(a) (b) (c)

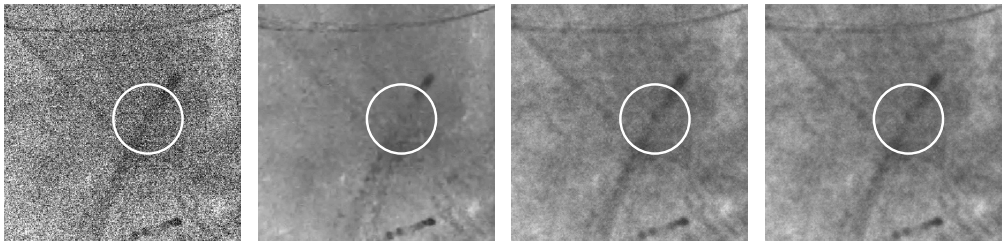
Fig. 2 (a) is an ROI of the input image, (b) and (c) are denoised versions using the proposed algorithm. In (b) noise characteristics have been used for patch matching, and in (c) noise characteristics have not been used for patch matching.

Table 1 Validation of the 20 Sequences acquired from the 3 clinical sites with respect to CNR. The CNR for each site is represented as Median ± Standard deviation

		Site 1	**Site 2**	**Site 3**	**Mean**
	Input	0.73 ± 0.12	0.29 ± 0.1	0.79 ± 0.1	0.60
Proposed	n=3	1.84 ± 0.27	0.76 ± 0.23	1.49 ± 0.11	1.37
	n=7	**2.67 ± 0.52**	1.12 ± 0.29	**1.82 ± 0.13**	1.91
	n=9	**2.8 ± 0.57**	1.24 ± 0.29	**1.86 ± 0.14**	**2.00**
	n=11	**2.88 ± 0.6**	1.31 ± 0.29	**1.88 ± 0.14**	**2.05**
	n=15	**2.95 ± 0.63**	1.39 ± 0.3	**1.9 ± 0.14**	**2.10**
	n=20	**2.99 ± 0.65**	1.43 ± 0.3	**1.91 ± 0.14**	**2.13**
	VBM3D	2.2 ± 0.32	**1.9 ± 0.24**	1.76 ± 0.04	1.93

By studying those cases where VBM3D has performed better than the proposed method in terms of CNR, we have found that VBM3D can lead to oversmoothed regions, i.e., a loss of texture creating an unnatural image appearance. An example can be seen in the ROI of Fig 3(b) (sequence 6 of site 2), where the body of the catheter is completely missing. Other regions around wires are affected by artefacts as well. Since the regions are very smooth, such artefacts may go unnoticed. On the other hand, the proposed method does not result in such a loss of structures (Fig. 3(c)). Even when a maximum of 20 patches are used for denoising, details, e.g., devices, are preserved in the case of the proposed method (Fig. 3(d)). A proper value for n has been found by analysing how the CNR values changed depending on the number of patches used (see Table 1). We have found that $n = 9$ provides the best trade-off between denoising performance and computational effort.

In Fig. 4, we show ROIs of several images before and after processing using the proposed method (Fig. 4(b)) and VBM3D (Fig. 4(c)). In the ROI in Fig. 4 (Col. 1), the denoising performance of the methods has been analysed for a sequence that has been used for the navigation of an ablation catheter inside a kidney. It can be observed that VBM3D produces images that can be characterized by flat local regions and sharp edges. This results in the "patchy" appearance of the image background. For instance, in Fig 4 (Col. 2), catheter shafts appear to be missing in some regions and in addition, the edges of the catheters appear to have undergone noticeable deformation. Unfortu-



(a) (b) (c) (d)

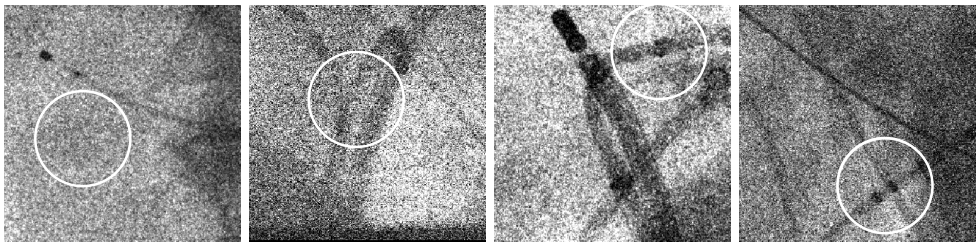
Fig. 3 (a) is an ROI of input sequence 6 from site 2. (b) is the ROI of VBM3D processed image. (c) and (d) are ROIs of the denoised images using the proposed method where $n = 9$ and $n = 20$, respectively. The circles highlight the region where the body of the catheter is missing in (b).

nately, such artefacts could mislead device detection algorithms further along an image processing chain. On the other hand, the proposed method produces smooth local regions with structures intact. In certain cases it may not reach the contrast that can be obtained with VBM3D as shown in Fig. 4(b) (Col. 3 and Col. 4). For a more comprehensive comparison, we have also obtained qualitative evaluations of the denoised scenes from twelve independent X-ray image quality experts (Fig. 5). The evaluations are in line with our quantitative evaluation. According to the experts, the denoising performance of the proposed method is superior to VBM3D. They have also found the proposed method to outperform VBM3D consistently in terms of producing artefact-free and realistically appearing images facilitating better catheter visibility.

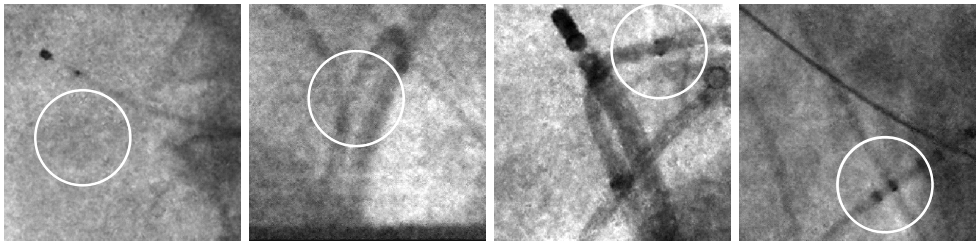
4 Discussion and Conclusion

In this paper, we have proposed a patch-based, blind spatio-temporal denoising method for ultra-low dose fluoroscopy sequences. The method includes the characteristics of noise to match patches, build a stack of temporally aligned patches and denoise them using low rank approximation. Most importantly, each patch can be denoised independently. This makes the algorithm highly parallelizable and a suitable candidate for the architecture of modern graphics processing units.

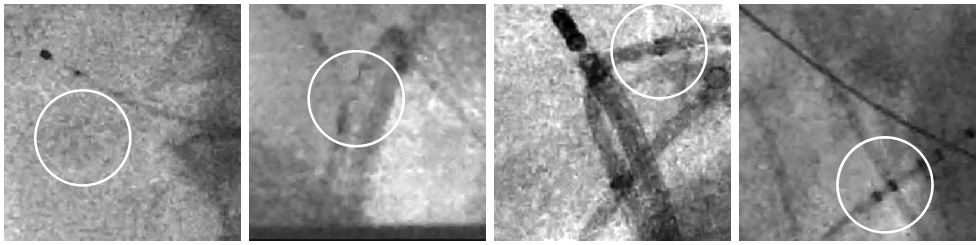
The state-of-the-art block matching methods [13,14] make use of a guide image (initial estimate) to tackle mismatches due to the presence of noise, i.e., they follow a two-step process, where an initial estimate is used to improve patch matching and the overall denoising performance. But at high noise levels, even this ad-hoc strategy of matching patches fails as the guide image retains noise patterns that resemble structures. These noise patterns in turn get preserved in the final denoised estimate in the form of artefacts. Since VBM3D has been designed to produce smooth images, severe thresholding of Haar coefficients may be carried out. This can lead to images where small regions are almost flat or where edges are no longer smooth but appear jagged.



(a) ROIs of input images



(b) ROIs of denoised images (proposed method)



(c) ROIs of denoised images (VBM3D)

Fig. 4 Each column shows the input as well as the results of the processing using the proposed method and the VBM3D. The shown ROIs are from (columns left to right) sequence 1 of site 3, sequence 8 of site 2, sequence 1 of site 1 and sequence 5 from site 2, respectively. The highlighted regions in columns 1 and 2 indicate places where VBM3D has introduced artefacts. The highlighted regions in columns 3 and 4 show places where the proposed method has slightly lower contrast compared to VBM3D.

Taking into account the noise properties not only helps in reducing such artefacts but also removes the necessity of computing a guide image. This may provide us with the opportunity to use our algorithm in a real-time environment. The evaluation on the EP sequences from clinical site 1 suggests that an improvement in the CNR by a factor of around 3.88 with respect to the input could be achieved. In this case, the proposed method also outperforms VBM3D in terms of CNR by a factor of 1.27 and also in terms of visual appearance. For sequences from clinical site 2, the improvement in the CNR of the proposed method (4.38) is comparatively lower than VBM3D (6.00). For the sequences from clinical site 3, the average improvement in the CNR is similar to the proposed method (2.33) and VBM3D (2.21). The steep improvement in the case of VBM3D for clinical site 2 could be due to the fact that VBM3D has been designed to provide smooth regions. Unfortunately, X-ray images are not composed of smooth regions as anatomical structures are not flat. This is

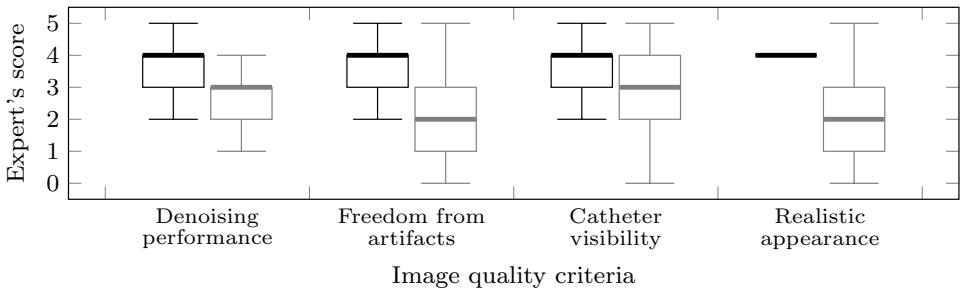


Fig. 5 Qualitative evaluation of the proposed method (black) and VBM3D (grey) based on the visual inspection of twelve X-ray image quality experts.

evident from the qualitative evaluation by X-ray image quality experts, who have found the improvement after denoising using the proposed method to be superior and artefact-free when compared to VBM3D. Most importantly, the experts have stated that the results resemble higher dose X-ray images. This suggests that the standard X-ray dose could be reduced further by a significant amount. As the proposed method depends mainly on the characteristics of noise, it is expected to be applicable to all ultra-low dose procedures.

Although, the proposed method ensures a significant reduction in noise, the images are not noise-free. One of the reasons for this could be a variation in the standard deviation of the expected noise $\sigma_{\eta'}$ that is obtained after applying a noise variance stabilization transform using the system parameters. The system parameters, system gain, electronic noise and offset, are usually obtained from calibrations. At ultra-low dose levels, it may be possible that the system does not behave as expected and consequently the calibrated values may vary. Another (usually minor) drawback is that the denoised images may suffer from mild blurring around instrument edges. This is due to the lack of perfect matches due to the 3D motion of objects. In such cases, the application of low rank approximation will result in the averaging of pixels that may not be perfectly matched. Our future work would be to target this issue by performing constrained low rank approximations.

Acknowledgement and Disclaimer: This work was supported by Siemens Healthcare GmbH. The concepts and results presented in this paper are based on research and not commercially available. *Ethical approval:* All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

References

1. Albarqouni, S., Baust, M., Conjeti, S., Al-Amoudi, A., Navab, N.: Multi-scale graph-based guided filter for de-noising cryo-electron tomographic data. In: BMVC, pp. 17–1 (2015)

2. Amiot, C., Girard, C., Chanussot, J., Pescatore, J., Desvignes, M.: Spatio-temporal multiscale denoising of fluoroscopic sequence. IEEE Trans Med Imaging **35**(6), 1565–1574 (2016)
3. Boracchi, G., Foi, A.: Multiframe raw-data denoising based on block-matching and 3-d filtering for low-light imaging and stabilization. In: Proc. Int. Workshop on Local and Non-Local Approx. in Image Processing, vol. 1, pp. 277–284 (2008)
4. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, vol. 2, pp. 60–65. IEEE (2005)
5. Cerciello, T., Bifulco, P., Cesarelli, M., Fratini, A.: A comparison of denoising methods for x-ray fluoroscopic images. Biomed Signal Process Control **7**(6), 550–559 (2012)
6. Chaudhury, K.N., Singer, A.: Non-local euclidean medians. IEEE Signal Process Lett **19**(11), 745–748 (2012)
7. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1256–1272 (2017)
8. Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. IEEE Trans. Image Process. **18**(10), 2221–2229 (2009)
9. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Bm3d image denoising with shape-adaptive principal component analysis. In: Signal Processing with Adaptive Sparse Structured Representations (2009)
10. Fritz, S.L., Mirvis, S.E., Osher Pais, S., Roys, S.: Phantom evaluation of angiographer performance using low frame rate acquisition fluoroscopy. Med Phys **15**(4), 600–603 (1988)
11. Goceri, E., Goksel, B., Elder, J.B., Puduvalli, V.K., Otero, J.J., Gurcan, M.N.: Quantitative validation of anti-ptbp1 antibody for diagnostic neuropathology use: Image analysis approach. Int J Numer Method Biomed Eng (2017)
12. Goodman, T.: Ionizing radiation effects and their risk to humans. Image wisely radiation safety in adult medical imaging. Disponível na internet a **22** (2010)
13. Guo, Q., Zhang, C., Zhang, Y., Liu, H.: An efficient svd-based method for image denoising. IEEE Trans. Circuits Syst. Video Technol. **26**(5), 868–880 (2016)
14. Kostadin, D., Alessandro, F., KAREN, E.: Video denoising by sparse 3d transform-domain collaborative filtering. In: Eur. Signal Process. Conf, vol. 149 (2007)
15. Lebrun, M., Leclaire, A.: An implementation and detailed analysis of the k-svd image denoising algorithm. Image Processing On Line **2**, 96–133 (2012)
16. Lee, M.S., Park, S.W., Lee, S.Y., Kang, M.G.: Motion-adaptive 3d nonlocal means filter based on stochastic distance for low-dose x-ray fluoroscopy. Biomed Signal Process Control **38**, 74–85 (2017)
17. Lefkimmiatis, S., Papandreou, G., Maragos, P.: Photon-limited image denoising by inference on multiscale models. In: Proc Int Conf Image Proc, pp. 2332–2335. IEEE (2008)
18. Luisier, F., Blu, T., Unser, M.: Sure-let for orthonormal wavelet-domain video denoising. IEEE Trans. Circuits Syst. Video Technol. **20**(6), 913–919 (2010)
19. Makitalo, M., Foi, A.: Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. IEEE Trans Image Process **22**(1), 91–103 (2013)
20. Parker, D.L., Clayton, P.D., Tarbox, L.R., VonBehren, P.L.: Optimal dose utilization with variable x-ray intensity in digital radiography. In: Application of Optical Instrumentation in Medicine XI, pp. 102–110. International Society for Optics and Photonics (1983)
21. Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, pp. 2774–2781 (2014)
22. Wagner, M., Yang, P., Schafer, S., Strother, C., Mistretta, C.: Noise reduction for curvelinear structures in real time fluoroscopy applications using directional binary masks. Med Phys **42**(8), 4645–4653 (2015)
23. Weickert, J., Scharr, H.: A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. J Vis Commun Image Represent **13**(1-2), 103–118 (2002)
24. Yang, K., Huang, S.Y., Packard, N.J., Boone, J.M.: Noise variance analysis using a flat panel x-ray detector: A method for additive noise assessment with application to breast ct applications. Med Phys **37**(7), 3527–3537 (2010)
25. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process (2017)

5.3. Preliminary results of DSA denoising based on a weighted low-rank approach using an advanced neurovascular replication system

5.3.1. Paper Summary

Digital subtraction angiography (DSA) is essential to assess the structure of blood vessels as well as the flow of blood through them in detail. Acquiring high-quality DSA images usually involves the use of undiluted contrast agent and high X-ray dose. Unfortunately, the high amount of iodine in the contrast agent puts a burden on the patients' kidneys, while the high X-ray dose may expose patients and medical staff to a considerable amount of radiation. Lowering the concentration of the contrast agent or the X-ray dose usually means sacrificing image quality. However, to retain the required image quality in such a situation, a spatio-temporal denoising approach has been proposed and published in [3].

Sequences acquired in the context of neurovascular DSA are usually composed of spatially aligned X-ray images. Since the background is stationary, the sequences are of low-rank and variation is introduced by the inflow and outflow of the contrast agent. Although there are several methods for denoising X-ray sequences, the availability of literature focusing specifically on the denoising of DSA sequences is limited. Moreover, the methods which make use of inaccurate noise models [190], fail to preserve the temporal dynamics of contrast flow [191], or appear ad-hoc with respect to the choice of thresholds [192].

To denoise a frame while taking into account the low-rank nature of the sequence, a constrained weighted rank-1 approximation of the stack comprising the frame to be denoised and its temporal neighbors is computed. The weights are used to compensate for the mismatches between the spatially aligned frames due to the inflow and outflow of the contrast agent. In other words, they are chosen to prevent mismatched pixels from contributing towards the low-rank approximation. The rank-1 approximation is performed by a weighted singular value decomposition. As there are several local optima for the singular vectors and values, a-priori information such as the smoothness of the approximants are used to perform constrained row- and column-wise approximations. The method has been evaluated using a vascular flow phantom emulating cranial arteries into which contrast agent can be manually injected (Vascular Simulations Replicator, Vascular Simulations, Stony Brook NY, USA). For the evaluation, image sequences acquired at different dose levels as well as different contrast agent concentrations have been used.

A quantitative analysis has suggested that the proposed method has resulted in significant improvement with respect to CNR and also that the dose as well as the concentration of the contrast agent could be reduced by about 75%. A visual analysis has indicated that the method yields images that share the characteristics of typical DSA images. The use of constrained low-rank approximations prevents the occurrence of smoothing artifacts that are introduced by patch-based denoising methods due to the presence of the flowing contrast agent. This is because the pulsatile motion of the contrast agent will lead to temporal mismatches. In the case of the proposed method, these temporal mismatches in the vessel regions will result in low temporal weights. As a consequence, denoising will focus on in-plane neighboring pixels. Thus, the texture of the flowing contrast agent will be retained, but the denoising performance in these regions may be lower.

Preliminary Results of DSA Denoising based on a weighted low-rank Approach using an advanced neurovascular Replication System

Sai Gokul Hariharan[1] · **Christian Kaethner**[2] · **Norbert Strobel**[2,3] · **Markus Kowarschik**[1,2] · **Julie DiNitto**[4] · **Shadi Albarqouni**[1] · **Rebecca Fahrig**[2,5] · **Nassir Navab**[1,6]

Abstract

Purpose. 2D digital subtraction angiography (DSA) has become an important technique for interventional neuroradiology tasks, such as detection and subsequent treatment of aneurysms. In order to provide high-quality DSA images, usually undiluted contrast agent and a high X-ray dose are used. The iodinated contrast agent puts a burden on the patients' kidneys while the use of high-dose X-rays expose both patients and medical staff to a considerable amount of radiation. Unfortunately, reducing either the X-ray dose or the contrast agent concentration usually results in a sacrifice of image quality.

Materials and Methods. To denoise a frame, the proposed spatio-temporal denoising method utilizes the low-rank nature of a spatially aligned temporal sequence where variation is introduced by the flow of contrast agent through a vessel tree of interest. That is, a constrained weighted rank-1 approximation of the stack comprising the frame to be denoised and its temporal neighbors is computed where the weights are used to prevent the contribution of non-similar pixels towards the low-rank approximation. The method has been evaluated using a vascular flow phantom emulating cranial arteries into which contrast agent can be manually injected (Vascular Simulations Replicator, Vascular Simulations, Stony Brook NY, United States). For the evaluation, image sequences acquired at different dose levels as well as different contrast agent concentrations have been used.

Results. Qualitative and quantitative analyses have shown that with the pro-

[1]Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2]Siemens Healthineers AG, Advanced Therapies, Forchheim, Germany
[3]Fakultät für Elektrotechnik, Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt, Schweinfurt, Germany
[4]Siemens Medical Solutions, Hoffman Estates, United States
[5]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[6]Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America

posed approach, the dose and the concentration of the contrast agent could both be reduced by about 75%, while maintaining the required image quality. Most importantly, it has been observed that the DSA images obtained using the proposed method have the closest resemblance to typical DSA images, i.e., they preserve the typical image characteristics best.
Conclusion. Using the proposed denoising approach, it is possible to improve the image quality of low-dose DSA images. This improvement could enable both a reduction in contrast agent and radiation dose when acquiring DSA images, thereby benefiting patients as well as clinicians. Since the resulting images are free from artifacts and as the inherent characteristics of the images are also preserved, the proposed method seems to be well suited for clinical images as well.

1 Introduction

Interventional X-ray guided imaging is becoming more and more important to support the treatment of increasingly sophisticated endovascular procedures. For some procedures, e.g., the coiling of aneurysms in the context of interventional neuroradiology, high-quality angiographic images, digital subtraction angiography (DSA) in particular, are essential to assess the structure of blood vessels as well as the blood flow through them in detail. In DSA, a digital subtraction of non-contrast-enhanced (mask image) from contrast-enhanced X-ray images (fill images) is performed to remove obstructing anatomical background information and obtain a clear view of the vasculature. In order to account for the exponential attenuation of the body, the digital subtraction is usually carried out after applying a logarithmic transformation on the mask and fill images, respectively.

As of today, acquiring high-quality DSA images usually involves the use of undiluted contrast agents and a high X-ray dose. The strong iodinated contrast agent puts a burden on the patients' kidneys, while the high X-ray dose may expose both patients and medical staff to a considerable amount of radiation. Unfortunately, a reduction in the X-ray dose or the contrast agent concentration results in a sacrifice of image quality—due to a drop in the contrast-to-noise-ratio (CNR). Fortunately, with the help of sophisticated denoising techniques, the image quality can be improved. This, in turn, could facilitate a reduction in X-ray dose and may make it possible to lower the concentration of the contrast agent used.

At low X-ray dose levels, the resulting images are corrupted by both signal-independent electronic noise and signal-dependent quantum noise. The problem of denoising images corrupted by signal-dependent noise is usually approached by either directly considering the statistics of the particular noise model [14, 15], or by using a variance stabilization transform (VST) to perform the denoising task in a more modular way [8]. With the latter approach, the signal-dependency is removed by rendering the noise variance constant

throughout the image [23]. This allows the reuse of any denoising technique designed for the removal of Gaussian noise.

Spatial denoising algorithms, designed for Gaussian noise, that are based on solving partial differential equations, such as diffusion-based filtering [20, 24], have made a sizeable impact on biomedical imaging mainly due to their performance at a low computational cost. Cerciello et al. [4], on the other hand, consider the well engineered patch-based method BM3D [11], as one of the best denoising algorithms for biomedical images when combined with a VST. Additional gains can be obtained by spatio-temporal methods that for example make use of wavelets [1,13], dictionary learning [7,12] and non-local self similarities between spatio-temporal neighborhoods of patches [3,8]. An important addition to the latter is a particularly well known spatio-temporal version of BM3D [11], known as VBM3D [10]. More recently, learning-based methods have been shown to deliver promising results [5,25,26]. Since these methods require pairs of spatially aligned noisy and noise-free images for training, they cannot be straightforwardly applied to real X-ray images. This is due to the non-availability of corresponding noisy and noise-free pairs of clinical images as patients cannot be irradiated twice to obtain them. In addition, acquiring exactly spatially aligned X-ray images at different dose levels may not be possible due to the motion introduced by the internal organs as well as breathing.

Although there are several methods for denoising X-ray sequences, the availability of literature that focuses specifically on the denoising of DSA sequences is surprisingly limited. Bugunović et al. [2] have presented two denoising approaches, one based on a 1D Wiener filtering of temporal data and another based on 3D wavelet denoising via a wavelet shrinkage technique. However, their assumption that noise in the high-dose images follows a Gaussian distribution, may not be appropriate for X-ray images [8]. Another interesting approach has been proposed by Niu et al. [19], who have taken advantage of the low-rank nature of DSA image sequences by applying a singular value decomposition on the matrix comprising vectorized digitally subtracted images and subsequently thresholding the singular values based on a predefined level. However, a drawback of this approach may be a compromise in the temporal resolution of image regions with high temporal dynamics due to the thresholding of a significant number of singular values in order to achieve a high denoising performance. The resulting images may be unsuitable for analyzing the flow of contrast, especially in regions with aneurysms. Menger et al. [18], on the other hand, have calculated individual cut-off thresholds for each pixel of a DSA series. Unfortuntaly, the choice of these thresholds are ad-hoc and not easy to replicate. Finally, using a sophisticated noise reduction algorithm, Söderman et al. [21] have suggested that the dose required to acquire DSA sequences can be reduced by 75% without the loss of image quality. However, they have not analyzed the impact of a reduction in the contrast agent concentration on the achieved image quality.

Sequences acquired in the context of neuro DSA are usually composed of spatially aligned X-ray images. Therefore, the sequences are of low-rank

and variation is introduced by the inflow and outflow of the iodinated contrast agent (consistent background with moving contrast agent). Since the denoising approach presented in [8] relies on accumulating and utilizing similar spatio-temporal patches to perform denoising, it is not well suited to DSA. This is because the inflow and outflow of the contrast agent changes the appearance of patches. More specifically, there will be some pixels that are already well opacified in some frames of the fill images, whereas in the other frames the corresponding pixels are either yet to be filled with contrast agent or are already clear of it. To tackle this problem, we have developed a denoising method that involves the application of an iterative weighted low-rank approximation to stacks of spatially aligned (full-resolution) images. The purpose of the weights is to prevent the contributions of temporally mismatched pixels towards the low-rank approximation. The approach results in denoised mask and fill images which, after subtraction result in DSA images with superior image quality.

2 Methodology

In X-ray based digital imaging techniques, the occurance of noise can have various sources, e.g., quantum noise, electronic noise and quantization noise. In addition to this, the detector gains may be switched based on the clinical imaging scenario. To handle these different factors, we have proposed a denoising approach that is related to the modular denoising strategy used in [8].

For denoising a frame (single image) $z_f \in \mathbb{R}^{N\times M}$ (with N rows and M columns at frame number f) in a sequence $\boldsymbol{z} \in \mathbb{R}^{N\times M\times F}$ of spatially aligned F frames, we have made use of z_f and its $T-1$ neighboring frames (T frames in total). Since DSA images are viewed and analyzed only after the entire sequence has been acquired, the previous as well as future frames (associated with the current frame) can be used to denoise the current frame. To this end, the T frames have been combined into a 3D matrix $\boldsymbol{y_f} \in \mathbb{R}^{N\times M\times T}$. The matrix $\boldsymbol{y_f}$ comprises z_f and its neighboring $T-1$ frames $z_{f'}$, where f' represents the indices of the neighboring frames of z_f, t represents the current temporal location of z_f in $\boldsymbol{y_f}$ ($y_{f,t}$) and t' represents the temporal locations of $z_{f'}$ in $\boldsymbol{y_f}$ ($y_{f,t'}$) (Fig. 1a). A schematic overview of the complete denoising workflow is presented in Fig. 1. After setting up an input stack of X-ray frames around a particular image, we have stabilized the noise variance across the frames in $\boldsymbol{y_f}$ using the generalized Anscombe transform (GAT) [16] to obtain $\boldsymbol{Y_f} \in \mathbb{R}^{N\times M\times T}$ (Fig. 1b). In the second step, we have computed pixel-wise weights $\boldsymbol{W_f} \in \mathbb{R}^{N\times M\times T}$ for all the frames in $\boldsymbol{Y_f}$ with respect to the current frame $Y_{f,t}$ (Fig. 1c). After that, we have applied a row-wise spatio-temporal denoising to $\boldsymbol{Y_f}$ using an iterative weighted low-rank approximation to obtain the denoised current frame in the GAT domain $\tilde{Y}_{f,t}$ (Fig. 1d). Similarly, stacks corresponding to the neighboring frames $\boldsymbol{Y_{f'}}$ are denoised row-wise to obtain the denoised neighboring frames in the GAT domain $\tilde{Y}_{f',t}$ that are then used to construct the denoised stack $\tilde{\boldsymbol{Y}}_{\boldsymbol{f}}$ (Fig. 1e). Then, we have performed column-wise spatio-temporal denoising to the denoised stack $\tilde{\boldsymbol{Y}}_{\boldsymbol{f}}$ using the

iterative weighted low-rank approximation (Fig. 1f) to obtain the denoised current frame $\hat{Y}_{f,t}$. Finally, we have applied an inverse noise VST to return to the X-ray image domain $\hat{z}_f$ (also represented as $\hat{y}_{f,t}$ in $\hat{\boldsymbol{y}}_{\boldsymbol{f}}$) (Fig. 1g). The main elements of the algorithm are described in detail in the subsequent sections.

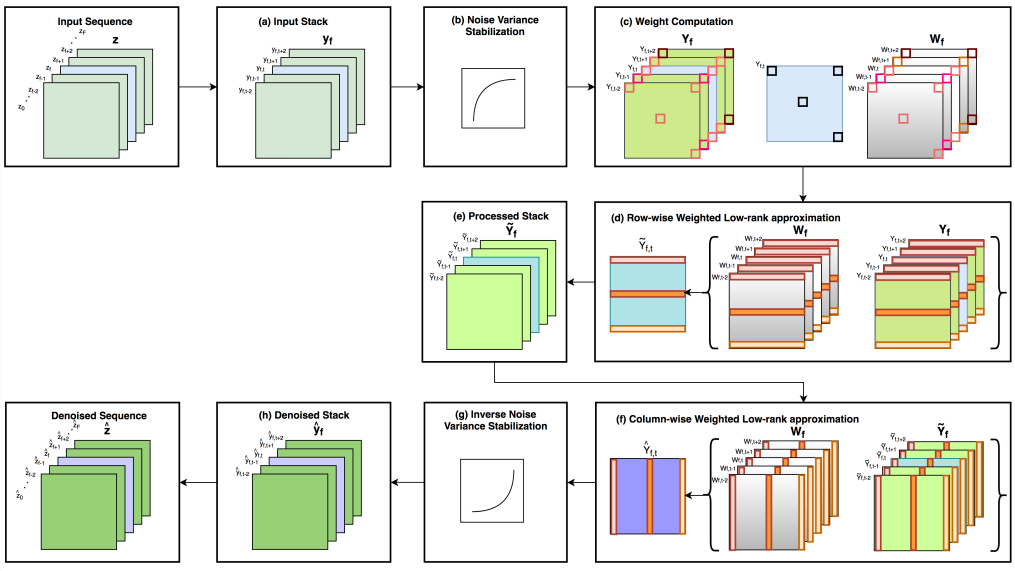


Fig. 1 Schematic workflow visualization of the proposed denoising approach using a weighted low-rank approximation.

2.1 Imaging Model and Noise Variance Stabilization

We have made use of the imaging model mentioned in [9], where X-ray images are assumed to be corrupted by signal-dependent quantum noise and signal-independent electronic noise. The model can be represented by:

$$z[r,c] = \alpha \cdot x[r,c] + g + \eta[r,c], \tag{1}$$

where α is the system gain, g is the system offset, and $z[r,c]$, $x[r,c]$ and $\eta[r,c]$ represent the observed gray value, quanta and electronic noise at the particular location $[r,c]$, respectively. The quantum noise associated with x can be modeled using a Poisson distribution and the overall electronic noise η can be modeled using a Gaussian distribution with zero mean and a specific standard deviation σ_n. The variance of mixed noise σ_z^2 can be expressed as

$$\sigma_z^2[r,c] = \alpha^2 \cdot \sigma_x^2[r,c] + \sigma_n^2. \tag{2}$$

By taking the system parameters into account, the noise variance of the mixed signal-dependent noise can be stabilized to a constant $\sigma_{\eta'}^2$ using the generalized Anscombe transform (GAT) [23]. More formally, as shown in Fig. 1b, we have applied the GAT on $\boldsymbol{y_f}$ to obtain $\boldsymbol{Y_f}$.

2.2 Weight Computation

In order to account for potential mismatches among pixels in successive frames $Y_{f,t'}$ in $\boldsymbol{Y_f}$ when compared to the current frame $Y_{f,t}$, weights are utilized as a part of the low-rank approximation. When the compared pixels are dissimilar, low weights are assigned. Accordingly, highly similar pixels are assigned high weights. The weights for frame $Y_{f,t'}$ at the location $[r, c]$ are derived by

$$w_{f,t',r,c} = e^{-|w_{\mu_{f,t',r,c}} \times w_{s_{f,t',r,c}}|} \times e^{-|w_{\sigma_{f,t',r,c}} \times w_{s_{f,t',r,c}}|} \tag{3}$$

and are based on the mean $\mu_{d_{f,t',r,c}}$, the standard deviation $\sigma_{d_{f,t',r,c}}$ and the smoothing factor $w_{s_{f,t',r,c}}$ associated with the difference patch

$$d_{f,t',r,c} = Y_{f,t',r,c} - Y_{f,t,r,c} \tag{4}$$

that is computed from the $k \times k$ spatial neighborhoods around $[r, c]$ in the current frame $Y_{f,t,r,c}$ and neighboring frames $Y_{f,t',r,c}$. The different components are computed as follows:

$$w_{\mu_{f,t',r,c}} = \mu_{d_{f,t',r,c}}, \tag{5}$$

$$w_{\sigma_{f,t',r,c}} = \mid \sigma_{d_{f,t',r,c}} - \sqrt{2} \times \sigma_{\eta'} \mid, \tag{6}$$

$$w_{s_{f,t',r,c}} = \frac{\max(\boldsymbol{Y})^2}{\mu_{Y_{f,t',r,c}} \times \mu_{Y_{f,t,r,c}}}. \tag{7}$$

In an ideal case, where there is a perfect match, $d_{f,t',r,c}$ should contain noise alone. Therefore, the standard deviation $\sigma_{d_{f,t',r,c}}$ and the mean $\mu_{d_{f,t',r,c}}$ of the difference patch $d_{f,t',r,c}$ should approximately be equal to $\sqrt{2} \times \sigma_{\eta'}$ and 0, respectively. Since the gray values associated with contrast agent are low, the smoothing factor $w_{s_{f,t',r,c}}$ contributing to such pixels should be low in order to have minimal temporal averaging in the case of a mismatch. Finally, when either $w_{\sigma_{f,t',r,c}}$ or $w_{\mu_{f,t',r,c}}$ in Eq. 3 is high, then we consider this a mismatch and the assigned weight will be low and vice versa.

2.3 Weighted Low-rank Approximation

In data samples with missing or corrupted entries, the use of non-uniform weights is required to account for any difference in the samples [22].

Given a data matrix $A \in \mathbb{R}^{M \times T}$ constructed from T data samples of size $M \times 1$ and a weight matrix $W \in \mathbb{R}^{M \times T}$ containing non negative weights for each of the $M \times T$ elements, the best rank R approximation of the data matrix can be described by a weighted singular value decomposition given by

$$[U, \Sigma, V] = \min_{U,\Sigma,V} ||A - U\Sigma V^*||_W^2, \tag{8}$$

where $U \in \mathbb{R}^{M \times R}$ and $V \in \mathbb{R}^{T \times R}$ are the rank R singular matrices, $\Sigma \in \mathbb{R}^{R \times R}$ is a diagonal matrix containing singular values and $||\ ||_W$ represents the

weighted L_2 norm with respect to W [17]. As there are several local optima for U, V and Σ in Eq. 8, Das et al. [6] have proposed the use of prior information about the data or the approximants. For example, in the case of images, the smoothness of the approximants can usually be assumed and this leads to

$$[U, \Sigma, V] = \min_{U,\Sigma,V} \left\{ ||A - U\Sigma V^*||_W^2 + \frac{\alpha_u}{2}||B_u U||_F^2 + \frac{\alpha_v}{2}||B_v V||_F^2 \right\}, \quad (9)$$

where B_u and B_v are linear operators set to second order finite difference matrices for reflecting a-priori information, α_u and α_v are regularizers and $||\ ||_F$ represents the Frobenius norm. In our case, since the frames in $\boldsymbol{Y_f}$ are spatially aligned, $\boldsymbol{Y_f}$ is expected to have a low-rank where variation is introduced by the flowing contrast agent. By using the weights $\boldsymbol{W_f}$ to account for noise as well as the mismatches introduced due to the inflow and outflow of contrast agent, the denoising of $\boldsymbol{Y_f}$ can be performed with the help of an iterative weighted rank-1 approximation. Unfortunately, due to the large matrix sizes of the X-ray images, it is computationally expensive to compute a weighted rank-1 approximation of $\boldsymbol{Y_f}$. Therefore, we have approximated the weighted rank-1 approximation of a 3D matrix with two 2D rank-1 approximations applied row-wise and column-wise. At the outset, the data matrix and the weight matrix for performing the weighted rank-1 approximation row-wise on $\boldsymbol{Y_f}$ are built. For each row r, the data matrix $A_{f,r} \in \mathbb{R}^{M \times T}$ is obtained by taking a horizontal slice of $\boldsymbol{Y_f}$ at the row location r (represented as stripes in Fig. 1d). The corresponding weight matrix $W_{f,r}$ is computed from $\boldsymbol{W_f}$ accordingly. Subsequently, based on an assumption that the singular vectors should be smooth for an image, the weighted-1 approximation of $A_{f,r}$ is solved by

$$[U_{f,r}, \Sigma_{f,r}, V_{f,r}] = \min_{U_{f,r},\Sigma_{f,r},V_{f,r}} \Big\{ ||A_{f,r} - U_{f,r}\Sigma_{f,r}V_{f,r}^*||_{W_{f,r}}^2 + \dots \\ \dots \frac{\alpha_u}{2}||B_u U_{f,r}||_F^2 + \frac{\alpha_v}{2}||B_v V_{f,r}||_F^2 \Big\}. \quad (10)$$

This problem involves an iteratively performed alternating optimization to estimate the left and the right approximants [6]. Once the row-wise denoised estimates for the entire sequence are available (Fig. 1e), a column-wise weighted rank-1 approximation of the frames of $\hat{Y}$ is computed—similar to the above mentioned row-wise rank-1 approximation—using the associated data and weight matrices (Fig. 1f). After the final denoised estimate $\hat{Y}$ is computed, an inverse GAT is applied to return to the image domain (Fig. 1h).

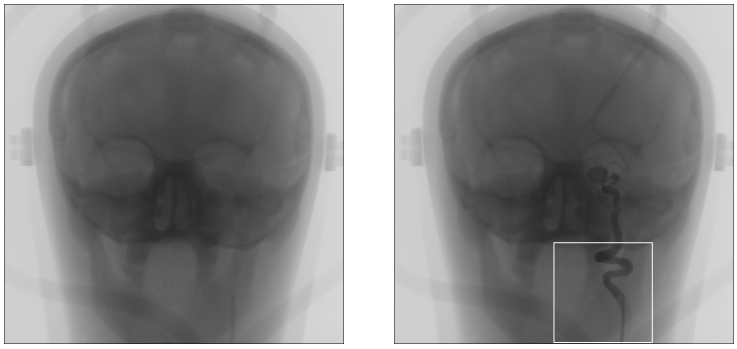
2.4 Denoising of DSA Images

Instead of denoising a DSA sequence as proposed in [18,19], the mask frame(s) and fill frames of a sequence are first denoised independently using the stacks

of frames generated specifically for each of the mask and fill images, respectively. Subsequently, the DSA images are computed by subtracting the logarithm transformed pixels of the denoised mask frame from the corresponding logarithm transformed pixels of the denoised fill frames. As all frames of the X-ray sequence (including frames free from contrast agent) are denoised, the mask frame is chosen from one of the denoised low-dose mask frames (frames without contrast agent). This avoids the necessity of acquiring a high-dose mask image. In fact, the use of a low-dose mask instead of a high-dose mask might even give superior results as the spatio-temporal denoising would result in similar noise characteristics.

3 Material

In order to validate the proposed method, we have used X-ray images of a vascular flow phantom emulating cranial arteries (Fig. 2) into which contrast agent can be manually injected (Vascular Simulations Replicator, Vascular Simulations, Stony Brook NY, United States). We have obtained X-ray sequences at 3 different concentration levels of the contrast agent (25%, 50% and 100% of the standard contrast agent concentration) and 4 different dose levels (12%, 18%, 27%, 100% of the standard dose level). We have compared our results quantitatively and qualitatively against the well established denoising method VBM3D [10] (in the noise variance stabilized domain [16]) and a recently published denoising method based on a low-rank approximation (LRA) that has been designed for live fluoroscopic images [8]. Since the use of 9 spatio-temporal patches of size 9×9, for computing the denoised image, has been found to offer a good trade-off between denoising performance and computational effort [8], we have kept this choice for denoising the mask and the fill images, respectively. In addition, we have compared the denoised results with the images acquired using standard settings (100% dose level and 100% contrast agent concentration).

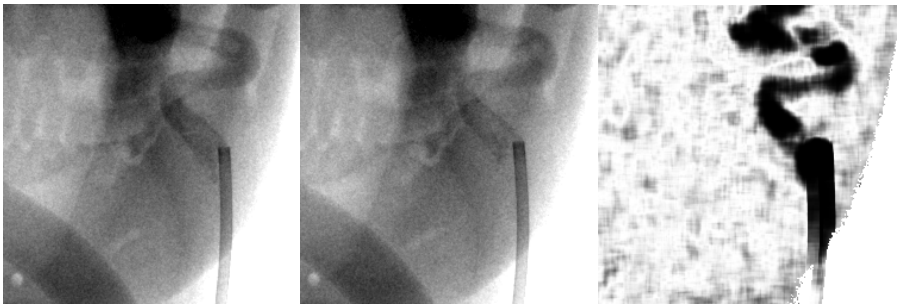


(a) Mask image (b) Fill image

Fig. 2 Visualization of a mask and a fill image acquired at standard dose level. The highlighted region in (b) has been used in Fig. 3 for visualizing the weights computed between two successive fill frames.

4 Results

In order to highlight the impact and the sensitivity of the weights used in the proposed denoising algorithm (introduced in Eq. 3), the pixel-wise weights computed from two subsequent example frames of the acquired X-ray sequence are shown in Fig. 3. They have been computed for the frame shown in Fig. 3a with respect to the one displayed in Fig. 3b. It can be observed that even small differences in the gray values between the corresponding pixels have been captured by the weights visualized in Fig. 3c.



(a) (b) (c)

Fig. 3 Visualization of the computed weights (c) for a region of interest (ROI) (a) with respect to ROI (b). The two frames to which (a) and (b) belong to have been acquired at different points in time. Dark and bright colors represent low and high weights, respectively.

For the proposed method, the choice of $k = 13$ when calculating weights among $k \times k$ spatial neighborhoods results in a reliable comparison of the statistical properties of associated patches and hence results in accurate weights. A smaller value for k would result in inconsistencies in the computed estimates. To obtain singular vectors with a similar smoothness in the spatial as well as the temporal direction, setting B_u and B_v to second order finite difference operators with the same accuracy is preferable. Therefore, they have been set to the maximum number of frames used for denoising T. Finally, setting both α_u and α_v to 0.0001 has been found to be a suitable choice for reducing the noise level while at the same time retaining the sharpness of the image. Therefore, these parameters have been used for further computations. In Table 1, a quantitative analysis with respect to the average CNR for all DSA sequences acquired at different dose levels and contrast agent concentrations is presented. It can be observed from Table 1 that for the proposed method, the average CNR increases with an increase in T and is comparable with those of VBM3D and LRA for $T = 9$. In fact, the proposed method outperforms the compared methods for $T = 10$ with respect to average CNR.

As the goal is to study the possibility of a reduction in the dose level and contrast agent concentration, the results computed for the 12% dose level and the 25% contrast agent concentration have been visualized in Fig. 4. In accordance to the quantitative results shown in Table 1, a progressive improvement

Table 1 Quantitative analysis with respect to average CNR for DSA sequences acquired using different dose levels and contrast agent concentrations and processed using VBM3D [10], LRA [8] and the proposed method for different values of T, respectively. The best values are highlighted using bold face.

Contrast agent concentration: 25%

Dose	Input	Proposed (T)						VBM3D	LRA
		2	4	6	8	9	10		
12%	5.51	10.30	14.37	16.97	19.13	20.63	21.25	**22.43**	20.76
18%	7.29	12.16	17.05	20.15	22.82	24.30	25.70	**26.40**	24.76
27%	16.41	29.26	40.15	48.76	57.29	60.71	65.55	**86.81**	69.62

Contrast agent concentration: 50%

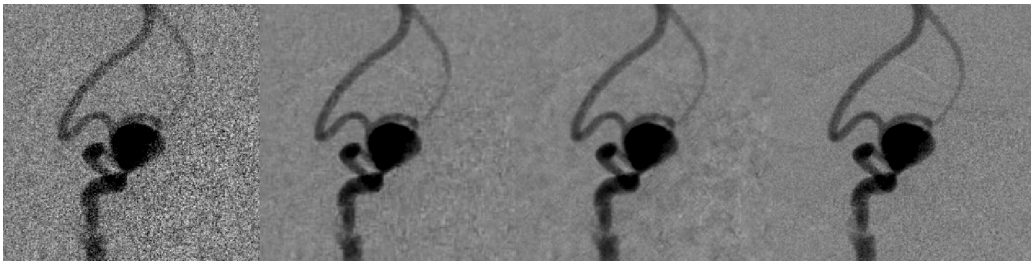
Dose	Input	Proposed (T)						VBM3D	LRA
		2	4	6	8	9	10		
12%	11.43	20.59	29.67	36.67	42.56	45.07	47.71	47.18	**49.45**
18%	15.56	27.77	40.29	**48.63**	**54.94**	**57.58**	**60.22**	46.40	45.47
27%	20.44	34.68	47.07	56.98	67.73	72.01	75.48	**82.57**	61.90

Contrast agent concentration: 100%

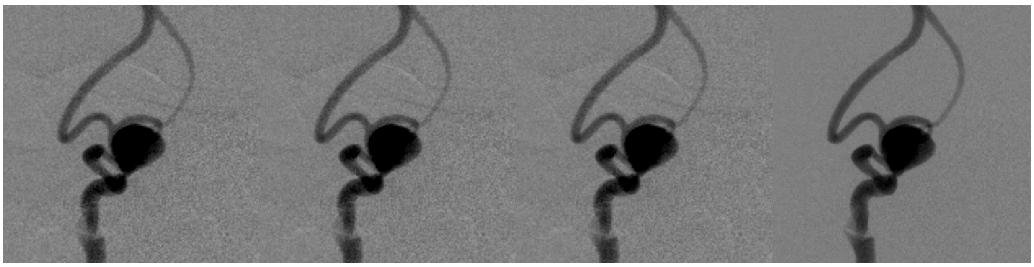
Dose	Input	Proposed (T)						VBM3D	LRA
		2	4	6	8	9	10		
12%	15.32	28.87	41.51	51.57	59.12	**62.37**	**65.62**	59.13	61.62
18%	19.29	33.37	47.11	57.40	66.46	**70.08**	**75.33**	68.86	64.64
27%	23.76	44.54	61.96	75.56	**84.48**	**88.29**	**94.06**	84.07	82.69

in the image quality can be seen for the proposed method with an increase in T (Fig. 4d-g). However, no significant differences in terms of the image impression can be observed for $T > 8$. A comparison to an unprocessed reference ROI acquired with 100% of the standard dose level and 100% of the contrast agent concentration (Fig. 4h) indicates that the proposed method results in realistic images, especially with respect to the underlying noise pattern, whereas, VBM3D (Fig. 4b) and LRA (Fig. 4c) fail to do this even though their average CNR was higher than that of the proposed method. It should be noted that the width of the gray level window used to display Fig. 4a-g has been set to around 25% of that used for Fig. 4h in order to make the images visually comparable. Although a significant improvement in the CNR up to 21.25 (for $T = 10$) can be obtained using the proposed method, the resulting images (Fig. 4g) have a somewhat higher amount of noise and a lower CNR than the images acquired at 100% dose level with 100% contrast agent concentration (CNR: 31.28) (Fig. 4h).

In Fig. 5, we have presented the results of another low-dose and low contrast agent concentration setting - 27% dose level and 25% contrast agent concentration. The gray level windowing has been set similar to Fig. 4. In the case of VBM3D (Fig. 5b), there is an improvement with respect to background artifact reduction when compared to Fig. 4b. However, this improvement is still insufficient to match the appearance of the reference image acquired using the standard settings (Fig. 5h). From Fig. 5b and Fig. 5c, it can be seen that the results of VBM3D and LRA contain artifacts and moreover there is a sig-



(a) Unprocessed (b) VBM3D (c) LRA (d) Proposed $T = 6$



(e) Proposed $T = 8$ (f) Proposed $T = 9$ (g) Proposed $T = 10$ (h) Reference

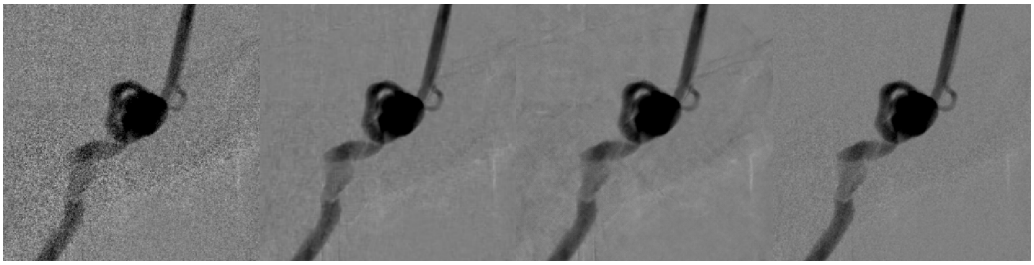
Fig. 4 Visualization of an (a) unprocessed ROI acquired using 12% of the standard dose with 25% contrast agent concentration and the corresponding denoised ROIs obtained using (b) VBM3D [10], (c) LRA [8] and (d-g) the proposed method for $T = \{6, 8, 9, 10\}$ along with an (h) unprocessed reference ROI acquired at 100% dose level with 100% contrast agent concentration.

nificant difference in the noise characteristics when compared to the reference image (Fig. 5h). The use of the proposed method has resulted in DSA images (Fig. 5g) with the highest resemblance to the DSA images acquired using standard settings (Fig. 5h). From a quantitative perspective, the average CNR of the processed (27% dose level and 25% contrast agent concentration) sequence (CNR: 65.55, $T = 10$) is a little lower than that of the sequence acquired using the standard setting (CNR: 72.70).

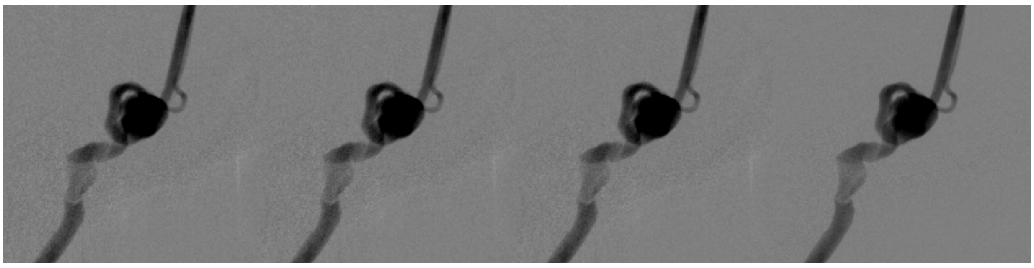
For a more comprehensive comparison, we have also obtained qualitative evaluations of the denoised scenes from eleven independent X-ray image quality experts (Fig. 6). The evaluations are in line with our quantitative evaluation. According to the experts, all the methods preserve the flow of the contrast agent. They have also found that the proposed method outperforms the other methods in terms of denoising performance as well as producing artifact free and realistically appearing DSA images.

5 Discussion and Conclusion

We have proposed a novel method to perform a spatio-temporal denoising of DSA sequences based on a weighted low-rank approximation. Since the method utilizes photons received for acquiring past and future frames in a noise variance stabilized domain by carefully choosing the pixels that con-



(a) Unprocessed (b) VBM3D (c) LRA (d) Proposed $T = 6$



(e) Proposed $T = 8$ (f) Proposed $T = 9$ (g) Proposed $T = 10$ (h) Reference

Fig. 5 Visualization of an (a) unprocessed ROI acquired using 27% of the standard dose with 25% contrast agent concentration and the corresponding denoised ROIs obtained using (b) VBM3D [10], (c) LRA [8] and (d-g) the proposed method for $T = \{6, 8, 9, 10\}$ along with an (h) unprocessed reference ROI acquired at 100% dose level with 100% contrast agent concentration.

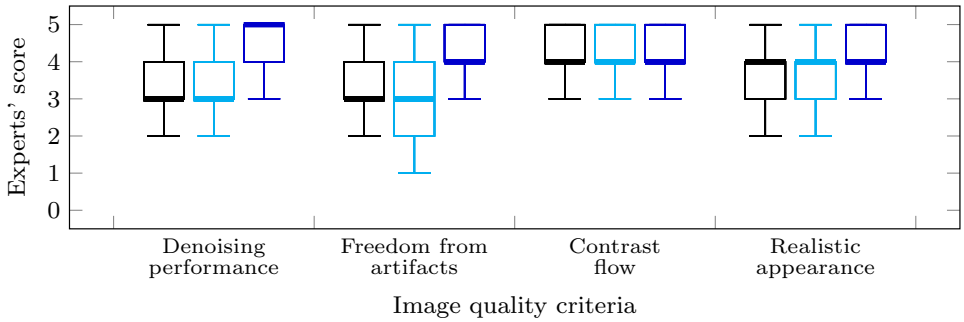


Fig. 6 Qualitative evaluation of VBM3D [10] (black), LRA [8] (cyan) and the proposed method (blue) for $T = 9$ based on the visual inspection of the processed sequences by eleven independent X-ray image quality experts. The thick lines inside the boxes represent the median values, the range of the boxes represent the middle 50% of the values, and the whiskers represent the ranges for the bottom 25% and the top 25% of the data values, excluding outliers. Due to the distributions of the data values, both the whiskers are not present for some of the cases.

tribute towards the denoising, the appearance of the resulting processed images resembles the appearance of the images acquired at a higher-dose setting. We have also analyzed the possibility of a reduction in the X-ray dose as well as the concentration of the contrast agent based on a study involving a vascular flow phantom (Vascular Simulations Replicator, Vascular Simulations, Stony Brook NY, United States).

A quantitative analysis with respect to CNR for $T = 9$—a good trade-off between performance and and computational time [8]—shows that VBM3D [10], LRA [8] and the proposed method result in a similar overall improvement by a factor of 3.89, 3.61 and 3.71, respectively. However, a visual analysis suggests that only the use of the proposed method yields images that share the characteristics of typical DSA images. This is because, the denoising methods proposed in [10] and [8] modify the properties of the underlying noise. In the case of [10], artifacts that resemble flat structures are introduced at very low-dose levels, e.g., 12% of the standard dose level. This could be a consequence of not taking the noise characteristics into account when finding similar patches that are used for denoising. Even though the method proposed in [8] makes use of an X-ray imaging model while matching patches, it may be difficult to obtain well matched patches due to the inflow and outflow of the contrast agent. Moreover, the characteristics of noise in the denoised image may have changed due to the different steps, such as the low-rank approximation of stacked patches and the aggregation of the denoised patches. This effect gets enhanced when performing the digital subtraction. Since the proposed method makes use of constrained row-wise and column-wise low-rank approximations, it does not suffer from the smoothing artifacts introduced by a non-constrained low-rank approximation of misaligned patches. This, in turn, reduces blurring artifacts that are otherwise visible in patch-based denoising approaches. It should also be noted that in the case of pulsatile motion of the flowing contrast agent, the temporal mismatches in the vessel regions will result in low temporal weights. As a consequence, denoising will be based on in-plane neighboring pixels. In such cases, the texture of the flowing contrast agent will be retained, but the denoising performance may be affected.

Finally, a visual analysis suggests that by denoising using the proposed method, it may be possible to enhance the appearance of DSA images acquired at 27% of the standard dose level and 25% of the standard contrast agent concentration and make them similar to standard DSA images acquired at 100% dose level and 100% contrast agent concentration. In fact, both the qualitative and the quantitative analyses of the phantom study suggest that a reduction in the dose as well as the contrast agent concentration by about 75% may be possible. In our future work, we will evaluate the performance of the proposed algorithm on clinical X-ray sequences and also focus on improving the performance of the proposed approach by combining it with a DSA specific motion compensation algorithm.

Acknowledgement, Disclaimer and Conflict of Interest: This work was supported by Siemens Healthineers AG. The concepts and results presented in this paper are based on research and are not commercially available. The authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Amiot, C., Girard, C., Chanussot, J., Pescatore, J., Desvignes, M.: Spatio-temporal multiscale denoising of fluoroscopic sequence. IEEE Trans Med Imaging **35**(6), 1565–1574 (2016)
2. Bogunović, H., Lončarić, S.: Denoising of Time-Density Data in Digital Subtraction Angiography, pp. 1157–1166. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
3. Buades, A., Coll, B., Morel, J.M.: Nonlocal image and movie denoising. Int. J. Comput. Vision **76**(2), 123–139 (2008). DOI 10.1007/s11263-007-0052-1. URL http://dx.doi.org/10.1007/s11263-007-0052-1
4. Cerciello, T., Bifulco, P., Cesarelli, M., Fratini, A.: A comparison of denoising methods for x-ray fluoroscopic images. Biomed Signal Process Control **7**(6), 550–559 (2012)
5. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1256–1272 (2017)
6. Dasa, S., Neumaiera, A.: Fast regularized low rank approximation of weighted data sets. www.mat.univie.ac.at/%7Eneum/software/birsvd/svd%5Fincomplete%5Fdata.pdf (2011)
7. Deng, X., Liu, Z.: Image denoising based on steepest descent OMP and K-SVD. In: Signal Processing, Communications and Computing (ICSPCC), 2015 IEEE International Conference on, pp. 1–5 (2015). DOI 10.1109/ICSPCC.2015.7338821
8. Hariharan, S.G., Strobel, N., Kaethner, C., Kowarschik, M., Demirci, S., Albarqouni, S., Fahrig, R., Navab, N.: A photon recycling approach to the denoising of ultra-low dose x-ray sequences. International journal of computer assisted radiology and surgery **13**(6), 847–854 (2018)
9. Hariharan, S.G., Strobel, N., Kowarschik, M., Fahrig, R., Navab, N.: Simulation of realistic low dose fluoroscopic images from their high dose counterparts. In: Bildverarbeitung für die Medizin 2018, pp. 80–85. Springer (2018)
10. Kostadin, D., Alessandro, F., KAREN, E.: Video denoising by sparse 3d transform-domain collaborative filtering. In: European Signal Processing Conference, vol. 149. Tampere, Finland (2007)
11. Lebrun, M.: An analysis and implementation of the BM3D image denoising method. Image Processing On Line **2**, 175–213 (2012). DOI 10.5201/ipol.2012.l-bm3d
12. Lebrun, M., Leclaire, A.: An implementation and detailed analysis of the K-SVD image denoising algorithm. Image Processing On Line **2**, 96–133 (2012). DOI 10.5201/ipol.2012.llm-ksvd
13. Luisier, F., Blu, T., Unser, M.: SURE-LET for orthonormal wavelet-domain video denoising. Circuits and Systems for Video Technology, IEEE Transactions on **20**(6), 913–919 (2010). DOI 10.1109/TCSVT.2010.2045819
14. Luisier, F., Blu, T., Unser, M.: Image denoising in mixed poisson–gaussian noise. IEEE Transactions on image processing **20**(3), 696–708 (2011)
15. Luisier, F., Vonesch, C., Blu, T., Unser, M.: Fast interscale wavelet denoising of poisson-corrupted images. Signal Processing **90**(2), 415–427 (2010)
16. Makitalo, M., Foi, A.: Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. IEEE transactions on image processing **22**(1), 91–103 (2013)
17. Markovsky, I., Niranjan, M.: Approximate low-rank factorization with structured factors. Computational Statistics & Data Analysis **54**(12), 3411–3420 (2010)
18. Menger, N., Elsässer, T., Chen, G.H., Manhart, M.: Noise reduction in low dose dsa imaging using pixel adaptive svd-based approach. In: Bildverarbeitung für die Medizin 2017, pp. 37–42. Springer (2017)
19. Niu, K., Li, Y., Schafer, S., Royalty, K., Wu, Y., Strother, C., Chen, G.H.: Ultra low radiation dose digital subtraction angiography (dsa) imaging using low rank constraint. In: SPIE Medical Imaging, pp. 94,121O–94,121O. International Society for Optics and Photonics (2015)
20. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence **12**(7), 629–639 (1990)
21. Söderman, M., Holmin, S., Andersson, T., Palmgren, C., Babić, D., Hoornaert, B.: Image noise reduction algorithm for digital subtraction angiography: clinical results. Radiology **269**(2), 553–560 (2013)

22. Srebro, N., Jaakkola, T.: Weighted low-rank approximations. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 720–727 (2003)
23. Starck, J.L., Murtagh, F.D., Bijaoui, A.: Image processing and data analysis: the multiscale approach. Cambridge University Press (1998)
24. Weickert, J., Scharr, H.: A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. J Vis Commun Image Represent **13**(1-2), 103–118 (2002)
25. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Generative adversarial networks for noise reduction in low-dose ct. IEEE Trans Med Imaging **36**(12), 2536–2545 (2017)
26. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process (2017)

5.4. Summary

In this chapter, two analytical methods for the spatio-temporal denoising of low-dose X-ray images have been presented. The methods use the imaging model presented in Chapter 4, especially the noise characteristics, to identify similar pixels and patches while computing a low-rank approximation. Quantitative and qualitative analysis indicate that the results obtained are well denoised and appear more realistic. Although these results are promising, applying them in real-time would involve cost-performance trade-offs as they rely on solving a complex optimization problem. Therefore, in Chapter 6, advanced techniques that are computationally more efficient and facilitate real-time processing – while facilitating similar performance – are presented.

6. Learning-based Denoising of X-ray Images

6.1. Introduction

Denoising is considered to be a suitable technique to reduce the X-ray dose level while maintaining the required image quality. Though analytical denoising approaches provide suitable results, they involve cost-performance trade-offs during the application time – since they rely on solving complex optimization problems. In this chapter, computationally inexpensive learning-based denoising algorithms that involve the use of model-based X-ray simulations during the training phase, are presented and thoroughly evaluated.

6.2. An analytical approach for the simulation of realistic low-dose fluoroscopic images

6.2.1. Paper Summary

Learning-based denoising algorithms have been shown to outperform analytical methods. These approaches often rely on spatially aligned pairs of noisy and noise-free images. However, in the context of clinical X-ray imaging, it is not possible to acquire noise-free images during clinical interventions. In clinical practice, it is at the most possible to find X-ray images taken at a higher dose in cases where superior image quality has been required. Unfortunately, the spatially aligned low-dose counterparts are usually very hard to get due to organ motion and patient motion, e.g., caused by breathing. Most importantly, as per the ALARA (as low as reasonably achievable) principle, patients must not be exposed to unnecessary radiation. A possibility to address this issue is realistic noise simulation applied to higher dose images to generate lower dose counterparts. The contribution of this paper is a novel noise simulation approach based on an X-ray image formation model to generate low-dose images from their higher dose counterparts [4].

In the past, researchers have tried adding noise to high-dose images to simulate low-dose images based on the assumption that the sequences are corrupted by either Poisson noise or Gaussian noise or both [193, 194, 195, 196]. Unfortunately, they have usually relied on image formation models that may not fully account for how X-ray images are acquired in practice. Researchers have also taken into account the influence of system parameters on the noise level through extensive calibrations while simulating noise [197, 198]. However, as these methods consider the noise-corrupted gray values of high-dose images as ground truth to calculate the variance of noise that is added to generate low-dose images, there may be inconsistencies in the added signal-dependent noise.

In order to preserve the characteristics of noise associated with low-dose images, the proposed approach takes into consideration the system parameters associated with low- and high-dose X-ray image acquisitions, such as system gain and electronic noise (Chap-

ter 4), in order to perform a noise variance stabilization and subsequently add noise in the variance-stabilized domain. The method has been evaluated by comparing several corresponding regions of the associated real and simulated low-dose images – obtained from two different imaging systems of two anthropomorphic phantoms – visually as well as statistically, using a two sample Kolmogorov-Smirnov Test at 5% significance. In addition to being visually similar and having small differences with respect to the mean values and the standard deviations, the hypothesis that the corresponding regions belong to the same distribution has been accepted in 81.43% of the cases based on 80 pairs of real and simulated low-dose regions.

The results suggest that the simulated low-dose images obtained using the proposed method are almost indistinguishable from real low-dose images. The use of an image formation model along with the system parameters used for imaging helps in reducing the extensive calibrations otherwise needed to construct look-up tables for finding the associations between mean intensity values and their associated noise variance. Moreover, the versatility of the method makes it easily adaptable to different imaging systems. This in turn leads to an increased diversity of the training data for potential learning-based methods.

An analytical approach for the simulation of realistic low-dose fluoroscopic images

Sai Gokul Hariharan[1,2] · **Norbert Strobel**[2,3] · **Christian Kaethner**[2] · **Markus Kowarschik**[1,2] · **Rebecca Fahrig**[2,5] · **Nassir Navab**[1,4]

Abstract

Purpose. The quality of X-ray images plays an important role in computer assisted interventions. Although, learning-based denoising techniques have been shown to be successful in improving the image quality, they often rely on pairs of associated low- and high-dose X-ray images that are usually not possible to acquire at different dose levels in a clinical scenario. Moreover, since data variation is an important requirement for learning-based methods, the use of phantom data alone may not be sufficient. A possibility to address this issue is a realistic simulation of low-dose images from their related high-dose counterparts.

Method. We introduce a novel noise simulation method based on an X-ray image formation model. The method makes use of the system parameters associated with low- and high-dose X-ray image acquisitions, such as system gain and electronic noise, to preserve the image noise characteristics of low-dose images.

Results. We have compared several corresponding regions of the associated real and simulated low-dose images - obtained from 2 different imaging systems - visually as well as statistically, using a two sample Kolmogorov-Smirnov Test at 5% significance. In addition to being visually similar, the hypothesis that the corresponding regions - from 80 pairs of real and simulated low-dose regions - belonging to the same distribution has been accepted in 81.43% of the cases.

Conclusion. The results suggest that the simulated low-dose images obtained using the proposed method are almost indistinguishable from real low-dose images. Since extensive calibration procedures required in previous methods can be avoided using the proposed approach, it allows an easy adaptation to different X-ray imaging systems. This in turn leads to an increased diversity of the training data for potential learning-based methods.

[1] Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2] Siemens Healthcare GmbH, Advanced Therapies, Forchheim, Germany
[3] Fakultät für Elektrotechnik, Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt, Schweinfurt, Germany
[4] Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America
[5] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

1 Introduction

Exposure to X-ray radiation has always been an important concern of patients as well as clinicians, who are exposed to a non-negligible amount of radiation over a period of their professional life time. The potential consequences can be classified into stochastic and deterministic risks. According to the ALARA (as low as reasonably achievable) principle, radiation dose can be lowered if the required image quality necessary to perform clinical tasks can be retained. Fortunately, the use of sophisticated image processing techniques, denoising in particular, may help in improving the image quality. If one can utilize such techniques to maintain the required image quality, the X-ray radiation dose may be reduced.

Analytical denoising techniques, such as block matching 3D (BM3D) [5] or wavelet coefficients' thresholding approach based on Stein's unbiased risk estimate (SURE) and linear expansion of thresholds (LET) [13], are effective spatial denoising techniques for X-ray images [2], in particular, when combined with noise variance stabilization [14]. Furthermore, the joint processing of successive X-ray images taken during fluoroscopic procedures provides additional gains [1,11,8]. Unfortunately, even such well-engineered approaches could fail to produce realistic images at very low-dose levels (<25% standard dose) where the images are severely corrupted by noise. On the other hand, learning based methods have been recently found to solve such complicated problems, if sufficient training data is available. Cascade of shrinkage fields [17] and trainable nonlinear reaction diffusion [4], have been shown to outperform the well-engineered classical techniques, such as BM3D, in the case of optical color images. Recently, deep learning has also facilitated the design of X-ray image denoising methods [16] as well as blind denoising methods for optical color images [22]. Wolterink et al. [20] have employed a generative adversarial network for denoising low-dose CT volumes by transforming them to a higher dose level. Unfortunately, these methods require pairs of images, which are often not available in clinical practice.

In the past, researchers have tried adding noise to high-dose images to simulate low-dose images based on the assumption that the sequences are corrupted by either Poisson noise or Gaussian noise or even both [16,1,3]. Unfortunately, they have usually relied on image formation models that may not fully account for how X-ray images are acquired in practice, e.g., when different gain factors are used depending on the tube settings. Often the assumption made is that the detector gain is 1. This may not always be true, as a higher gain is usually applied when the dose is low to make a better utilization of the available dynamic range of the Digital/Analog converter. Moreover, at very low-dose levels and high gain factors, even a small amount of electronic noise starts to matter, and the noise itself may also vary with respect to the detector gain that has been used. Finally, when developing and assessing algorithms, researchers have often relied on sequences stored in the DICOM (Digital Imaging and Communications in Medicine) format for generating noise corrupted images [16]. In general, DICOM image sequences are tone mapped versions of the Poisson-Gaussian corrupted input sequences and no longer represent actual pixel values, whose brightness is proportional to the amount of the received X-ray quanta. Unless system dependent parameters are

taken into account, it may not be possible to simulate realistic noisy low-dose sequences from their high-dose counterparts. In this paper, we propose a novel method to generate realistic low-dose X-ray images from unprocessed (i.e., non-linear post-processing has not been applied) high-dose X-ray images, while ensuring that the characteristics of the simulated and the corresponding real low-dose images are similar. The realistic simulations of low-dose images could be used to develop learning-based denoising approaches in order to enhance the image quality of low-dose X-ray images. This in turn would benefit patients as well as the clinicians as it would result in a reduction in the exposure to X-ray radiation.

2 Related Work

To address the issue of the lack of low-dose X-ray images, researchers have been considering the possibility of simulating them. In the following paragraphs, we briefly present some commonly used approaches.

Modelling noise has been an important prerequisite for designing noise reduction filters. Considering photon noise to be a fundamental limitation of images acquired at low-light levels, Kuan et al. [12] model the images to be corrupted by Poisson noise. Based on the degradation model [12], the corrupted pixel value $g[i,j]$ is modeled as:

$$g[i,j] = \frac{\text{Poisson}_a(f[i,j])}{a}, \tag{1}$$

where, $f[i,j]$ is the original pixel value, $\text{Poisson}_a(.)$ is a Poisson random generator and a is a proportionality factor that controls the severity of the noise level. According to [12], a is chosen arbitrarily. They suggest that a suitable range for a is between 0 and 1 (0 being the worst case). As the emission of photons from an X-ray source can be described as a Poisson process, Chan et al. [3] have adapted the model presented in Eq. 1 to X-ray imaging. According to their model, the probability of receiving k photons at the detector, given that the mean rate of arrival is n_0 photons, is represented by

$$p(k) = \frac{e^{-n_0} \cdot n_0^k}{k!}. \tag{2}$$

As the variance equals the mean for a Poisson process, the standard deviation of quantum noise is $\sqrt{n_0}$. Chan et al. [3] also assume that quantum mottle will mask the effects of other degradation factors, such as electronic noise and quantization noise. In the case of attenuation, the average number of received photons n is approximated as $\rho \cdot n_0$, where ρ is given as

$$\rho = e^{-\mu l} \tag{3}$$

with μ being the X-ray attenuation factor and l the path length. Finally, they map the photon count n to a gray value d using a linear function $f(.)$ as follows:

$$d = f(\alpha \cdot n), \tag{4}$$

where α is the gain of the imaging system. Based on this model, they corrupt the image by first decreasing the photon count by a degradation factor ρ - which could be the ratio of dose levels - then generate a new Poisson random number using the new mean $\rho \cdot n$ that represents the average photon count incident on the detector and finally rescale the value by $1/\rho$ to obtain the corrupted pixel to achieve the effect of the automatic gain control. Chan et al. [3] claim their model to be similar to the degradation model of Kuan et al. [12], if $\rho = a$ and the maximum number of photon counts is equal to the number of gray values. In what follows, we refer to this approach as DPN-Method (direct Poisson noise estimate). Unfortunately, this method may not be accurate enough as it involves corrupting a scaled actual X-ray image - which is already corrupted by some amount of Poisson noise - with more Poisson noise. Moreover, the influence of electronic noise is completely neglected. While permissible in a higher dose setting, this assumption may not be applicable for very low-dose scenarios.

Holdsworth et al. [10] have quantified noise in digital images by calculating the relationship between standard deviation and intensity values from bright field and dark field charged coupled device (CCD) images. Veldkamp et al. [19] have measured the overall noise associated with different pixel values for different mAs settings using raw X-ray images of a LucAl phantom. Then, by performing linear interpolations between the measured data, they have obtained a look-up table with pixel values and associated noise levels. For the simulation of low-dose image, they first calculated a pixel in the reduced dose image $r[i, j]$ by scaling the pixel intensity values in the input image $y[i, j]$ by a dose reduction quotient c and then added noise of the appropriate amount $m[i, j]$ as shown below:

$$r[i, j] = \frac{y[i, j]}{c} + m[i, j]. \tag{5}$$

$m[i, j]$ is white noise from a Gaussian distribution with variance $\sigma^2_{\text{add}}[i, j]$. This variance is computed from the variance associated with $y[i, j]$, $\sigma^2_y[i, j]$, and the variance associated with the corresponding mean pixel value of an acquired low-dose image, $\sigma^2_{\text{low}}[i, j]$ as follows:

$$\sigma^2_{\text{add}}[i, j] = \sigma^2_y[i, j] \frac{c^2 - p^2}{c^2 \cdot p^2}, \tag{6}$$

where

$$p^2 = \frac{\sigma^2_y[i, j]}{\sigma^2_{\text{low}}[i, j]}. \tag{7}$$

In the rest of the paper, this approach is referred to as SDAGN-Method (signal-dependent additive Gaussian noise).

Gislason-Lee et al. [6] have extended the method proposed in [19] by considering different beam energies and dose levels for constructing the look-up table. They have analyzed flat field X-ray images of polymethyl methacrylate blocks acquired at different beam energies and dose levels to obtain the relationship between variance of noise, the mean pixel intensity and the X-ray dose for different beam energies, i.e., for each beam energy they have plotted variance of noise as

a function of the mean pixel intensity at all dose levels and have calculated a linear fit. With the help of the obtained look-up table, they have simulated low-dose X-ray images using the method proposed in [19]. Although the methods proposed in [19] and [6] inherently take into account the system parameters with the help of extensive calibrations, the gray value $r[i,j]$ that is used for calculating the variance of noise to be added is already corrupted. This means that using the gray value $r[i,j]$ as the mean estimate of the pixel may be incorrect and this may cause inconsistencies in the added signal-dependent noise.

3 Material and Methods

Noise can have various sources, e.g., quantum noise, electronic noise or quantization noise, and, depending on the imaging situation, different gains are applied. To take this into account while simulating low-dose X-ray images, Gislason-Lee et al. [6] and Veldkamp et al. [19] have made use of extensive calibrations. On the other hand, we have introduced a low-dose simulation method based on an image formation model instead of extensive calibrations.

3.1 Image Formation and Noise Model

The intensity of an X-ray beam is the product of the number of photons in the beam and its energy. As the beam traverses matter, it is attenuated and scattered, resulting in a reduction of the intensity of the beam.

The formation of images from X-ray quanta can be assumed to follow a linear model [21]. The mean gray value $\bar{y}[i,j]$ at a particular location $[i,j]$ has a linear relationship with respect to the incident mean X-ray air kerma $\bar{x}$ at that location. That is, the measured mean quanta $\bar{x}[i,j]$ is scaled by the detector gain α and shifted by the overall system offset p (assuming flat field correction). This can be represented using:

$$\bar{y}[i,j] = \alpha \cdot \bar{x}[i,j] + p. \tag{8}$$

The quantum nature of X-ray photons can be modeled using a Poisson distribution. The presence of electronic noise, e.g., due to read out noise and dark current, can be modeled using a Gaussian distribution with zero mean. The formation of an image can thus be represented as:

$$y[i,j] = \alpha \cdot \big(x[i,j] + \eta_a[i,j]\big) + p + \eta_q[i,j] = \alpha \cdot x[i,j] + p + \eta[i,j]. \tag{9}$$

Here, x, η_a, η_q and y represent the quanta, electronic noise due to analog sensor read out, electronic noise due to quantization (analog to digital conversion) and the noise-corrupted detector pixel value, respectively. The overall additive noise can be represented as η ($\eta = \eta_a + \eta_q$). The gain is typically chosen such that the resulting pixel values occupy as much of the available dynamic range as possible before the quantization takes place. The choice of the detector gain determines the amount of Poisson noise present in the images. The variance associated with the gray values can be represented as:

$$\sigma_y^2[i,j] = \alpha^2 \cdot \sigma_x^2[i,j] + \sigma_n^2. \tag{10}$$

Since the variance of Poisson noise is related directly to the mean value, Eq. 10 can be rewritten as

$$\sigma_y^2[i,j] = \alpha^2 \cdot \bar{x}[i,j] + \sigma_\eta^2. \tag{11}$$

This noise model is also in line with [6] and [19], i.e., the relationship between the mean intensity value and the noise variance for different gray values can be represented using the associated system parameters. This could reduce the extensive calibrations that are required for their low-dose X-ray image simulations.

3.2 Generation of Low-dose Images from corresponding High-dose Images

Let us assume that we acquire a high-dose image $y_h[i,j]$ as well as a low-dose image $y_l[i,j]$ of the same object using the same tube voltage. From Eq. 9, the images can be modeled as:

$$y_h[i,j] = \alpha_h \cdot x_h[i,j] + p_h + \eta_h[i,j], \tag{12}$$

and

$$y_l[i,j] = \alpha_l \cdot x_l[i,j] + p_l + \eta_l[i,j]. \tag{13}$$

The mean dose at a particular pixel in the high-dose image should be a scaled version (by a factor c) of the mean dose of the corresponding pixel in the low-dose image, i.e.,

$$\bar{x}_l[i,j] = \bar{x}_h[i,j] \cdot c. \tag{14}$$

From Eq. 8, Eq. 12, Eq. 13 and Eq. 14, the mean gray values of the low-dose and high-dose can be related as follows:

$$\bar{y}_{h\to l}[i,j] = \frac{\alpha_l}{\alpha_h} \cdot c \cdot \bar{y}_h[i,j] + p_l - c \cdot \frac{\alpha_l}{\alpha_h} \cdot p_h. \tag{15}$$

From Eq. 15, it can be observed that the mean pixel values of the low-dose image $\bar{y}_{h\to l}$ can be derived from the mean pixel values of the high-dose image by scaling and adding an offset. This is the first step of the algorithm as shown in Fig. 1. The use of these transform parameters helps in retaining the mean of low-dose images, i.e, the mean gray values of $y_{h\to l}$ are the same as found in y_l. In our implementation, the outcomes of Eq. 14 and Eq. 15 are rounded to the next integer value. This transformation can also be performed by analyzing several corresponding regions of the low- and high-dose images and obtaining the scaling factor as well as the offset. The transformation also scales the electronic noise due to analog sensor read out η_a to match that of low-dose images.

The next step is to add noise to the linearly transformed high-dose image $\bar{y}_{h\to l}$ in such a way that its noise level matches the one found in the original (acquired) low-dose images [9]. We perform this in the generalized Anscombe domain by adding an appropriate amount of white Gaussian noise. The generalized Anscombe

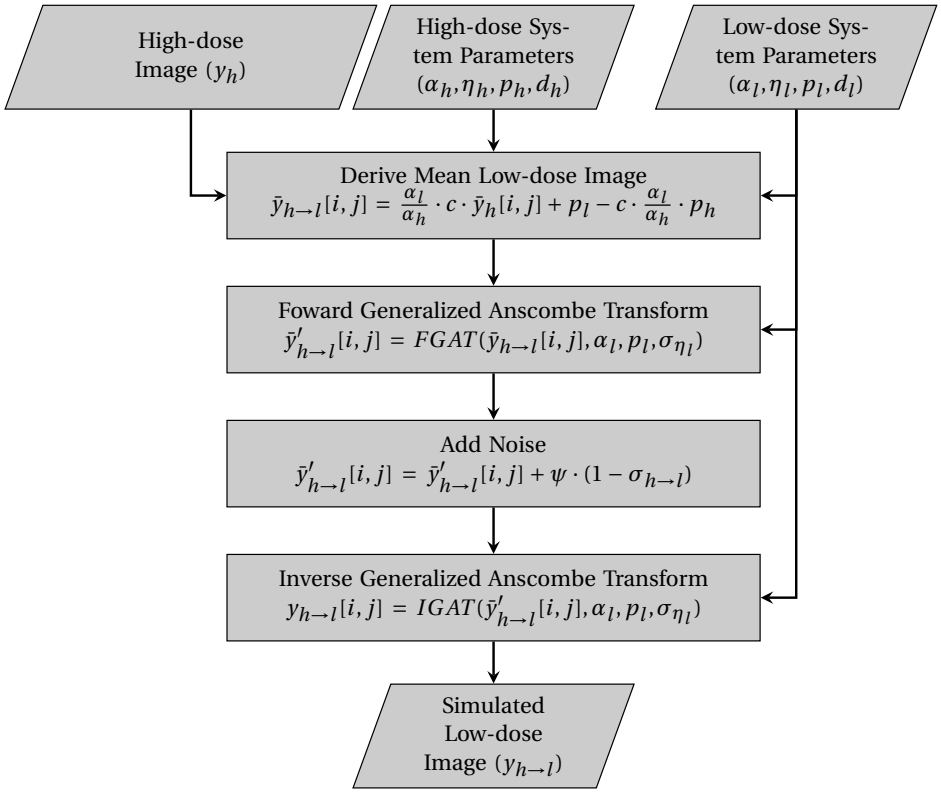


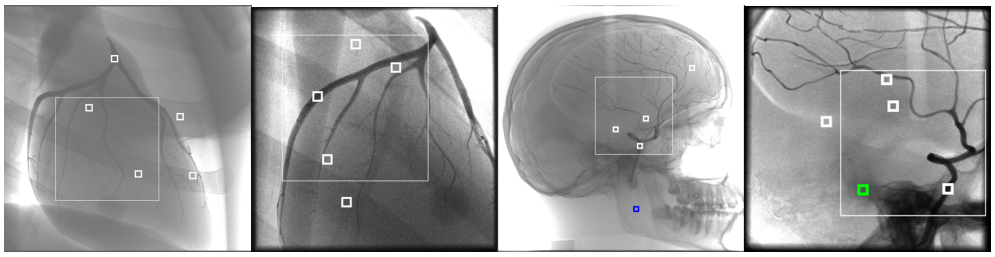
Fig. 1 Workflow of the proposed noise simulation approach.

transform (GAT) has been designed to stabilize signal-dependent mixed Poisson-Gaussian noise to a known constant (usually unit variance), i.e., the noise in the transformed image has a normal distribution [18]. The transform is applied element-wise to the image and the parameters required for the transform are the system gain, the system offset and the standard deviation of electronic noise. The system parameters of the low-dose image, α_l, p_l and σ_{η_l}, are used to apply the GAT on $\bar{y}_{h\to l}$ to obtain $\bar{y}'_{h\to l}$. Since the signal to noise ratio is much higher in high-dose images, the noise present in $\bar{y}'_{h\to l}$ ($\sigma_{h\to l}$) will, in general, be very low (variance much less than 1). The amount of noise in the image is measured by analyzing flat patches. Taking into account the amount of noise present, a noise mask if first computed with Gaussian noise of the standard deviation $1-\sigma_{h\to l}$. If binning of detector elements is involved, the noise mask is filtered accordingly, e.g, in the case of 2×2 binning of the detector elements, the noise mask is scaled by a factor of 2 and then a 2×2 average filter is applied to maintain the required standard deviation. Afterwards, the noise mask is added to $\bar{y}'_{h\to l}$ in order to make the standard deviation of noise equal to one. The corrupted image in the GAT domain is represented as $y'_{h\to l}$. Finally, an inverse GAT is applied on $y'_{h\to l}$ using the system parameters of the low-dose image, α_l, σ_{η_l} and p_l, to get back to the image domain $y_{h\to l}$.

3.3 Material

Unprocessed coronary and cerebral angiography X-ray images have been chosen to validate the proposed method. The X-ray images (after detector correction) have been acquired using a thorax phantom at different dose levels (100%, 81%, 50%, 28%, 15% and 7.5%) without moving the phantom. Similarly, X-ray images using a brain angiogram phantom have been acquired at dose levels of 100%, 66%, 33% and 22%. Based on the specified phantoms and dose levels, the X-ray images have been acquired using two different angiographic X-ray imaging systems (Siemens Healthineers AG, Erlangen, Germany). The first one is a general-purpose angiographic X-ray system, referred to as GP-System hereafter. The second one has been designed with low-dose applications in mind and has been abbreviated as LD-System in the rest of the manuscript. The sequences acquired at 100% dose - the clinical standard - have been used as a basis for the simulation of the low-dose images. It has also been verified that the images have been acquired at the same tube voltage. The system parameters used during the acquisition have been used for transforming the high-dose images y_h to low-dose $y_{h\rightarrow l}$.

Visual and statistical comparisons of simulated images obtained using DPN-Method [3], SDAGN-Method [19] and the method proposed in this paper have been performed. The high-dose images used for simulating the low-dose images along with the regions used for comparison have been highlighted in Fig. 2. The statistical comparison has been carried out by analyzing the standard deviation and the mean of several corresponding flat regions in the acquired low-dose and the simulated low-dose images for all the mentioned dose fractions. In addition to this, the corresponding regions of the simulated low-dose images generated from 100% dose images and the actual low-dose images have been compared using a two-sample Kolmogorov-Smirnov Test [15] at 5% significance level. The two-sample Kolmogorov-Smirnov (KS) test is a non-parametric hypothesis test that evaluates the difference between the cumulative distribution functions of the distributions of two sample data vectors.



(a) (b) (c) (d)

Fig. 2 100% dose images used for simulating low-dose images. (a) and (b) are the images of the thorax phantom obtained using GP-System and LD-System, respectively. (c) and (d) are the images of the brain angiogram phantom obtained using GP-System and LD-System, respectively. The regions used for evaluation have been highlighted. The regions highlighted with blue and green bounding boxes have been used for assessing the impact of the region size on the evaluation (Fig. 5 and Fig. 6).

Since, an open source implementation of DPN-Method [3] and SDAGN-Method [19] is not available, we have implemented the methods ourselves. In the case of DPN-Method [3], we do not perform a rescaling by $1/\rho$ as we work with raw images that do not involve an automatic gain control. For SDAGN-Method [19], we have used the system parameters associated with y_l (Eq. 12) and y_h (Eq. 13), to obtain σ_{low} and σ_y, respectively, as per Eq. 10.

4 Results

Since the reduction of the X-ray dose used for imaging is our main motivation, we have decided to focus on the visualization of the images with the lowest available dose levels in this study, i.e., 7.5% for the thorax phantom and 22% for the brain angiogram phantom. Figure 3a and Fig. 3b show two example regions of interests (ROIs) of the acquired 7.5% dose image of the thorax phantom (see Fig. 2a and Fig. 2b) and the simulated images using the methods presented in [3], [19] and the proposed method for the two imaging systems, respectively. Similarly, Fig. 4a and Fig. 4b show two ROIs of the acquired and the simulated results for the brain angiogram phantom (see Fig. 2c and Fig. 2d). In the case of GP-System, it can be observed from Fig. 3a and Fig. 4a that the images simulated using DPN-Method [3] in the case of GP-System may contain either too much noise or too little noise. For LD-System, it can be seen in Fig. 4b and Fig. 3b that DPN-Method [3] performs comparatively better. The simulated images obtained using SDAGN-Method [19] are more similar to the acquired low-dose images with a slight difference in the appearance of the noise. In fact, the images generated by the proposed method have the highest visual similarity to the corresponding acquired low-dose images.

Prior to the statistical evaluations, we have determined a suitable region size by analyzing its impact on the computed standard deviations. In Fig. 5a and Fig. 5b, we have visualized the standard deviations computed using different region sizes for two exemplary region locations, a homogeneous region (highlighted with a blue bounding box in Fig. 2c) and a region with mild texture (highlighted with a green bounding box in Fig. 2d), respectively. It can be seen from Fig. 5a and Fig. 5b that when the region size is small, there may be inconsistencies in the computed estimates. Therefore, it may not be possible to estimate reliably the gray-level statistics of the region. However, if the size of the chosen region is too large, a reliable estimation of the local statistics may no longer be possible due to an overlap with the texture of neighboring regions. Therefore, the size of the region in the range [20 × 20 pixels, 50 × 50 pixels] may be a suitable choice for analyzing the local statistics of the generated images. In Fig. 6a and Fig. 6b, we have shown the influence of the region size - using all the selected regions - on the average percentage in deviation of the standard deviation of the measurements when compared to the ground truth, for DPN-Method [3], SDAGN-Method [19] and the proposed method in the cases of GP-System and LD-System, respectively. It can be observed that DPN-Method [3], has the maximum deviation, whereas there is a marginal difference in the values of SDAGN-Method [19] and the proposed method. In addition, no significant influence of the region size can be seen, in the cases of SDAGN-

(a) 7.5% dose images of the thorax phantom for GP-System.

(b) 7.5% dose images of the thorax phantom for LD-System.

Fig. 3 Results for the simulation of low-dose (7.5%) images of the thorax phantom derived from 100% dose images acquired using (a) GP-System and (b) LD-System. The first column from the left has the ROIs of the acquired low-dose images and columns two, three and four show the simulated results using DPN-Method [3], SDAGN-Method [19] and the proposed method, respectively.

(a) 22% dose images of the brain angiogram phantom for GP-System.

(b) 22% dose images of the brain angiogram phantom for LD-System.

Fig. 4 Results for the simulation of low-dose (22%) images of the brain angiogram phantom derived from 100% dose images acquired using (a) GP-System and (b) LD-System. The first column from the left has the ROIs of the acquired low-dose images and columns two, three and four show the simulated results using DPN-Method [3], SDAGN-Method [19] and the proposed method, respectively.

Method [19] and the proposed method. However, it must be noted that for the data obtained from LD-System, the standard deviation begins to stabilize only when the region size is around 35 × 35 pixels. Therefore, in order to have a sufficient sample size to perform a reliable analysis of the statistics of the simulated images - in

the presence of almost negligible to low amounts of texture - we have chosen the region size to be 35 × 35 pixels.

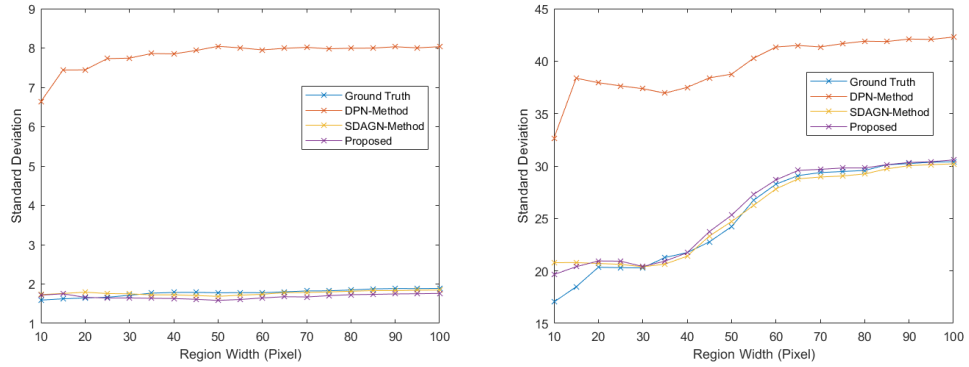


Fig. 5 Influence of a chosen region size (region width × region width) on the computed standard deviation in (a) the absence of variance due to texture (computed from the region highlighted with a blue bounding box in Fig. 2c) and (b) the presence of variance due to mild texture (computed from the region highlighted with a green bounding box in Fig. 2d).

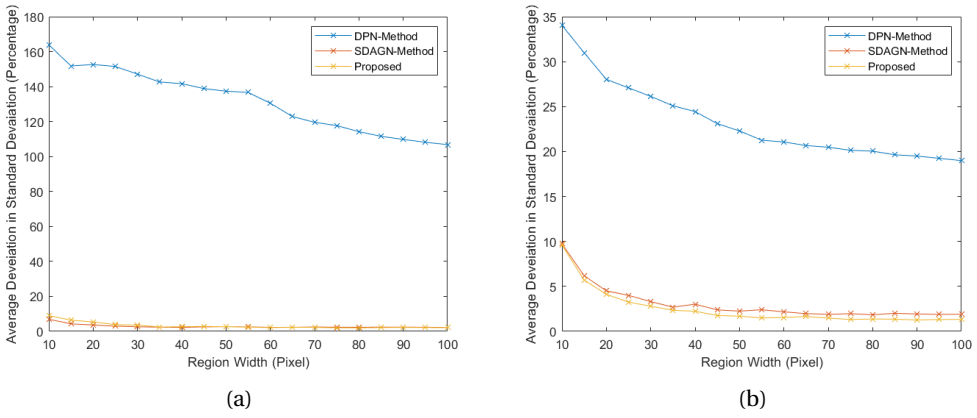


Fig. 6 Influence of a chosen region size (region width × region width) on the average percentage in deviation of the standard deviation in the case of (a) GP-System and (b) LD-System.

We have presented the results of the statistical/quantitative analysis - with respect to means and standard deviations of several corresponding regions - of the images in Fig. 3 and Fig. 4, in Tab. 1 and Tab. 2, respectively. The quantitative analysis of the standard deviations supports the visual analysis of the images. In the case of DPN-Method [3], the difference in the standard deviations of the regions is comparatively low only for LD-System (see Tab. 2). However, for SDAGN-Method [19] and the proposed method, the standard deviations are close to the

acquired low-dose images in almost all the cases. From the analysis of the deviation in the mean values, it can be noticed that all the three methods ensure the preservation of the mean. In Tab. 3, we have presented the average percentage of deviations in the means and the standard deviations for the 80 regions of interests for the three methods and the two systems along with the results of the KS test. There is a significant difference in the percentage of deviation from the expected standard deviation in the case of the simulated low-dose (SLD) images obtained using DPN-Method [3] for GP-System (130.23%) when compared to LD-System (26.44%). But, in the case of SDAGN-Method [19], these deviations are comparatively less and more similar for the two systems (2.56% for GP-System and 3.01% for LD-System). However, the least deviation (2.49% for GP-System and 2.39% for LD-System) can be achieved for the simulated images generated using the proposed method. The average percentage of deviation in the means is similar and almost negligible at around 1.5% for all the compared methods. Finally, the results of the KS test show that in the case of the proposed method the hypothesis, that the regions belong to the same distribution, has been accepted for 81.43% of the analyzed cases, whereas, in the cases of DPN-Method [3] and SDAGN-Method [19] the hypothesis has been accepted only for 10% and 41.43% of the cases, respectively.

Table 1 Comparison of the low-dose (LD) images of the thorax phantom acquired at 7.5% dose with the corresponding simulated low-dose (SLD) images with respect to mean μ and standard deviation σ across several regions.

GP-System

Region	LD (7.5%)		SLD DPN-Method [3]		SLD SDAGN-Method [19]		SLD Proposed	
	μ	σ	μ	σ	μ	σ	μ	σ
1	63.13	11.77	**63.07**	12.94	62.46	**11.60**	63.59	11.42
2	17.56	7.44	**17.57**	6.06	17.76	7.55	18.36	**7.47**
3	38.71	9.00	39.08	9.18	**38.73**	**8.96**	39.00	8.75
4	26.36	8.42	26.47	7.62	**26.39**	8.10	26.90	**8.35**
5	8.89	7.21	**9.15**	4.58	9.37	**7.38**	9.90	6.92

LD-System

Region	LD (7.5%)		SLD DPN-Method [3]		SLD SDAGN-Method [19]		SLD Proposed	
	μ	σ	μ	σ	μ	σ	μ	σ
1	18.28	9.81	19.08	6.42	**19.00**	**9.66**	19.87	10.08
2	30.00	11.16	29.11	8.09	29.00	**11.50**	**30.33**	10.82
3	45.88	13.06	46.31	10.48	**45.91**	**12.83**	46.62	12.25
4	31.80	11.35	32.15	8.94	**32.13**	10.91	33.08	**11.37**
5	16.55	11.19	16.18	8.15	**16.80**	11.24	17.91	**11.20**

Table 2 Comparison of the low-dose (LD) images of the brain angiogram phantom acquired at 22% dose with the corresponding simulated low-dose (SLD) images with respect to mean μ and standard deviation σ across several regions.

GP-System

Region	LD (7.5%)		SLD DPN-Method [3]		SLD SDAGN-Method [19]		SLD Proposed	
	μ	σ	μ	σ	μ	σ	μ	σ
1	22.29	1.79	21.72	6.73	21.78	1.82	**21.94**	**1.80**
2	10.74	1.29	10.57	4.55	10.67	1.22	**10.80**	**1.26**
3	6.01	1.95	**6.03**	3.73	6.09	1.85	6.07	**1.97**
4	11.90	1.29	11.51	4.68	11.64	**1.30**	**11.75**	1.31
5	15.30	1.41	**15.30**	5.68	15.24	**1.35**	15.17	1.32

LD-System

Region	LD (7.5%)		SLD DPN-Method [3]		SLD SDAGN-Method [19]		SLD Proposed	
	μ	σ	μ	σ	μ	σ	μ	σ
1	52.44	29.60	52.70	30.27	53.98	28.12	**52.68**	**29.12**
2	80.32	9.90	81.43	13.86	81.88	**9.99**	**81.36**	9.57
3	116.88	9.97	116.76	16.43	116.05	**9.96**	**116.85**	9.57
4	149.73	10.91	150.80	18.35	**149.16**	10.65	150.99	**10.77**
5	121.83	18.27	**122.15**	21.89	122.79	**18.40**	122.41	**18.14**

Table 3 Comparison of the simulated and acquired low-dose images with respect to the average percentage in deviation of the means μ, the average percentage in standard deviations σ and the success rate of Kolmogorov-Smirnov test at 5% significance (h) for 80 regions of interests at different dose levels and different systems.

System	SLD DPN-Method [3]			SLD SDAGN-Method [19]			SLD Proposed		
	μ	σ	h	μ	σ	h	μ	σ	h
1	1.25	130.23	5.71	**1.08**	2.56	28.57	1.58	**2.49**	**80**
2	**1.07**	26.44	14.29	1.50	3.01	54.29	1.66	**2.39**	**82.86**

5 Discussion and Conclusion

In this paper, we have proposed an effective method to simulate low-dose X-ray images from their higher-dose counterparts by taking into account an X-ray image formation model and the corresponding system parameters used for imaging. The method involves the corruption of the scaled down pixel values of an image in a noise variance stabilized domain. The use of an image formation model along with the system parameters used for imaging helps in reducing the extensive calibrations otherwise needed to construct look-up tables to find the associations between mean intensity values and their associated noise variance. Moreover, compared to strategies that require extensive calibrations, the method is easily adaptable to different imaging systems.

Existing noise simulation methods that assume a noise model with unit system gain fail to provide satisfactory results, when the associated system gain differs from 1. This is because at low-dose levels a higher system gain is used and

therefore the influence of noise (including electronic noise) is significantly higher. Methods that do not account for system parameters require prior calibration and look-up tables for determining the choice of noise to be added. Even though, the requirement of such extensive calibrations could be substituted and simplified by the use of an X-ray image formation model, methods that involve the addition of noise based on the intensity value of a corrupt pixel may not give accurate results, for example, in terms of noise characteristics. Since the proposed method involves the addition of additive white Gaussian noise in a noise variance stabilized domain, the mean pixel intensity values are not required. The results of the KS Test at 5% significance level performed for 80 corresponding regions of the real and simulated low-dose images suggests that the proposed method generates statistically similar regions at an average rate of 81.43%, whereas, the methods [3] and [19] may produce such similar regions only for 10% and 41.43% of the cases, respectively. These results can be supported by the least average deviations of the standard deviations between real and simulated images (see Tab. 3).

Since the proposed method makes use of unprocessed X-ray images to simulate their low-dose counterparts, it can also be used to derive the low-dose counterparts of other images generated by an X-ray imaging system. For example, in computed tomography images, it could be used to generate the low-dose counterparts of the projection images before the line integrals are computed. A drawback of the proposed method is that a certain set of system parameters - in particular the relationship between dose and gray values (system gain) and the remaining electronic noise - need to be known. In general, these parameters can be estimated with the help of calibrations, but this requires significant effort as there are many possible detector entrance dose levels. Fortunately, these parameters can be estimated online using image processing techniques [7]. We also plan to extend our study to 3D C-arm applications and use the proposed method to generate training data for artificial neural networks for denoising low-dose X-ray images.

Acknowledgement, Disclaimer and Conflict of Interest: This work was supported by Siemens Healthineers AG. The concepts and results presented in this paper are based on research and are not commercially available. The authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Amiot, C., Girard, C., Chanussot, J., Pescatore, J., Desvignes, M.: Spatio-temporal multiscale denoising of fluoroscopic sequence. IEEE Trans Med Imaging **35**(6), 1565–1574 (2016)
2. Cerciello, T., Bifulco, P., Cesarelli, M., Fratini, A.: A comparison of denoising methods for x-ray fluoroscopic images. Biomed Signal Process Control **7**(6), 550–559 (2012)
3. Chan, C.L., Sullivan, B.J., Sahakian, A.V., Katsaggelos, A.K., Swiryn, S., Hueter, D.C., Frohlich, T.: Simulation of quantum mottle in digital angiographic images. In: Biomedical Image Processing, vol. 1245, pp. 104–111. SPIE (1990)

4. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1256–1272 (2017)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Bm3d image denoising with shape-adaptive principal component analysis. In: Signal Processing with Adaptive Sparse Structured Representations (2009)
6. Gislason-Lee, A.J., Kumcu, A., Kengyelics, S.M., Brettle, D.S., Treadgold, L.A., Sivananthan, M., Davies, A.G.: How much image noise can be added in cardiac x-ray imaging without loss in perceived image quality? J Electron Imaging **24**(5) (2015)
7. Hariharan, S.G., Kaethner, C., Strobel, N., Kowarschik, M., Fahrig, R., Navab, N.: Estimation of noise stabilization parameters for low-dose x-ray images corrupted by poisson-gaussian noise. Manuscript in preparation (2018)
8. Hariharan, S.G., Strobel, N., Kaethner, C., Kowarschik, M., Demirci, S., Albarqouni, S., Fahrig, R., Navab, N.: A photon recycling approach to the denoising of ultra-low dose x-ray sequences. Int J CARS (2018). DOI 10.1007/s11548-018-1746-2
9. Hariharan, S.G., Strobel, N., Kowarschik, M., Fahrig, R., Navab, N.: Simulation of realistic low dose fluoroscopic images from their high dose counterparts. In: Bildverarbeitung für die Medizin 2018, pp. 80–85. Springer (2018)
10. Holdsworth, D., Gerson, R., Fenster, A.: A time-delay integration charge-coupled device camera for slot-scanned digital radiography. Medical physics **17**(5), 876–886 (1990)
11. Kostadin, D., Alessandro, F., KAREN, E.: Video denoising by sparse 3d transform-domain collaborative filtering. In: Eur. Signal Process. Conf, vol. 149 (2007)
12. Kuan, D.T., Sawchuk, A.A., Strand, T.C., Chavel, P.: Adaptive noise smoothing filter for images with signal-dependent noise. IEEE transactions on pattern analysis and machine intelligence (2), 165–177 (1985)
13. Luisier, F., Blu, T., Unser, M.: A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. IEEE Trans Image Process **16**(3), 593–606 (2007)
14. Makitalo, M., Foi, A.: Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. IEEE Trans Image Process **22**(1), 91–103 (2013)
15. Massey Jr, F.J.: The kolmogorov-smirnov ttest for goodness of fit. J Am Stat Assoc **46**(253), 68–78 (1951)
16. Matviychuk, Y., Mailhé, B., Chen, X., Wang, Q., Kiraly, A., Strobel, N., Nadar, M.: Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy. Proc Int Conf Image Proc pp. 2330–2334 (2016)
17. Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, pp. 2774–2781 (2014)
18. Starck, J.L., Murtagh, F.D., Bijaoui, A.: Image processing and data analysis: the multiscale approach. Cambridge University Press (1998)
19. Veldkamp, W.J., Kroft, L.J., van Delft, J.P.A., Geleijns, J.: A technique for simulating the effect of dose reduction on image quality in digital chest radiography. Journal of digital imaging **22**(2), 114–125 (2009)
20. Wolterink, J.M., Leiner, T., Viergever, M.A., Isgum, I.: Generative adversarial networks for noise reduction in low-dose ct. IEEE Trans Med Imaging (2017)
21. Yang, K., Huang, S.Y., Packard, N.J., Boone, J.M.: Noise variance analysis using a flat panel x-ray detector: A method for additive noise assessment with application to breast ct applications. Med Phys **37**(7), 3527–3537 (2010)
22. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process (2017)

6.3. Learning-based X-ray Image Denoising Utilizing Model-based Image Simulations

6.3.1. Paper Summary

It is common practice to use advanced denoising techniques to ensure acceptable image quality at low X-ray dose levels. However, at very low dose levels, the application of conventional denoising techniques (such as block matching 3D [199] and weighted nuclear norm minimization [114]) may lead to undesirable artifacts or oversmoothing. In addition, applying these methods in real-time is non-trivial as they rely on solving a complex optimization problem [118]. On the other hand, supervised learning techniques, which are comparatively less expensive computationally during the inference phase, have outperformed conventional denoising techniques, provided aligned pairs of associated noise-free (high-dose X-ray) and noisy (low-dose X-ray) images are available [118]. Unfortunately, it is neither acceptable nor possible to acquire such image pairs in clinical practice.

The contribution of the following paper is a novel learning-based denoising strategy for X-ray images [5]. The approach involves the use of model-based simulations of low-dose X-ray images during the training phase [4] (Section 6.2). In addition, a data-driven normalization step based on an X-ray imaging model is proposed to make the approach suitable for images acquired at different dose levels. The normalization involves the application of the generalized Anscombe transform. It utilizes the imaging parameters associated with an X-ray image in order to stabilize the noise variance to a known constant [1] (Chapter 4). In other words, this normalization step is used to remove the dependency of the trained network on different amounts of signal-dependent noise in the input images acquired at different dose levels using different image acquisition protocols.

To train the denoising network, 1200 unprocessed clinical X-ray images acquired at 100% of the standard-dose have been used to simulate X-ray images corresponding to 25% and 30% of the standard X-ray dose. The method has been evaluated using 4475 clinical X-ray images (2425 at 25% and 2050 at 30% of the standard X-ray dose) and 400 X-ray images of four different anthropomorphic phantoms (200 at 25% and 200 at 30% of the standard X-ray dose).

A quantitative and a qualitative analysis carried out by image quality experts have shown that the proposed strategy outperforms well-established conventional X-ray image denoising methods. Most importantly, the experts have found the proposed method to yield well-denoised images that are sharp, free from artifacts and have a realistic appearance. In addition, the proposed strategy of applying noise variance stabilization prior to training and application of the network avoids the introduction of artifacts that may be present in the results of commonly used strategies – such as training a denoising network to transform low-dose X-ray images to higher dose images. Furthermore, the method results in an image quality that is superior to that of the standard-dose images used for training. This indicates that the performance of the method is not limited by the quality of data used during the training phase. Finally, the evaluation has also indicated that the proposed approach allows for a significant dose reduction without sacrificing important image information.

Author's Final Manuscript

Learning-based X-ray Image Denoising Utilizing Model-based Image Simulations

Sai Gokul Hariharan[1,2], Christian Kaethner[2], Norbert Strobel[2,3], Markus Kowarschik[1,2], Shadi Albarqouni[1], Rebecca Fahrig[2,4], and Nassir Navab[1,5]

[1] Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2] Siemens Healthineers AG, Advanced Therapies, Forchheim, Germany
[3] Fakultät für Elektrotechnik, Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt, Schweinfurt, Germany
[4] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[5] Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America

Abstract. X-ray guidance is an integral part of interventional procedures, but the exposure to ionizing radiation poses a non-negligible threat to patients and clinical staff. Unfortunately, a reduction in the X-ray dose results in a lower signal-to-noise ratio, which may impair the quality of X-ray images. To ensure an acceptable image quality while keeping the X-ray dose as low as possible, it is common practice to use denoising techniques. However, at very low dose levels, the application of conventional denoising techniques may lead to undesirable artifacts or oversmoothing. On the other hand, supervised learning techniques have outperformed conventional techniques in producing suitable results, provided aligned pairs of associated high- and low-dose X-ray images are available. Unfortunately, it is neither acceptable nor possible to acquire such image pairs during a clinical intervention. To enable the use of learning-based methods for the denoising of X-ray images, we propose a novel strategy that involves the use of model-based simulations of low-dose X-ray images during the training phase. We utilize a data-driven normalization step that increases the robustness of the proposed approach to varying amounts of signal-dependent noise associated with different X-ray image acquisition protocols. A quantitative and qualitative analysis based on clinical and phantom data shows that the proposed strategy outperforms well-established conventional X-ray image denoising methods. It also indicates that the proposed approach allows for a significant dose reduction without sacrificing important image information.

1 Introduction

X-ray guidance during interventional procedures has gained immense importance in the recent past. Since X-ray imaging involves ionizing radiation, the medical benefits are accompanied by health risks. Even though the X-ray radiation dose

could be lowered to reduce them, it would also lower the image quality and may result in images that are clinically unacceptable. To preserve the required image quality at lower dose levels, sophisticated denoising techniques can be used. Noise reduction has been an important requirement not only for medical images, but also for optical images. Among others, block matching 3D (BM3D) [3] and denoising based on weighted nuclear norm minimization (WNNM) [4] are well-engineered patch-based approaches that combine non-local self similarity with transform-based processing to provide a significant noise reduction. Though BM3D has been developed about a decade ago, it is still considered to be one of the most effective methods [2, 13]. However, as these methods rely on solving a complex optimization problem [13], applying them in real-time is non-trivial.

Deep learning-based denoising techniques have been applied to optical images [13] and medical images, e.g., fluoroscopic X-ray images [10]. Unfortunately, such methods require pairs of noise-free and noisy images. Acquiring such image pairs in a clinical setting is close to impossible due to the associated increase in ionizing radiation and patient motion. Fortunately, Lehtinen *et al.* [9] have recently shown that deep learning-based denoising techniques can be designed by training networks with pairs of spatially aligned noisy instances of images alone. However, even acquiring such pairs may be difficult in a clinical scenario. In principle, the need for aligned image pairs could be overcome by resorting to generative adversarial networks that make use of a cycle consistency loss [8]. However, high-dose X-ray images would again be needed, which may not be available. Since the results of the method are at the most as good as the high-quality images used for training [8], the lack of appropriate data is a limiting factor.

In this work, we propose a strategy that overcomes the above mentioned issues. It involves the training of a denoising neural network using different instances of simulated low-dose X-ray images. Since we desire an approach that is suitable for denoising X-ray images acquired at different dose levels, we introduce a model-based and data-driven normalization step. The normalization removes the dependency of the trained network on different amounts of signal-dependent noise in the input images acquired at different dose levels. To the best of our knowledge, this is the first work that has shown the benefits of considering an imaging model when designing learning-based X-ray image denoising algorithms.

2 Material and Methods

Depending on the imaging system and the situation, X-ray images can be corrupted by different amounts and types of noise, e.g., quantum noise, electronic noise and quantization noise. Moreover, the signal-to-noise-ratio is directly proportional to the radiation dose used. Based on these characteristics, we have designed a learning-based method to denoise X-ray images.

2.1 X-ray Imaging Model

The transformation of the received X-ray quanta collected at a flat-panel detector to a gray value in an image involves a series of steps, namely the absorption of

the X-rays in the scintillator as well as the generation, coupling and integration of the optical photons by evenly spaced photo-diodes [11]. It can be assumed that this process follows a linear model [12]. Thus, the observed noise-corrupted gray value $y \in \mathbb{R}^{M \times N}$ at row r and column c can be represented as

$$y[r, c] = \beta \cdot x[r, c] + g + n[r, c], \tag{1}$$

where $x \in \mathbb{R}^{M \times N}$, $n \in \mathbb{R}^{M \times N}$, β and g represent the charges (corrupted by quantum noise) at the photo-diodes, the electronic noise with standard deviation (STD) σ_n, the overall system gain and the (constant) system offset, respectively. The (mixed) noise variance of a detector pixel's gray value can be expressed as

$$\sigma_y^2[r, c] = \alpha \cdot (\bar{y}[r, c] - g) + \sigma_n^2, \tag{2}$$

a line with slope α and y-intercept σ_n^2. $\bar{y}[r, c]$ denotes the mean (noise-free) value of $y[r, c]$ and the parameter α depends on the gain mode of the detector [12]. The imaging parameters α and σ_n^2 can, for example, be obtained from the system specifications or using calibration measurements [11]. Once known, they can be taken into account to perform a noise variance stabilization (NVS) based on the generalized Anscombe transform (GAT) as suggested in [5, 6]:

$$y'[r, c] = t(y[r, c]) = \frac{2}{\alpha} \sqrt{\alpha \cdot y[r, c] + \frac{3}{8}\alpha^2 - \alpha \cdot g + \sigma_n^2}. \tag{3}$$

The resulting pixel values $y'[r, c]$ have signal-independent noise with unit variance.

2.2 Denoising of low-dose X-ray images

The learning process involves the transformation of noise-corrupted instances of an image (or a region) into another [9], where the different instances have the same underlying noise statistics. In order to get two such noise-corrupted instances, we have relied on realistic simulations of low-dose images y_{l_1} and y_{l_2} from a standard-dose X-ray image y_s. In the first step, y_s is scaled down – such that its gray value range matches the gray value range of a corresponding low-dose image – to obtain $y_{s \to l}$. Then, to simulate a low-dose image y_l from $y_{s \to l}$, signal-dependent and signal-independent noise need to be added [1]. We have simulated the signal-independent electronic noise as additive white Gaussian noise (AWGN) $\eta_e[r, c]$ in the image domain [1]. Since the gray values of $y_{s \to l}$ are not noise-free, we have simulated the signal-dependent quantum noise by adding filtered AWGN (correlated noise) $\eta_q[r, c]$ in the GAT domain [7]. Filtered AWGN is used to include the influence of the detector blur on quantum noise. The process is described by

$$y_l[r, c] = t^{-1}(t(y_{s \to l}[r, c]) + k_q * \eta_q[r, c]) + \eta_e[r, c], \tag{4}$$

where $t^{-1}(.)$ represents the inverse GAT. The filtering kernel k_q and the noise components η_q and η_e can be derived from the system's modulation transfer

function and noise power spectrum. As mentioned earlier, we have performed an NVS on y_{l_1} and y_{l_2} as per Eq. 3 to obtain y'_{l_1} and y'_{l_2}, respectively. Similarly, $y_{s\to l}$ can be normalized using the imaging parameters associated with y_{l_1} to get $y'_{s\to l}$. From y'_{l_1}, y'_{l_2} and $y'_{s\to l}$, we randomly select regions of size $K \times K$ around a particular location (e.g, around $[r, c]$) to get $l'_1 \in \mathbb{R}^{K\times K}$, $l'_2 \in \mathbb{R}^{K\times K}$ and $s' \in \mathbb{R}^{K\times K}$, respectively. Since X-ray images are always corrupted by some amount of noise, s' can be expressed as:

$$s'[r, c] = \bar{s'}[r, c] + \eta_{s\to l}[r, c], \tag{5}$$

where $\bar{s'}[r, c]$ is the noise-free version of $s'[r, c]$ and $\eta_{s\to l}[r, c]$ is the filtered white Gaussian noise component (associated with $s'[r, c]$ in the GAT domain) with STD $\sigma_{s\to l}$. Similarly, l'_1 and l'_2 can be expressed as:

$$\begin{aligned} l'_1[r, c] &= \bar{s'}[r, c] + \eta_{s\to l}[r, c] + \eta_1[r, c] = \bar{s'}[r, c] + \eta_{l_1}[r, c], \\ l'_2[r, c] &= \bar{s'}[r, c] + \eta_{s\to l}[r, c] + \eta_2[r, c] = \bar{s'}[r, c] + \eta_{l_2}[r, c], \end{aligned} \tag{6}$$

where, η_1 and η_2 are noise matrices with filtered AWGN of STD $\sqrt{1-\sigma^2_{s\to l}}$ and the STD of noise in η_{l_1} and η_{l_2} is 1.

The network $D(l'_1, \theta)$ is then trained with P randomly chosen pairs of noisy regions $\{l'^{(i)}_1, l'^{(i)}_2\}_{i=1}^{P}$ based on minimizing the function

$$l(\theta) = \frac{1}{P}\sum_{i=1}^{P}\left\|\left(D(l'^{(i)}_1, \theta) - l'^{(i)}_2\right)\right\|_2 = \frac{1}{P}\sum_{i=1}^{P}\left\|\left(l'^{(i)}_1 + R(l'^{(i)}_1, \theta) - l'^{(i)}_2\right)\right\|_2, \tag{7}$$

where θ represents the parameters of D and the output of $D(l'_1, \theta)$ has been rewritten as $l'_1 + R(l'_1, \theta)$ ($R(l'_1, \theta) \in \mathbb{R}^{K\times K}$ is the difference between the output and the input of the network). Using Eq. 6, Eq. 7 can be rewritten as

$$\begin{aligned} l(\theta) &= \frac{1}{P}\sum_{i=1}^{P}\left\|\left(\bar{s'}^{(i)} + \eta^{(i)}_{l_1} + R(l'^{(i)}_1, \theta) - (\bar{s'}^{(i)} + \eta^{(i)}_{l_2})\right)\right\|_2 \\ &= \frac{1}{P}\sum_{i=1}^{P}\left\|\left(\eta^{(i)}_{l_1} + R(l'^{(i)}_1, \theta) - \eta^{(i)}_{l_2}\right)\right\|_2. \end{aligned} \tag{8}$$

Due to the use of a large amount of simulations, the random nature of noise and the dependency of the output of the network mainly on l'_1 (and its independence from l'_2), it is not possible for the network to result in $R(l'_1, \theta) = \eta_{l_2} - \eta_{l_1}$ to reach an average minimum loss of 0. On the other hand, it may be possible to obtain $R(l'_1, \theta) \sim -\eta_{l_1}$, which would result in an average loss of around 1 as STD of noise in $\eta_{l_1} \sim 1$. This implies that the best possible solution to Eq. 8 is the generation of a noise-free region.

Since the proposed method involves pairs of low-dose images as input, we have referred to it as L2L (low2low as analogy to noise2noise [9]). As an alternative, l'_2 can also be replaced by s' in Eq. 7, which would result in the convergence of the average loss to $\sigma_{\eta_{s\to l}}$. We have referred to this approach as L2H (low2high).

2.3 Material

In terms of the network architecture, we have used the U-Net architecture and the training strategy proposed in [9]. To train networks for the different denoising approaches, we have made use of 1200 unprocessed clinical X-ray images (M=N=896) acquired at 100% of the standard X-ray dose and simulated X-ray images corresponding to 25% and 30% of the standard X-ray dose. We have trained the networks (one network for each approach) for 1000 epochs using quadratic regions of width K=128. Then, we have evaluated the trained networks on 4475 clinical X-ray images (2425 at 25% and 2050 at 30% of the standard X-ray dose) and 400 X-ray images of 4 different anthropomorphic phantoms (200 at 25% and 200 at 30% of the standard X-ray dose). To show the impact of using an NVS, we have presented the results of the networks trained for L2L and L2H with and without an NVS. As a benchmark, we have also compared the results of the learning-based approaches with those of WNNM [4] and BM3D [3]. Since these methods have been designed specifically for AWGN, we have applied an NVS (Eq. 3) before using them. For the phantom images, quantitative analyses have been performed with respect to peak signal-to-noise-ratio (PSNR), structural similarity index (SSIM) and contrast-to-noise ratio (CNR). Noise-free ground-truth (GT) images have been generated by temporally averaging 500 noisy static images. Since it is not possible to acquire a GT for clinical images, we have performed a quantitative evaluation using CNR alone. In addition, a blind qualitative evaluation of fifteen scenes by six independent X-ray image quality experts has been conducted.

3 Results

In Fig. 1, we have presented the input as well as the processed regions of interest (ROIs) of an angiographic head phantom acquired at 30% of the standard X-ray dose. It can be seen that, even after applying an NVS, BM3D has resulted in noticeable artifacts (Fig. 1d) and WNNM has yielded a comparatively blurred output (Fig. 1e). The learning-based approaches, L2L as well as L2H, have also resulted in mild artifacts when an NVS is not used (Fig. 1f and Fig. 1g). However, combining these approaches with an NVS has improved their performance significantly (Fig. 1h and Fig. 1i). In particular, L2L with an NVS (Fig. 1i) has achieved a superior denoising performance compared to the other methods. Moreover, the result of L2L with NVS (Fig. 1i) is more similar to its noise-free counterpart (Fig. 1a) and also less noisy compared to the corresponding unprocessed standard-dose ROI (Fig. 1b). In Fig. 2, it can be seen that for the clinical images acquired at 30% of the standard-dose the results are in accordance with those from the phantom study. Since the use of L2L has resulted in less artifacts when compared to L2H in the phantom study, we have focused on L2L for this comparison.

In Tab. 1, we have presented the results of the quantitative analyses for the X-ray images acquired at 25% and 30% of the standard dose level. It can be observed that all the methods have resulted in a substantial improvement over the input. The learning-based approaches L2L and L2H have performed better

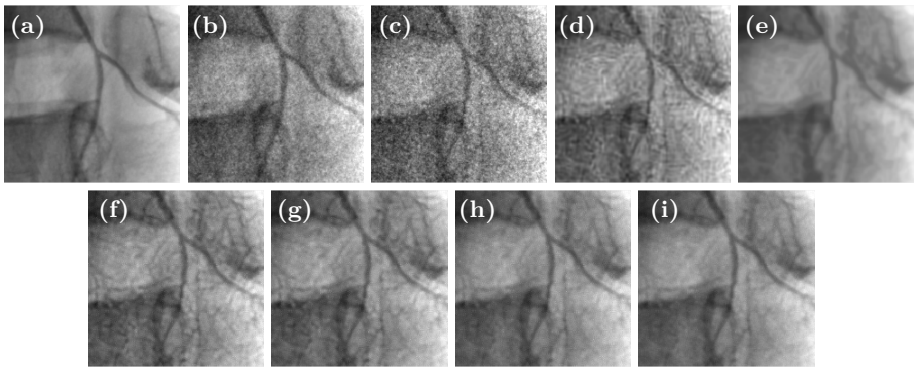


Fig. 1: Visualization of X-ray acquisitions (selected ROI) of an angiographic head phantom: (a) GT, (b) standard-dose, (c) 30% standard-dose and the results of processing (c) using: (d) BM3D with NVS (e) WNNM with NVS, (f) L2H without NVS, (g) L2L without NVS, (h) L2H with NVS and (i) L2L with NVS.

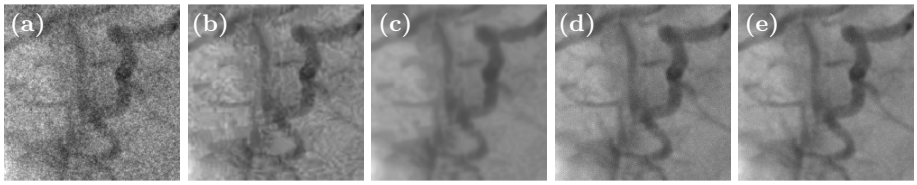


Fig. 2: Visualization of selected ROIs of an angiographic sequence acquired at 30% of the standard dose level: (a) input, (b) BM3D with NVS (c) WNNM with NVS, (d) L2L without NVS and (e) L2L with NVS.

when an NVS is involved. In fact, L2L with NVS has outperformed the other methods with respect to mean PSNR and SSIM. On the other hand, WNNM with NVS has resulted in a mean CNR that is even higher than that of the GT images indicating that the resulting images have been oversmoothed. These findings support the visual analysis.

The result of the experts' evaluation presented in Fig. 3 indicates that the learning-based approaches outperform BM3D and WNNM. Even though WNNM has obtained high scores for denoising performance and freedom from artifacts, it has received low scores for sharpness and realistic appearance. On an average, L2L with NVS has been found to produce visually superior results that are well denoised, sharp, artifact-free and most importantly realistic.

4 Discussion and Conclusion

We have presented a learning-based denoising approach that uses an imaging model to generate training data as well as to normalize the input data. Since the

Table 1: Quantitative evaluation of the denoising methods. The mean values for the different metrics are presented and the values closest to GT are highlighted.

Dose	Category	Metric	Input	BM3D NVS	WNNM NVS	L2H	L2L	L2H NVS	L2L NVS	GT
25%	Phantom	PSNR	25.69	29.92	32.54	30.28	30.67	31.70	**32.63**	Inf
		SSIM	0.927	0.971	0.984	0.973	0.975	0.980	**0.985**	1
		CNR	16.39	45.84	84.72	36.98	38.92	40.90	**52.16**	58.69
	Clinical	CNR	15.81	43.96	94.57	34.83	35.70	36.12	46.02	-
30%	Phantom	PSNR	26.75	30.93	33.06	31.26	31.66	32.65	**33.65**	Inf
		SSIM	0.941	0.977	0.985	0.978	0.980	0.984	**0.988**	1
		CNR	18.75	50.99	119.86	39.92	41.98	44.36	**57.07**	63.59
	Clinical	CNR	17.25	42.63	83.58	35.71	36.55	37.01	45.67	-

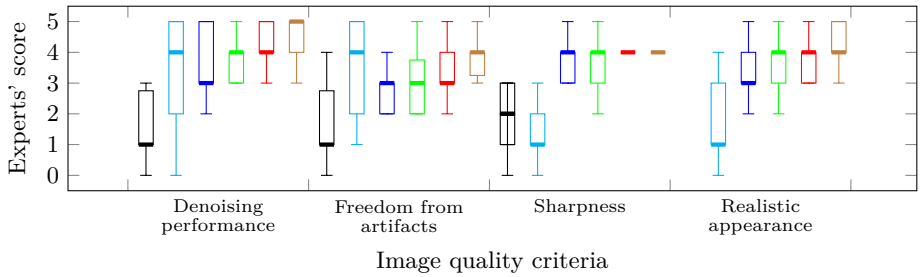


Fig. 3: Qualitative evaluation of BM3D with NVS (black), WNNM with NVS (cyan), L2H without NVS (blue), L2L without NVS (green), L2H with NVS (red) and L2L with NVS (brown), where 0 is the worst and 5 is the best possible score.

well established patch-based denoising techniques BM3D [3] and WNNM [4] have been designed to produce smooth images, they underperform on low-dose X-ray images corrupted by high amounts of noise – even when used with an NVS. The learning-based approaches L2L and L2H (irrespective of whether an NVS is used) have produced visually superior results for low-dose X-ray images when compared to BM3D and WNNM. A possible explanation could be the use of a sufficiently large amount of realistic training data that is usually important for learning the required features. Furthermore, a significant gain in the performance of both L2L and L2H can be achieved when the data is normalized using an NVS. This is due to the fact that the system gain is not consistent across different dose levels. Moreover, if the presence of signal-dependent noise is not taken into account, denoising may result in an uneven smoothing across different gray values thereby producing artifacts. Therefore, the proposed strategy of applying an NVS prior to the training as well as during the application not only avoids such artifacts, but also makes the network robust to different dose levels. This reiterates the fact that considering simple noise models, such as a Gaussian noise model [10], to simulate training data may not always be a suitable strategy. Since L2L makes use of pairs of low-dose simulations, the network is exposed to more instances of

noisy data when compared to L2H. This factors into a marginal improvement of L2L over the latter. In addition, the method results in an image quality that is superior to that of the standard-dose images used for training. This indicates that unlike [8], the performance of the method is not limited by the quality of the data used during the training phase.

Finally, quantitative and qualitative evaluations with respect to the images used in the study suggest that the proposed approach allows for a significant dose reduction without sacrificing important image information. However, more experiments are needed to thoroughly analyze the dose saving potential of the proposed method. A further reduction in the dose may be difficult to achieve in practice using spatial denoising alone as ultra low-dose images can be severely corrupted by detector artifacts. In such cases, it would be necessary to include temporal information to achieve an optimal denoising performance.

Acknowledgement and Disclaimer: This work was supported by Siemens Healthineers AG. The concepts and results presented in this paper are based on research and are not commercially available.

References

1. Borges, L.R., Guerrero, I., Bakic, P.R., Foi, A., Maidment, A.D., Vieira, M.A.: Method for simulating dose reduction in digital breast tomosynthesis. IEEE Trans Med Imaging 36(11), 2331–2342 (2017)
2. Cerciello, T., Bifulco, P., Cesarelli, M., Fratini, A.: A comparison of denoising methods for x-ray fluoroscopic images. Biomed Signal Process Control 7(6), 550–559 (2012)
3. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans Image Process 16(8), 2080–2095 (2007)
4. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: Proc IEEE Conf Comput Vis Pattern Recognit. pp. 2862–2869 (2014)
5. Hariharan, S.G., Kaethner, C., Strobel, N., Kowarschik, M., DiNitto, J., Albarqouni, S., Fahrig, R., Navab, N.: Preliminary results of dsa denoising based on a weighted low-rank approach using an advanced neurovascular replication system. Int J Comput Assist Radiol Surg 14(7), 1117–1126 (2019)
6. Hariharan, S.G., Strobel, N., Kaethner, C., Kowarschik, M., Demirci, S., Albarqouni, S., Fahrig, R., Navab, N.: A photon recycling approach to the denoising of ultra-low dose x-ray sequences. Int J Comput Assist Radiol Surg 13(6), 847–854 (2018)
7. Hariharan, S.G., Strobel, N., Kaethner, C., Kowarschik, M., Fahrig, R., Navab, N.: An analytical approach for the simulation of realistic low-dose fluoroscopic images. Int J Comput Assist Radiol Surg pp. 1–10 (2019)
8. Kang, E., Koo, H.J., Yang, D.H., Seo, J.B., Ye, J.C.: Cycle-consistent adversarial denoising network for multiphase coronary ct angiography. Med Phys 46(2), 550–562 (2019)

9. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2Noise: Learning image restoration without clean data. In: Proc Int Conf Mach Learn. pp. 2965–2974 (2018)
10. Matviychuk, Y., Mailhé, B., Chen, X., Wang, Q., Kiraly, A., Strobel, N., Nadar, M.: Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy. Proc Int Conf Image Proc pp. 2330–2334 (2016)
11. Siewerdsen, J., Antonuk, L., El-Mohri, Y., Yorkston, J., Huang, W., Boudry, J., Cunningham, I.: Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (amfpis) for diagnostic radiology. Med Phys 24(1), 71–89 (1997)
12. Yang, K., Huang, S.Y., Packard, N.J., Boone, J.M.: Noise variance analysis using a flat panel x-ray detector: A method for additive noise assessment with application to breast ct applications. Med Phys 37(7), 3527–3537 (2010)
13. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process 26(7), 3142–3155 (2017)

6.4. Robust Learning-based X-ray Image Denoising - Potential Pitfalls, their Analysis and Solutions

6.4.1. Paper Summary

In Sec. 6.3, it has been shown that learning-based denoising techniques such as the one presented in [176], can also be applied to X-ray images by generating spatially aligned noisy instances utilizing model-based simulations of low-dose X-ray images. Although several noise simulation approaches have been proposed, they are often found to be based on rather simplistic noise models or prone to inconsistencies [196, 197, 198]. For example, X-ray images have been modeled to have Poisson noise alone or noise is added to already corrupted pixels with signal-dependent noise. To overcome the above mentioned limitations and to simulate realistic low-dose X-ray images, an approach for simulating signal-dependent quantum noise using Gaussian noise in a noise variance-stabilized domain has been introduced in Sec. 6.3 [5] as a part of the learning-based denoising strategy. In fact, the proposed learning-based denoising approach outperforms well-established analytical and learning-based denoising approaches qualitatively as well as quantitatively. However, methods that are trained on a particular noise model may be sensitive to a change in noise characteristics. That is, they may yield sub-optimal outcome when there is a significant difference in the characteristics of noise between the training and application data. In practice, such a situation may, for example, arise when there is a change in the operating conditions of the imaging system.

The main focus of the following paper [6] is on investigating the behavior of the algorithm with respect to a change in the noise characteristics. To this end, the X-ray imaging model presented in [5] (Sec. 6.3) is extended by incorporating an apparent, noise shaping blurring filter kernel. This filter function is used to generate correlated noise that is added in a noise variance-stabilized domain. Based on this extended model, several experiments are performed to understand the cause of image degradation while utilizing a learning-based denoising approach as well as identifying solutions to prevent the degradation. In order to train networks corresponding to different noise models, 1200 clinical X-ray images acquired at 100% of the standard-dose level have been utilized. For the evaluation, 1400 X-ray images (200 per dose level) of four anthropomorphic phantoms acquired at different low-dose levels, specifically 10%, 15%, 20%, 25%, 30% and 50%, as well as 100% of the standard-dose level, have been used. Finally, a general evaluation of the method has been performed using 3250 clinical X-ray images acquired at different fractions of the standard dose level (2050 at 30%, 1200 at 25%, 800 at 20%, 600 at 15% and 600 at 10%).

A thorough analysis of the filter function's impact on the denoising result has revealed the importance of using a filter function that is closely related to the imaging system's noise power spectrum. In addition, it has been demonstrated that the behavior of the denoising network can be explained well by drawing an analogy to adaptive sub-band thresholding-based denoising approaches. Moreover, the analysis has also shown that the outcome of the network becomes predictable when noise models involved during the training and application phases are taken into account. This emphasizes the fact that appropriate noise modeling plays a crucial role in developing learning-based denoising methods. In particular, a correct blurring filter function and accurate simulations during the training phase are essential to obtain superior denoising results on low-dose image data.

Author's Final Manuscript

Robust Learning-based X-ray Image Denoising - Potential Pitfalls, their Analysis and Solutions

Sai Gokul Hariharan[1,2], Christian Kaethner[2], Norbert Strobel[3], Markus Kowarschik[1,2], Rebecca Fahrig[2,4], Nassir Navab[1,5]

[1]Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2]Siemens Healthineers AG, Advanced Therapies, Forchheim, Germany
[3]Institute of Medical Engineering Schweinfurt, University of Applied Sciences, Würzburg-Schweinfurt, Schweinfurt, Germany
[4]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[5]Whiting School of Engineering, Johns Hopkins University, Baltimore, United States of America

Abstract.
Purpose: Since guidance based on X-ray imaging is an integral part of interventional procedures, continuous efforts are taken towards reducing the exposure of patients and clinical staff to ionizing radiation. Even though a reduction in the X-ray dose may lower associated radiation risks, it is likely to impair the quality of the acquired images, potentially making it more difficult for physicians to carry out their procedures.
Method: We present a robust learning-based denoising strategy involving model-based simulations of low-dose X-ray images during the training phase. The method also utilizes a data-driven normalization step – based on an X-ray imaging model – to stabilize the mixed signal-dependent noise associated with X-ray images. We thoroughly analyze the method's sensitivity to a mismatch in dose levels used for training and application. We also study the impact of differing noise models used when training for low and very low-dose X-ray images on the denoising results.
Results: A quantitative and qualitative analysis based on acquired phantom and clinical data has shown that the proposed learning-based strategy is stable across different dose levels and yields excellent denoising results, if an accurate noise model is applied. We also found that there can be severe artifacts when the noise characteristics of the training images are significantly different from those in the actual images to be processed. This problem can be especially acute at very low dose levels. During a thorough analysis of our experimental results, we further discovered that viewing the results from the perspective of denoising via thresholding of sub-band coefficients can be very beneficial to get a better understanding of the proposed learning-based denoising strategy.
Conclusion: The proposed learning-based denoising strategy provides scope for significant X-ray dose reduction without the loss of important image information if the characteristics of noise is accurately accounted for during the training phase.

1. Introduction

Guidance through X-ray imaging is an essential component for many minimally invasive interventional procedures which facilitate faster patient recovery, reduce the risk of infection, and lower healthcare costs. Since X-ray imaging involves ionizing radiation, these benefits are accompanied by some degree of health risks. As a consequence, it is crucial that the applied radiation dose is kept as low as possible. Unfortunately, lowering the X-ray radiation dose may also reduce the resulting image quality and may yield images that are clinically unacceptable. Fortunately, the application of sophisticated denoising techniques has shown considerable potential to obtain the required image quality at lower X-ray dose levels [1].

Prominent conventional denoising approaches can be broadly classified as: (i) transform-based techniques that involve sophisticated thresholding of coefficients in a transform domain [2–4], (ii) methods that take advantage of self-similar structures in images [5–15], and (iii) variational approaches that are based on partial differential equations [16, 17]. For example, block matching 3D (BM3D) [7] (still considered one of the most effective methods [18–20]) and denoising based on a weighted nuclear norm minimization (WNNM) [8] are well-known patch-based approaches that combine non-local self similarity with transform-based processing.

More recently, learning-based methods have been shown to outperform the traditional denoising techniques for optical images [19, 21] as well as for X-ray and Computed Tomography images [22, 23]. Unfortunately, such methods usually require spatially registered pairs of noise-free and noisy images whose acquisition in a clinical setting is practically close to impossible due to the associated additional radiation exposure and patient motion. As an alternative, it has been shown that deep learning-based denoising with a similar performance is feasible using pairs of spatially aligned noisy instances of optical images alone [24]. However, acquiring pairs of associated, spatially exactly registered X-ray images in a clinical scenario is also difficult in practice. In principle, the need for aligned image pairs could be overcome by resorting to generative adversarial networks that make use of a cycle consistency loss [25]. However, in such a case high-dose (HD) X-ray images would be needed, which may be either not available or, if available, would be a limiting factor, since the results would at the most be as good as the image quality of the HD images used during training. Other approaches not relying on aligned image pairs are the self-supervised methods presented in [26, 27]. They are based on the assumption that the signal is correlated while the noise is conditionally pixel-wise independent both of the signal and of itself. Since X-ray images are impaired by a mixture of correlated Poisson and uncorrelated Gaussian noise [28], the methods are not suitable. In [29], it has been shown that learning-based denoising techniques such as the one presented in [24], can also be applied to X-ray images by generating spatially aligned noisy instances utilizing model-based simulations of low-dose (LD) X-ray images.

Although several noise simulation approaches have been proposed, they are often

found to be based on rather simplistic noise models or prone to inconsistencies. In [30], X-ray images have been modelled to be corrupted by Poisson noise alone without considering the effects of electronic noise. While permissible in a higher dose setting, this assumption is hard to justify at (very) low X-ray dose levels. In [31], the overall noise associated with different pixel values for different mAs settings has been derived using extensive measurements. Then, for the simulation of LD images, a pixel is first calculated in the reduced dose image by scaling the pixel intensity values in the input image by a dose reduction quotient. Subsequently, an appropriate amount of noise is added. This method has been extended in [32] by taking into consideration different beam energies and dose levels. Although both methods inherently consider the system parameters, such as the system gain and the electronic noise component, the gray value used for calculating the variance of noise to be added is already corrupted. Since using an actual gray value as the noise-free estimate for a pixel's gray value is likely to be incorrect, this approach may cause inconsistencies in the added signal-dependent noise. To overcome the above mentioned limitations and to simulate realistic LD X-ray images, an approach for simulating signal-dependent quantum noise using Gaussian noise in a noise variance stabilized domain has been introduced by the authors [29], where the noise is signal-independent and Gaussian-distributed [33, 34].

In figure 1, the results of applying the proposed denoising method (introduced in [29]) in the context of LD X-ray image denoising is shown. It can be seen that the method outperforms well established analytical and learning-based denoising approaches qualitatively as well as quantitatively – with respect to peak signal-to-noise-ratio (PSNR) and structural similarity index (SSIM). The strategy of training the network with pairs of images of the same scenes corrupted by the same underlying noise model, enables the network to learn the noise characteristics and remove it. It is to be noted that in the case of Low2High [29], where a network is trained with spatially aligned pairs of LD and HD images to transform an LD image into a HD image, the network needs to learn the characteristics of noise associated with LD and HD images to perform the denoising. Since X-ray images are affected by Poission noise, the noise characteristics in the LD and HD images are different. This complicates training and may yield inferior results (as shown in figure 1). On the other hand, methods that are trained on a particular noise model are sensitive to changes in noise characteristics. They may yield sub-optimal outcome when there is a significant difference in the characteristics of noise between the training and application data. This can, for example, happen, if the X-ray images used for training have been acquired from an imaging system that is different from the system used for getting the images to be processed, or when there is an unexpected change in the operating conditions of the imaging system that has not been accounted for. Three examples are shown in the third, fourth and the fifth column of the second row in figure 1, respectively. In these cases, the noise properties comprised colored noise and white noise, all not matching the properties of the imaging system. The resulting sub-optimal outcomes demonstrate that, it is important to investigate an algorithm's robustness with respect to a change in the dose and the noise characteristics.

To better account for the noise's correlation properties, we revisit the derivation of the X-ray imaging model presented in [29] and extend the model by incorporating an apparent blurring filter function. This filter function is used to generate correlated noise that is added in a noise variance stabilized domain. Based on this extended model, we perform several experiments to better understand the cause of image degradation while utilizing a learning-based denoising approach. We also identify solutions to prevent artifacts. In addition, we demonstrate that the method can be applied across multiple dose levels with good results. In particular, we show that using a correct filter function and accurate simulations during the training phase are essential to obtain superior denoising results on LD image data. A thorough analysis of the filter function's impact on the denoising outcome reveals the importance of using a filter kernel that is related to the imaging system's noise power spectrum (NPS). Our analysis also shows that the outcome of the network becomes predictable when noise models involved during the training and application phases are taken into account. In fact, by drawing an analogy to adaptive sub-band thresholding-based denoising approaches, we demonstrate that the behavior of the network can be explained well this way. Finally, the outcome of our analysis reveals that appropriate noise modeling plays a crucial role in designing learning-based denoising methods.

2. Material and Methods

Depending on the imaging system and the clinical situation, X-ray images can be corrupted by different amounts and types of noise. Considering these factors, a learning-based denoising strategy that involves an iterative training of a denoising network has been designed. The respective training and application workflows are presented in figure 2. In the training phase, the input is a batch of b randomly chosen standard-dose (SD) images $y_s^{(i)} \in \mathbb{R}^{M\times N}$. The superscript i denotes the i^{th} SD image. As a first step, for each $y_s^{(i)}$, two low-dose (LD) images $y_{l_1}^{(i)} \in \mathbb{R}^{M\times N}$ and $y_{l_2}^{(i)} \in \mathbb{R}^{M\times N}$ are simulated with different instances of noise. Then, $y_{l_1}^{(i)}$ and $y_{l_2}^{(i)}$ are normalized using a noise variance stabilization transform to get $y_{l_1}^{\prime(i)} \in \mathbb{R}^{M\times N}$ and $y_{l_2}^{\prime(i)} \in \mathbb{R}^{M\times N}$, respectively. Subsequently, corresponding regions $l_1^{\prime(i)} \in \mathbb{R}^{m\times n}$ and $l_2^{\prime(i)} \in \mathbb{R}^{m\times n}$ are chosen from $y_{l_1}^{\prime(i)}$ and $y_{l_2}^{\prime(i)}$, and taken as the input and the expected output of the denoising network, respectively. During the application phase, an acquired LD image $y_l \in \mathbb{R}^{M\times N}$ is first normalized using a noise variance stabilization transform and provided as input to the trained network. Finally, an inverse noise variance stabilization transform is applied to the output of the network to map the denoised image back to the actual image domain. The different aspects of the method are explained in the following sections.

During the training phase, pairs of corresponding image regions (ROIs) are first extracted at randomly chosen spatial locations from pairs of simulated LD images. The simulated LD images, derived from standard-dose (SD) images, are normalized using the generalized Anscombe transform (GAT).

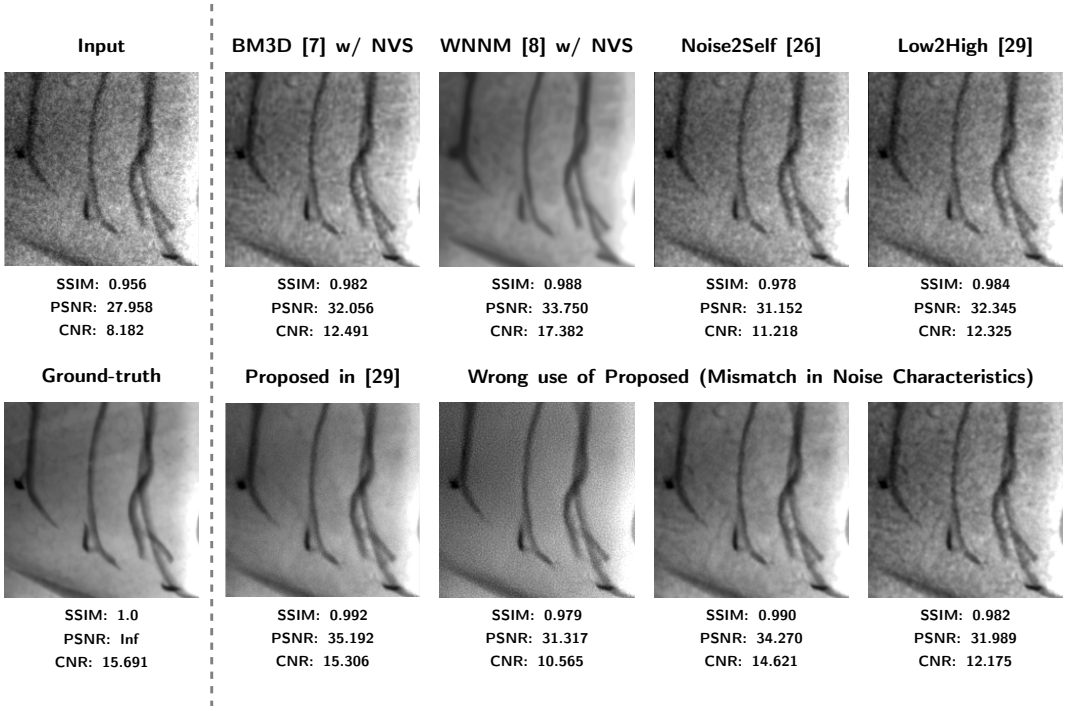


Figure 1: Visualization of selected regions of phantom images acquired at 30% of the standard-dose (SD) level. In addition to the input data, the results of different denoising methods and the GT for the phantom images are shown. Since the methods, block matching 3D (BM3D) and weighted nuclear norm minimization (WNNM), have been designed for additive Gaussian noise, they have been used after applying a noise variance stabilization (NVS). Noise-free ground-truth (GT) images have been generated by temporally averaging 500 static noisy images. The metrics structural similarity index (SSIM), peak-signal-to-noise ratio (PSNR) and contrast to noise ratio (CNR) have been used for performing a quantitative evaluation.

2.1. X-ray Imaging Model

The transformation of the received X-ray quanta – collected at an indirect-detection, flat-panel detector – into a pixel gray value can be described by a succession of stages [28]. Each stage may involve either a quantum gain or spatial spreading (blurring). It can be assumed that this process follows a linear model [35]. Thus, an observed noise-corrupted gray value $y \in \mathbb{R}^{M\times N}$ at row r and column c can be represented as

$$y[r,c] = \beta \cdot (x * k_q)[r,c] + g + n[r,c], \tag{1}$$

where $x \in \mathbb{R}^{M\times N}$ represents the charges (corrupted by quantum noise) at the photodiodes convolved $(*)$ with the stochastic spreading function k_q [28]. The variable $n \in \mathbb{R}^{M\times N}$ represents electronic noise with standard deviation (STD) σ_n sampled at row position r, and column position c, respectively. The overall scale factor and the (constant) system offset are given by β and g, respectively. The (mixed) noise variance

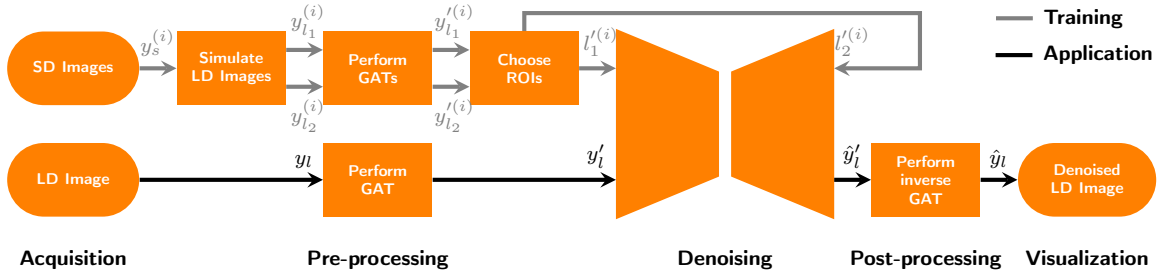


Figure 2: Schematic visualization of the workflows of the proposed strategy for training a denoising neural network (performed first) and its application (after training). During the training phase, pairs of simulated low-dose (LD) images $y_{l_1}^{(i)}$ and $y_{l_2}^{(i)}$ are first derived from standard-dose (SD) images $y_s^{(i)}$. The simulated LD images are subsequently normalized using the generalized Anscombe transform (GAT) to get $y_{l_1}^{\prime(i)}$ and $y_{l_2}^{\prime(i)}$, respectively. Then, pairs of corresponding image regions (ROIs) $l_1^{\prime(i)}$ and $l_2^{\prime(i)}$ are extracted at randomly chosen spatial locations from $y_{l_1}^{\prime(i)}$ and $y_{l_2}^{\prime(i)}$, respectively. These samples are then provided as the input and the target for the network, respectively. During the application phase, an acquired LD image y_l is normalized using the GAT (to obtain y_l') before being provided as an input to the denoising network. An inverse GAT is applied to the network' output $\hat{y}_l'$ in order to map it back to the image domain (where it is represented as $\hat{y}_l$).

of a detector pixel's gray value can be expressed as

$$\sigma_y^2[r,c] = \alpha \cdot \big(\bar{y}[r,c] - g\big) + \sigma_n^2. \tag{2}$$

This can be interpreted as a line with slope α and y-intercept $\sigma_n^2 - \alpha \cdot g$. In (2), $\bar{y}[r,c]$ denotes the mean (noise-free) value of $y[r,c]$ and the parameter α depends on the gain mode of the detector [35]. The imaging parameters α and σ_n^2 can, for example, be computed directly from an X-ray image [36] or obtained from the system specifications and calibration measurements [28]. Once known, they can be taken into account to perform a noise variance stabilization (NVS) based on the generalized Anscombe transform (GAT), $t_{\alpha,\sigma_n,g}(.)$, as suggested in [14, 29]:

$$\begin{aligned} y'[r,c] &= t_{\alpha,\sigma_n,g}\big(y[r,c]\big) \\ &= \frac{2}{\alpha}\sqrt{\alpha \cdot y[r,c] + \frac{3}{8}\alpha^2 - \alpha \cdot g + \sigma_n^2}. \end{aligned} \tag{3}$$

The resulting pixel values $y'[r,c]$ can be assumed to have signal-independent noise with unit variance. Figure 3 shows the power spectral densities (PSDs) and the noise power spectra (NPSs) derived in the GAT domain from stacks of 500 phantom images acquired at different dose levels. Since they have been computed in the GAT domain they are represented as PSD′ and NPS′, respectively. As expected, the NPS′ graphs remain similar across different dose levels, whereas, the PSD′ curves increase with dose. The increase is not linear due to the non-linear nature of the GAT.

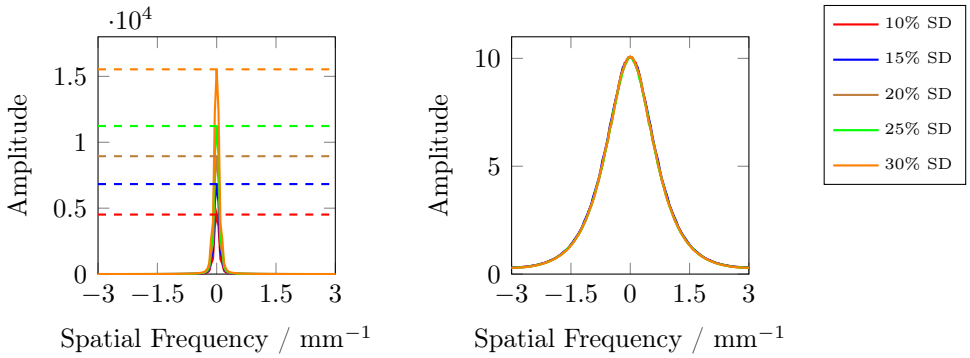


Figure 3: Cross-section visualization along the horizontal frequency axis of the power spectral densities estimated from stacks of phantom images acquired at different dose levels (left) as well as the associated noise power spectra (right). All the spectra are calculated in the generalized Anscombe transform domain. The peak amplitudes of the power spectral density graphs are represented using dashed lines.

The effect of the system blur introduced by k_q in (1) can be investigated by looking at the noise power spectra. We decided to analyze noise not in the image domain but in the noise variance stabilized domain because this is where we inject noise following Borges et al. [37]. To find the noise power spectra in the GAT domain, we have used flat-field images [38, 39]. First, we have transformed the flat-field images $y_b \in \mathbb{R}^{M\times N}$ using the GAT $t_{(.)}(.)$. Then, we have performed a power spectral analysis. In this context, we have subtracted an offset related to the electronic noise (computed from flat-field dark images) from the NPS of mixed noise to get the NPS of quantum noise in the GAT domain (referred to as NPS′). From this we get an estimate of the apparent blurring filter in the frequency domain by taking the square root of the individual frequency components of NPS′ (denoted as $\tilde{K}'_q$) [38, 39]. Subsequently, we have fitted a Gaussian mixture model $\hat{K}'_q$ to the estimated filter $\tilde{K}'_q$ (in the GAT domain) such that the mean squared error between the fitted model and the estimated filter is minimized. The fitted model is then represented as follows:

$$\hat{K}_q[u,v] = \sum_{i=1}^{4} \left(a_i e^{-\frac{(u-b_i)^2+(v-b_i)^2}{2\cdot w_i^2}} \right), \tag{4}$$

where a_i, b_i and w_i are the parameters of the Gaussian model. This process is shown in figure 4. The filter coefficients are normalized such that the sum of their square values is equal to one. In (4), the number of mixture models has been set to four as this resulted in the best fit for our experiments and analysis. It is to be noted that the Gaussian mixture model is only an example and other model fitting approaches can also be used.

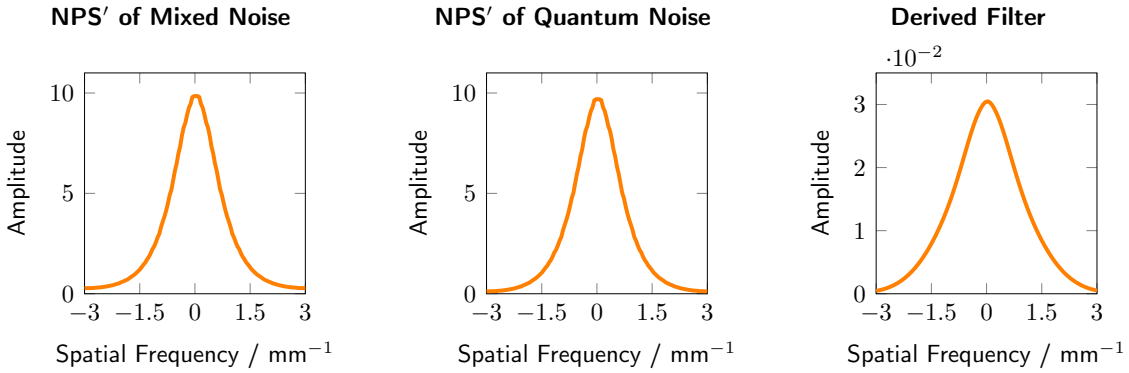


Figure 4: Cross-section visualization along the horizontal frequency axis of the noise power spectrum of the mixed noise (left), followed by the quantum noise (middle), computed by subtracting the offset related to the electronic noise, along with the normalized derived filter $\hat{K}'_q$ (right). All the spectra are computed in the generalized Anscombe transform domain and hence denoted as NPS′.

2.2. Simulation of Low-dose X-ray Images

The gray values of an SD X-ray image y_s are first scaled down by a dose reduction factor d such that its intensity range matches the gray value range of a corresponding LD image. The resulting image with a reduced gray level range is referred to as $y_{s\to l}$. Then, to obtain a realistic LD image y_l from $y_{s\to l}$, signal-dependent and signal-independent noise need to be added [40]. We have simulated the signal-independent electronic noise as additive white Gaussian noise (AWGN) $\eta_e[r,c]$ with standard deviation of $\sigma_e = \sigma_n\sqrt{1-\frac{1}{d^2}}$ and added it in the image domain [40]. Since the gray values of $y_{s\to l}$ are not noise-free, we have simulated the signal-dependent quantum noise by adding filtered AWGN $\eta_q[r,c]$ in the GAT domain [34]. The entire process is described by

$$y_l[r,c] = t^{-1}_{\alpha,\frac{\sigma_n}{d},g}\left(t_{\alpha,\frac{\sigma_n}{d},g}\left(y_{s\to l}[r,c]\right) + \eta_q[r,c]\right) + \eta_e[r,c], \tag{5}$$

where $t^{-1}_{(.)}(.)$ represents the inverse GAT. As per [41], the signal-dependent noise component η_q is obtained by filtering AWGN noise having STD σ_q with $\hat{K}'_q$, e.g., in the frequency domain. The noise variance in $y'_{s\to l}$ as per the applied GAT will be $\frac{1}{d}$. Since the resulting image is expected to have noise with unit variance and due to the chosen normalization of the filter coefficients, it holds that $\sigma_q = \sqrt{1-\frac{1}{d}}$.

2.3. Strategy for Training the Denoising Network

Training is performed in the GAT domain using P pairs of spatially aligned noisy image patches $l'^{(i)}_1$ and $l'^{(i)}_2$ that have the same signal component but different realizations of noise. In both cases, the noise has identical statistical properties. The P associated patch-pairs, each comprising $k \times k$ pixels, are randomly drawn from two simulated,

GAT-transformed LD images, y'_{l_1} and y'_{l_2}. The two underlying LD images, y_{l_1} and y_{l_2}, are derived from the same standard dose image, y_s following (5).

As a consequence of the steps outlined above, we can write

$$\begin{aligned} l_1'^{(i)}[r,c] &= \overline{s'^{(i)}}[r,c] + \eta_{l_1}^{(i)}[r,c], \\ l_2'^{(i)}[r,c] &= \overline{s'^{(i)}}[r,c] + \eta_{l_2}^{(i)}[r,c]. \end{aligned} \tag{6}$$

In (6), $\eta_{l_1}^{(i)}[r,c]$ and $\eta_{l_2}^{(i)}[r,c]$ are uncorrelated noise matrices with a standard deviation of one, and $\overline{s'^{(i)}}$ is the noise-free version of the i^{th} patch's signal component in the GAT domain.

Our goal now is to train a neural network estimating $\overline{s'}$ using paired noisy training examples l'_1 and l'_2. This is equivalent to finding a set of weights, θ, minimizing the following L_2 loss function

$$l(\theta) = \frac{1}{P}\sum_{i=1}^{P}\left\| \left(D(l_1'^{(i)}, \theta) - l_2'^{(i)} \right) \right\|_2 . \tag{7}$$

In this equation, the network $D(.)$ tries to predict the noisy patch $l_2'^{(i)}$ from the associated noisy patch $l_1'^{(i)}$. If we try to predict the noise-free signal of the i^{th} patch, $\overline{s'^{(i)}}$, from associated noisy regions, $l_1'^{(i)}$, then the best estimate is well known to be the conditional mean of $\overline{s'^{(i)}}$ given $l_1'^{(i)}$. In fact, this also holds if we try to predict noisy targets from noisy inputs [24]. In both cases, the output of the network is a prediction of the noise-free version of the input.

2.4. Application of the Denoising Network

In the application phase, the noise variance associated with an input image is normalized using the GAT, and the normalized image is provided as input to the trained network D (as shown in figure 2). An inverse GAT is applied to the output of the neural network to get the denoised image.

2.5. Analogy between learning-based and sub-band denoising

To gain some additional insights into how convolutional neural networks (CNNs) may perform denoising, let us recall that they comprise three main stages: First, the input image, a grayscale X-ray image in our case, is convolved (actually cross-correlated) with a set of (learned) filter kernels. This generates as many feature maps as there are filter kernels. Each feature map can be viewed as a collection of transform coefficients obtained by mapping image pixels within a certain receptive field onto (learned) basis functions. In the second step, each transform coefficient is thresholded, typically by means of a rectified linear activation unit (ReLU). Then, pooling, normally max-pooling, is applied. After this stage, the three-step procedure may be repeated on the reduced-sized feature maps effectively generating a convolutional pyramid.

Since classical, transform-based denoising approaches also involve convolutions, downsampling operations and thresholding, we can establish some interesting analogies between CNNs and sub-band transforms. We are particularly drawn to the *BayesShrink* soft-thresholding approach proposed by Chang et al. [3]. It facilitates the estimation of (sub-band) adaptive thresholds. This is important because adaptive thresholds are an integral part of CNNs. From a theoretical point of view, *BayesShrink* involves the soft-threshold T that minimizes the *Bayes risk* assuming a Generalized Gaussian prior. Based on the thresholded wavelet coefficients $\hat{X}$ and the actual wavelet coefficients X, the Bayes risk is defined as

$$r(T) = E(\hat{X} - X)^2. \tag{8}$$

The optimal threshold is the argument minimizing this function. Although no analytical solution could be derived, Chang et al. [3] found an approximation closely matching their numerically obtained results:

$$T_B(\sigma_X) = \frac{\sigma^2}{\sigma_X}. \tag{9}$$

Here σ^2 is the noise variance and σ_X is the standard deviation of the signal's wavelet coefficients. Both values can be calculated from the data within the individual sub-bands. This facilitates a data-driven, sub-band-adaptive thresholding approach. According to (9), we will get a small threshold if the noise in a sub-band is low compared to the signal. In such a case, most of the wavelet coefficients are kept. On the other hand, if the noise in a sub-band exceeds the signal, the resulting threshold is large effectively removing all but the biggest wavelet coefficients. In fact, it is possible to estimate σ_X from its context (context modeling) and obtain spatially adaptive thresholding [42]. Although wavelet denoising approaches are far less flexible than CNNs, they serve as a good working model for the explanations in the subsequent sections.

2.6. Material

To perform a thorough evaluation of the proposed algorithm, we have analyzed the robustness of the method as well as the potential causes of artifacts related to the training of the denoising network across different dose levels and noise characteristics both qualitatively and quantitatively. As network architecture, we have used the U-Net architecture [43] proposed in [24] and the training strategy proposed in [29]. Although different network architectures are available, we have chosen the U-Net architecture as it can be understood on a conceptual level, and, because it enables real-time image processing which is essential for guiding devices under X-ray. For generating training data corresponding to different (low) dose levels, we have utilized 1200 unprocessed clinical X-ray images (M=N=896 pixel) acquired at 100% standard dose (SD). We have trained a network (depending on the analysis) for 1000 epochs using square regions of width K=128 pixel. For the evaluation, we have taken 1400 X-ray images (200 per dose level) of four anthropomorphic phantoms acquired at different low dose (LD) levels, specifically at 10% SD, 15% SD, 20% SD, 25% SD, 30% SD and 50% SD, as well as

100% SD. For the quantitative evaluation, we have applied structural similarity index (SSIM) and peak signal-to-noise-ratio (PSNR) where the noise-free ground-truth (GT) images have been generated by temporally averaging 500 static noisy images. We have made use of phantom images for these evaluations instead of clinical images as it is basically impossible to obtain hundreds of static X-ray patient images in a clinical setting due to practical and ethical reasons. Such a stack of images is needed to arrive at noise-free references which are essential to perform a quantitative evaluation utilizing standard performance metrics such as PSNR and SSIM. For clinical images, where no ground truth is available, we have used contrast-to-noise ratio (CNR) – from pairs of homogeneous regions with different mean gray values.

To evaluate the stability of the proposed method across the different LD levels, we have first trained denoising networks for the different LD levels using associated images derived from the 100% SD clinical images (based on (5)). Then, we have applied the trained networks on X-ray images of the phantom acquired at the different LD levels. In addition, we have evaluated the performance of a mixed-dose denoising network. For training it, we used X-ray images associated with the different dose levels mentioned above. To obtain the X-ray images at different dose levels, we have relied on our model-based simulation approach.

When simulating X-ray images at various LD levels used for training, we need to add noise. This is a critical step that can give rise to denoising artifacts. In order to study the cause of such artifacts and to prevent them, we have investigated the impact of different noise characteristics. To this end, we colored AWGN differently by modifying $\hat{K}'_q$, when generating the training data. We have then trained multiple networks with the data generated from the unprocessed clinical X-ray images using modified blurring filter kernels

$$K'_{q,\delta}[u,v] = \sum_{i=1}^{4} \left(a_i e^{-\frac{(u-b_i)^2+(v-b_i)^2}{2\cdot\left(w_i\left(1+\frac{\delta}{100}\cdot\right)\right)^2}} \right), \tag{10}$$

where δ is the percentage by which the standard deviation of the Gaussian model has been changed. The parameter δ effectively controls the bandwidth of the apparent blurring filter. The smaller the value for δ, the narrower the bandwidth and vice versa. We have made sure that the filter coefficients have been properly normalized in all cases. In what follows, we have carefully studied how the networks react to different settings for $\delta \in [-60, -50, ..., 0, ..., 50, 60]$ depending on the X-ray dose of the input images. Here, $\delta = 0\%$ defaults to the original blurring filter kernel. When $\delta < 0\%$, the filter's bandwidth is reduced and fewer high-frequency components are retained, whereas for $\delta > 0\%$ the bandwidth increases. It should be noted that δ and the bandwidth of K'_q are not linearly related.

For the general evaluation of the approach, we have used 3250 clinical X-ray images (2050 at 30% SD, 1200 at 25% SD, 800 at 20% SD, 600 at 15% SD and 600 at 10% SD). As a benchmark, we have compared the results with those of other learning-based

denoising strategies, such as Noise2Self [26] and Low2High. Low2High is a commonly used strategy where spatially aligned pairs of simulated LD image and the associated SD image are used for training the denoising network [29]. In the case of Low2High, we have used pairs of simulated LD images (based on (5)) and the associated SD images. For Noise2Self, we have used 80% of the available acquired LD data for training the network.

3. Results

In this section, we present the results of our sensitivity analysis, both with respect to a mismatch in dose levels as well as with differing noise characteristics between the training images and the test images, respectively.

3.1. Influence of the Dose Level used for Training

To discuss the influence of the dose level utilized during the training phase, we present the results of applying the neural networks on X-ray images of an anthropomorphic thorax phantom acquired at 10% SD, 20% SD and 30% SD in figure 5. Prior to the application, we have trained the respective networks for 10% SD, 20% SD, and 30% SD as well as a mixture of LD levels (comprising LD images with dose levels ranging from 10% to 30% SD). As a reference, we also display the GT images associated with the input images.

Overall, the qualitative comparison of the denoised images acquired at the different dose levels shows that the denoising networks perform well independent of the particular training dose levels since there are only subtle differences between the denoised images. In the case of the X-ray images acquired at 10% SD, the network trained for 10% SD has resulted in well-denoised images with structures intact. Even in the case of the networks trained for 20% SD and 30% SD, the structures are restored clearly. However, these networks yield results with a higher amount of residual noise. The application of the denoising networks on images acquired at 20% SD shows that a training for 20% SD results in well-denoised and sharp images, whereas the use of a network trained for 10% SD results in slight blurring and a network trained for 30% SD results in some amount of residual noise. For the X-ray images acquired at 30% SD, the network trained for 30% SD has resulted in qualitatively good results, i.e., well-denoised and sharp images. Lowering the dose level used for training to 20% SD or even to 10% SD also yield good results demonstrating an overall stable denoising performance. However, it is to be noted that lower the dose level used for training, higher is the tendency of the network to generate slightly smoother and blurred images. In all the cases, the application of the network trained on mixed LD levels has resulted in denoised X-ray images resembling a balanced trade-off solution. This can be seen in the set of images provided as an example, where the results are most similar to the outcome achieved by applying the network trained for 20% SD.

Although the dose used during the training phase has some (small) impact on the results in terms of residual noise and spatial resolution, the differences between the input and the denoised results do not contain visible structures. As an example, the results of processing the images acquired at 10% SD and 30% SD using the networks trained for 10% SD and 30% SD are presented in figure 6. Even for the corner cases of this investigation presented here, the denoised results show that the structures are intact and no salient features are present in the images with the differences.

The quantitative evaluation of the results with respect to SSIM and PSNR is presented in Table. 1. The values confirm the qualitative evaluation based on the images shown in figure 5. Irrespective of the training dose levels and the application dose levels, the quality metrics are similar and high. The coefficient of variation (CV) between the quantitative values, computed by taking the ratio of the standard deviation of the values and the mean of the values, are below 0.38% for SSIM and 4.57% for PSNR across the applied dose levels and below 0.22% for SSIM and 2.05% for PSNR across networks trained for different dose levels. Irrespective of the input dose level, the network trained for 10% SD has resulted in the highest SSIM and PSNR, the network trained for 30% SD has resulted in the lowest and the network trained on mixed dose levels has resulted in a value that is in-between the above two cases.

3.2. Influence of the Noise Model used for Training

In figure 7 and figure 8, we present the impact of training denoising networks using different noise models. In this context, we have changed the power spectra of the noise added in the GAT domain during the training phase and have applied the trained networks on actual X-ray images acquired at 15% SD and 100% SD (second row in figure 7). These experiments have been performed using networks trained for the mixed LD levels. This has simplified the analysis while obtaining results that are only slightly different than what would have been obtained had the exact matches been used (see Sect. 3.1). From figure 7 it can be observed that when a network is trained using uncorrelated noise (flat power spectrum), it can produce results that are severely affected by artifacts especially for LD inputs. However, in the case of 100% SD inputs, using a flat noise power spectrum for training also produced acceptable results. For both the cases, the network trained for correlated noise derived using the proper filter kernel $\hat{K}'_q$, has resulted in substantial improvement.

In figure 8, we show the influence of the power spectra of correlated noise added in the GAT domain during the training phase by varying $\hat{K}'_q$ based on further image examples. The resulting networks have then been used during the application phase for the denoising of the actual anthropomorphic thorax phantom X-ray images to evaluate how a mismatch with respect to spectral properties would affect the denoising performance. When the trained networks are applied on an X-ray image acquired at 15% SD, it can be seen that the networks trained assuming correlated noise generated by applying a low-pass filter with a bandwidth that is too small, e.g., $\delta = -30\%$, have

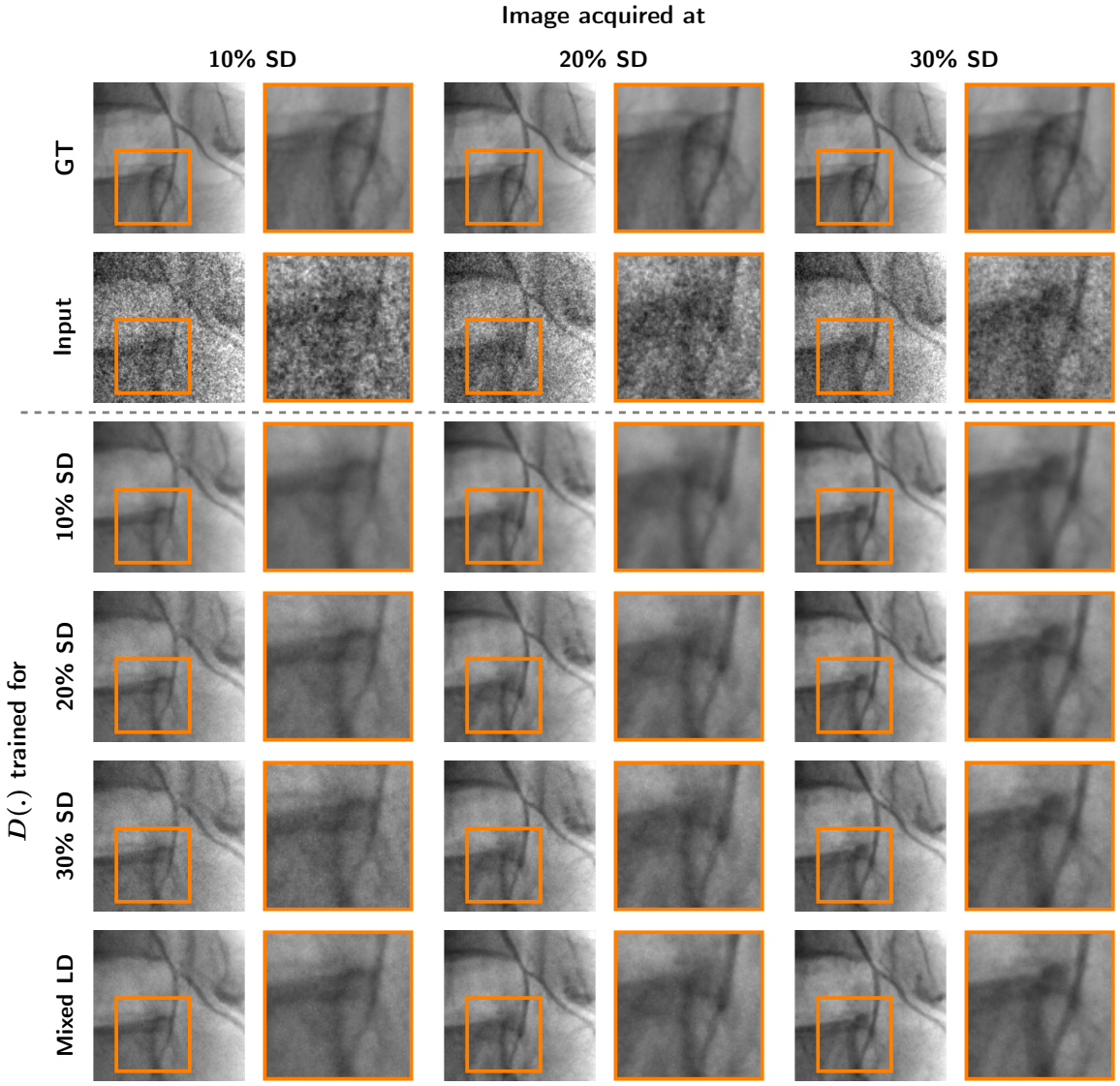


Figure 5: Visualizations of a region of interest of a phantom image acquired at different fractions of the standard X-ray dose (SD) level and processed using the networks $D(.)$ trained for different SD levels (last four rows). In addition to this, the (unprocessed) input and the GT data for the respective acquisitions are shown in the first two rows.

resulted in blurred images. The blurring can be severe. For example, in the case of $\delta = -60\%$, many details are lost in the denoised results. On the other hand, networks trained using correlated noise associated with a low-pass filters characterized by a very high cut-off frequency ($\delta >> 0\%$) have yielded artifacts that could be mistaken as anatomical structures, e.g., small blood vessels. Finally, we have found the resulting denoised images to be sufficiently sharp and free from artifacts only when the noise power spectra in the training and testing data are similar. On the contrary, for an X-ray fluoroscopy image acquired at 100% SD, where the SNR is comparatively higher, the

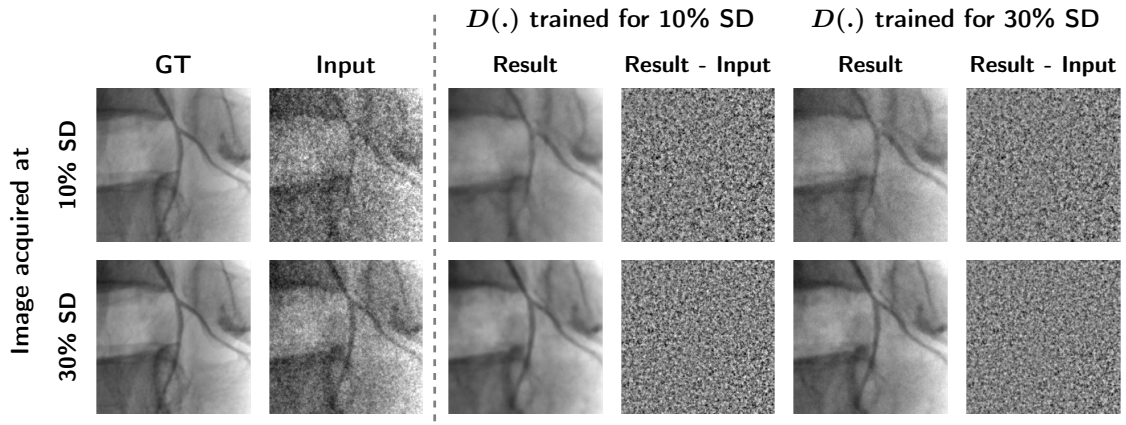


Figure 6: Visualization of a region of interest of phantom images acquired at 10% standard X-ray dose (SD) level and 30% SD. In addition to the ground truth (GT) and input data, the results of denoising using the networks $D(.)$ trained for 10% SD and 30% SD along with the respective differences to input images are shown.

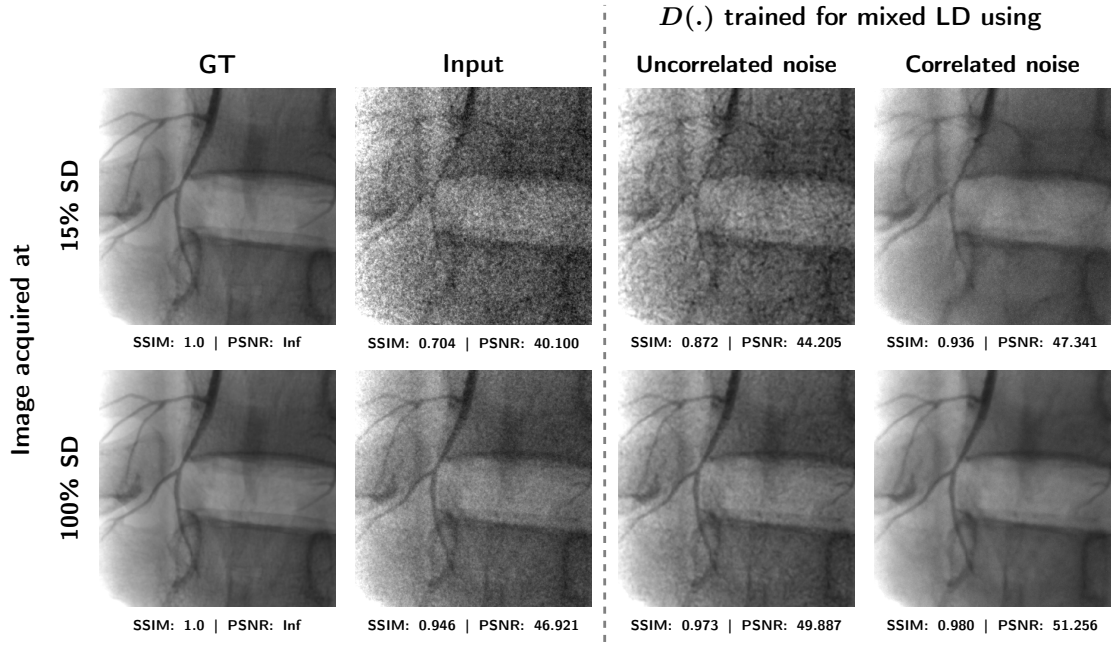


Figure 7: Visualization of the denoising results in the case of images acquired at 15% standard-dose (SD) and 100% SD for networks $D(.)$ trained for mixed LD with uncorrelated and correlated noise ($\delta = 0\%$). In addition to this, the respective ground truth (GT) and input images are displayed.

impact of the noise power spectrum's shape used during training has been less severe on the resulting denoised images. Compared to the observations at 15% SD, the structures are mostly intact even for larger deviations from $\delta = 0$. Similarly, the artifacts are less pronounced. These findings are also reflected in the quantitative analysis of our results

with respect to SSIM and PSNR (see figure 9). Since the SNR of an X-ray image is directly proportional to the X-ray dose, the metrics computed for higher dose images are higher. Across the dose levels, the highest values for the metrics are obtained for (and close to) $\delta = 0\%$. The results start deteriorating for larger deviations of δ from 0%. It can be seen from figure 8 and figure 9 that the worsening in the image quality within a dose level is directly related to the deviation in the shape of the filter kernel. In the

Table 1: Results of the quantitative evaluation regarding the influence of the dose level used for training a denoising network $D(.)$ with respect to structural similarity index (SSIM) and peak signal-to-noise-ratio (PSNR). SD represents the standard X-ray dose level. The coefficient of variation (CV) calculated across the rows and the columns of the results are included.

SSIM

		Image acquired at					CV
		10% SD	15% SD	20% SD	25% SD	30% SD	
	Input	0.906	0.935	0.949	0.959	0.967	
$D(.)$ trained for	10% SD	0.990	0.992	0.994	0.994	0.995	0.20%
	15% SD	0.989	0.991	0.993	0.994	0.995	0.24%
	20% SD	0.988	0.991	0.992	0.993	0.994	0.23%
	25% SD	0.986	0.990	0.992	0.993	0.994	0.32%
	30% SD	0.984	0.989	0.991	0.992	0.994	0.38%
	Mixed LD	0.988	0.990	0.992	0.993	0.994	0.24%
	CV	0.22%	0.10%	0.10%	0.08%	0.05%	

PSNR

		Image acquired at					CV
		10% SD	15% SD	20% SD	25% SD	30% SD	
	Input	25.334	26.900	28.266	29.240	30.381	
$D(.)$ trained for	10% SD	35.667	35.815	37.696	37.989	38.949	3.84%
	15% SD	35.168	35.505	37.184	37.638	38.455	3.82%
	20% SD	34.638	35.147	36.807	37.379	38.243	4.15%
	25% SD	34.210	34.924	36.400	36.853	37.739	3.99%
	30% SD	33.629	34.423	36.037	36.635	37.637	4.57%
	Mixed LD	34.670	34.911	36.388	36.818	37.797	3.64%
	CV	2.05%	1.39%	1.65%	1.44%	1.33%	

case of 15% SD, the CV computed across the SSIM values is 1.73% and that computed across the PSNR values is 1.80%. In the case of 30% SD, the CV is 0.90% for SSIM and 1.54% for PSNR. For a slightly higher dose of 50% SD, the CV is 0.53% for SSIM and 1.30% for PSNR. Finally, in the case of 100% SD, the CV is 0.26% for SSIM and 0.96% for PSNR. These results also support our previous visual analysis that the influence of the noise characteristics used during training is higher at LD levels and lower at HD levels. This means that lower the dose, more important it is to choose a filter kernel generating noise with the correct noise power spectrum, to avoid image degradation.

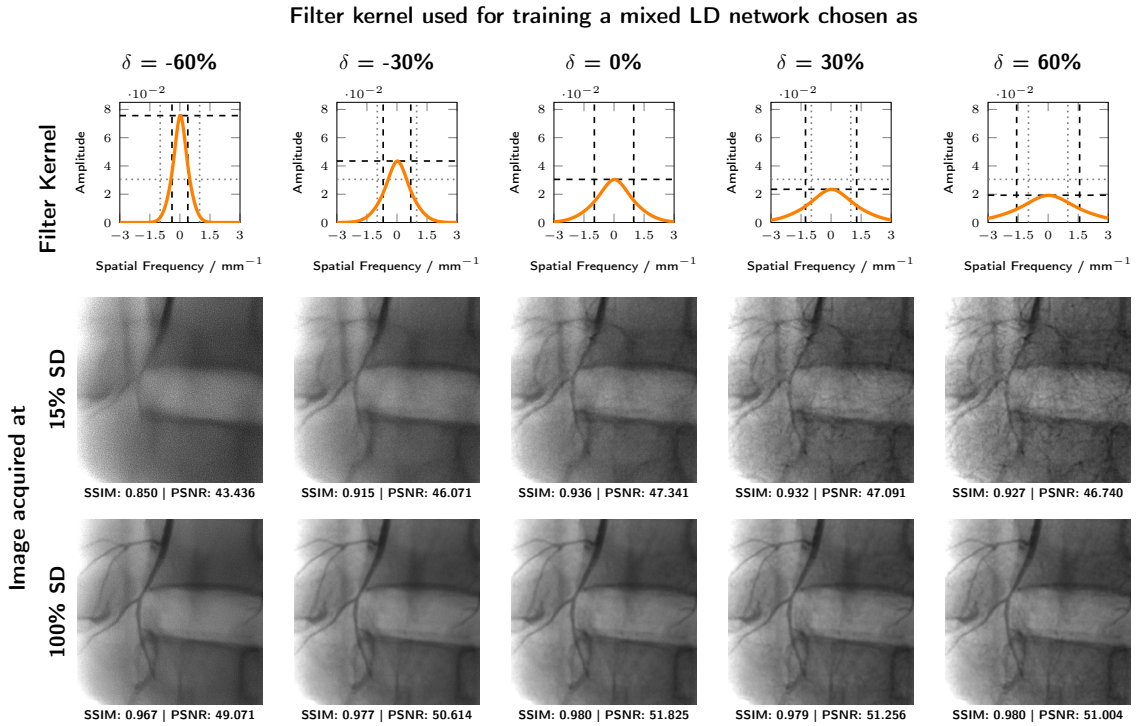


Figure 8: In the first row, we visualize the normalized filter kernels' cross-sections with different bandwidth parameters δ used to generate the power spectra of the noise added in generalized Anscombe transform (GAT) domain for training different mixed low-dose (LD) networks (one for each filter). Below are images showing the filters' impact on the resulting denoising network's performance for X-ray images acquired at 15% standard-dose (SD) and 100% SD (second column in figure 7). As characteristics of the filter kernels' δ, the maximum value (horizontal line(s), black dashed) and the full width at half maximum (FWHM, vertical line(s), black dashed) are shown and compared to the respective ones for $\delta = 0\%$ (horizontal and vertical lines, gray dotted). The values for the metrics structural similarity index (SSIM) and peak signal-to-noise-ratio (PSNR) are computed with respect to the GT shown in the first row of figure 7.

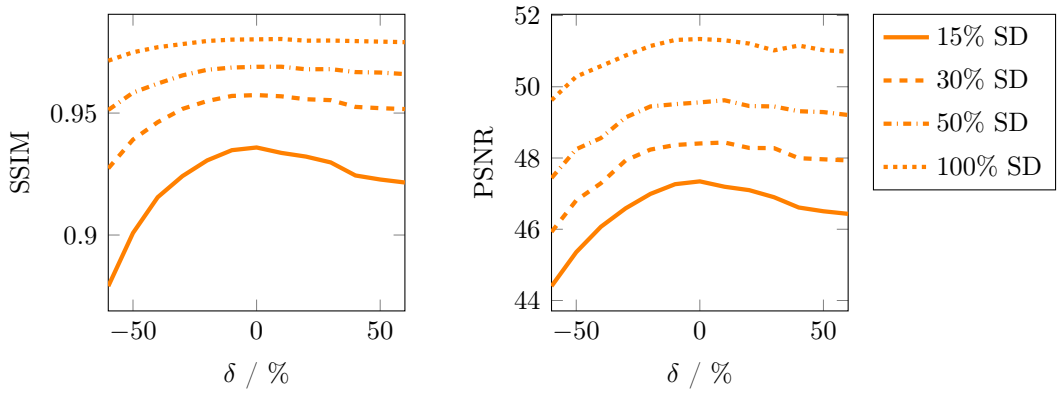


Figure 9: Quantitative evaluation of the kernel influence on the denoising performance for X-ray images acquired at different fractions of the standard-dose (SD) level based on structural similarity index (SSIM) and peak signal-to-noise-ratio (PSNR).

3.3. General Evaluation of the Method using Phantom and Clinical Images

In figure 10, we present the input as well as the processed regions of interests (ROIs) of clinical images acquired at 30% SD, 25% SD, 20% SD, 15% SD and 10% SD. Noise2Self, a method designed for AWGN, does not yield suitable results for all the cases. We have also found that in the case of clinical images, the differences in image quality between Low2High and the proposed method are not as pronounced as for the phantom images (in figure 1). Still, the proposed learning-based approach is superior as indicated by the CNR values. It does not display any visible artifacts (neither in clinical nor in phantom images), i.e., the background noise is low and, at the same time, anatomical structures and instruments appear sharp. In the clinical examples, it can also be seen that the proposed method and Low2High have improved the visibility of small vessels and instruments. This would, for example, help a physician in choosing a suitable path for navigating instruments and guide wires. In addition, the improved visibility of electrodes, catheters and their tips would be beneficial in electrophysiology and ablation procedures. However, when the proposed method is not applied correctly, it can result in sub-optimal denoising and even severe artifacts as shown in figure 11. This is also in line with the analysis made using the X-ray images of phantoms.

4. Discussion and Conclusion

In this paper, we have presented a robust learning-based X-ray image denoising approach. It uses a refined image acquisition model involving a blurring filter kernel to generate noise that is added when creating training data. The robust behavior of the method is also due to noise variance stabilization of the input data. By performing a quantitative evaluation of the influence of the dose levels associated with the training images, we have found that the proposed approach works well across multiple dose levels.

The qualitative evaluation of the influence of the dose levels used during training

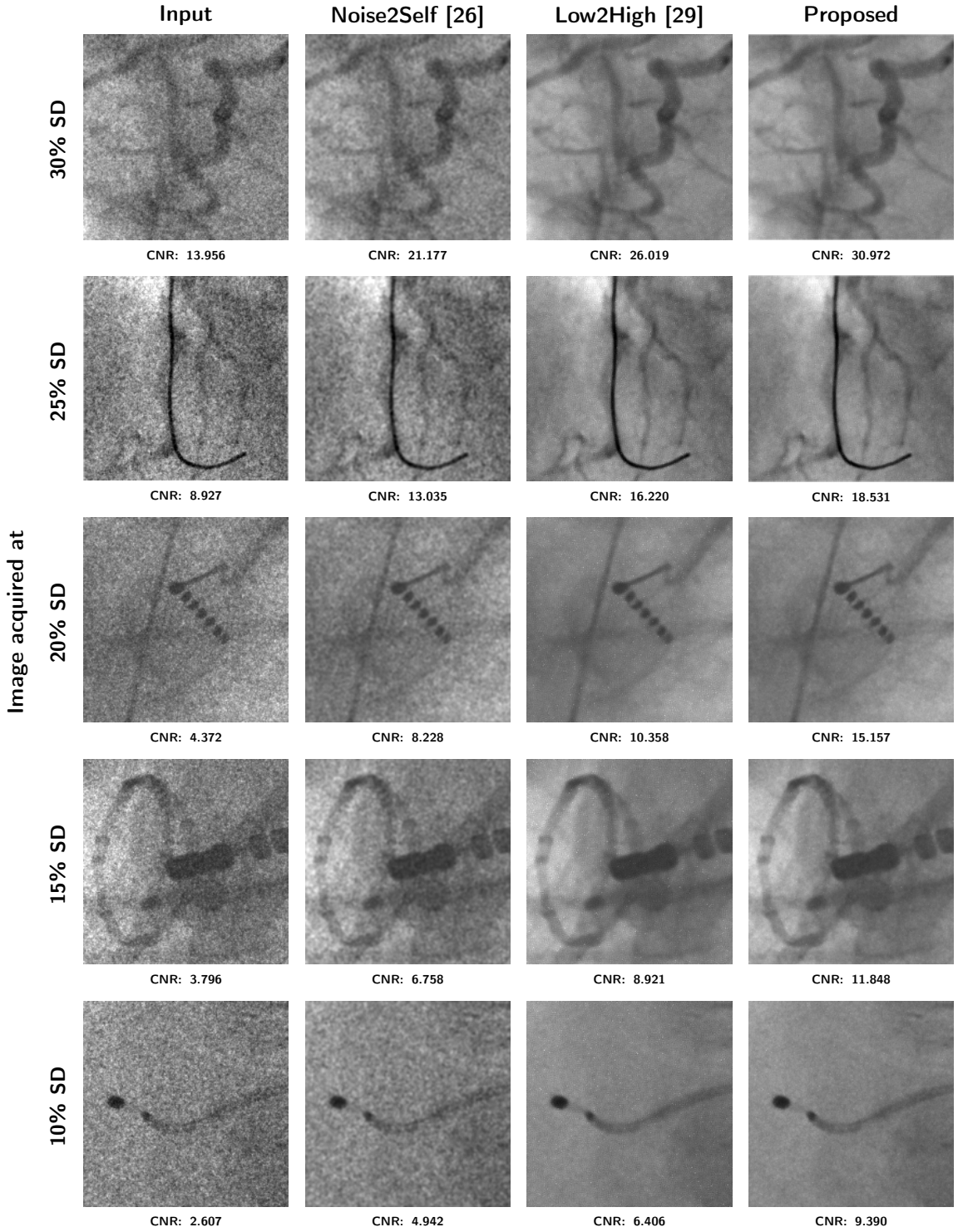


Figure 10: Visualization of selected regions of clinical images acquired at different fractions of the standard-dose (SD) level. In addition to the input data, the results of different denoising methods are shown. The results of the quantitative analysis with respect to contrast-to-noise-ratio (CNR) are also presented.

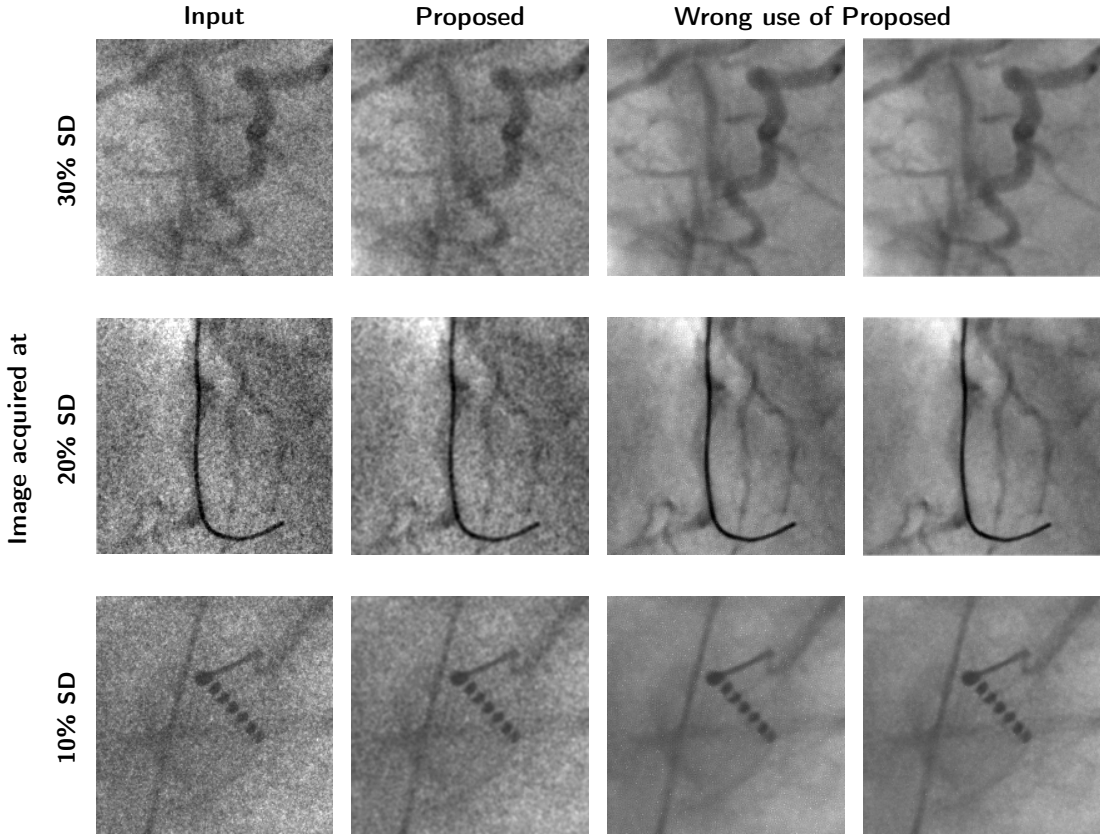


Figure 11: Visualization of selected regions of clinical images acquired at different fractions of the standard-dose (SD) level. In addition to the input data, the results of denoising using the proposed approach are presented. For this purpose, the denoising networks have been trained using the noise model associated with the input images (column 2) as well as different models (columns 3 and 4).

has shown that networks trained for a specific dose level have yielded superior results when applied on X-ray images acquired at the corresponding dose level. A network trained for a different dose level resulted in sub-optimal denoising performance (see figure 5). For example, we see the retention of noise after denoising when the denoisig network is trained on higher dose levels and applied on images acquired at lower dose levels. On the other hand, a network trained for a lower dose level has resulted in mild blurring when applied on images taken at higher dose levels.

In this context, we can recall our wavelet thresholding analogy and revisit (9), keeping in mind that the GAT normalizes the noise level to a fixed value σ, while the signal energy σ_X scales (non-linearly) with the dose. According to our analogy, the thresholds should be higher when the training of the network is performed with lower-dose inputs. This is because the values of the transform coefficients in the various sub-bands are smaller. If the network is then used for denoising higher-dose inputs, some of their signal parts will be strong enough to exceed the higher thresholds, i.e., those signals will remain unaffected. But there may also be instances where some signal details

are lost due to thresholding. Transform coefficients due to noise are, however, likely to be removed well. This yields well denoised results that may appear slightly blurred. If we train a denoising network with higher-dose input images, the thresholds will be smaller. This is because the numerator in (9) remains the same while the denominator increases. Hence, if we now use this network to denoise lower-dose images, for which higher thresholds would be more appropriate, there will be more noise left in the outputs as less of it gets thresholded away.

In the second part of our analysis, we have analyzed the impact of different noise models used during the training phase (see figure 8). If the training data features more low-frequency noise components compared to the actual data used during application, the trained network yields denoised images that are blurred. This can again be explained by applying the wavelet sub-band analogy together with (9). If our training data has more low frequency noise than expected, the thresholds in the associated low-frequency sub-bands will be higher. If we now denoise an image acquired at a lower dose featuring a regular frequency distribution using such a trained network, the signal components as well as the noise in these low-frequency sub-bands are likely to be removed. This may yield a very blurred result. In addition, as most of the noise in the training data is pushed into lower-frequency sub-bands, less noise will go to the higher frequency sub-bands. As a consequence, the thresholds in these sub-bands will be lower. The effect of this is that the denoised images will contain high-frequency noise. On the other hand, if the training data comprises more noise in the high frequency sub-bands than regular, there will be less noise in the lower-frequency sub-bands. This is because the amount of noise is normalized to a constant by the GAT. In such a case, the thresholds in the lower-frequency sub-bands will be too low, whereas they will be too high in the upper-frequency sub-bands. If we now process a lower dose X-ray image with normal frequency content using this network, we are left with too much noise in the lower-frequency sub-bands, while noise in the high-frequency noise will be removed significantly. Unfortunately, the left-over lower-frequency noise may resemble anatomical structures such as vessels. For input images with high enough dose, the various thresholds in the different sub-bands are less of a problem because their signal parts should in general be strong enough to avoid being eliminated by the thresholds. However, noise present in HD images will be removed considerably due to the normalization inherent in the GAT. A quantitative analysis of the results of the different trained networks – with respect to SSIM and PSNR – has revealed that the negative effects of using incorrect noise characteristics are particularly prominent at very LD levels, while the impact on higher dose levels is considerably lower. This analysis is in line with the explanation given above.

Extending our analogy to training with higher dose images would imply very small thresholds due to a fixed noise level σ and higher signal levels σ_X. In such a case, the thresholds will be rather small. Therefore, it may be possible to use less accurate noise models without adverse consequences. In other words, simple noise models, such as a Gaussian noise model [22] or uncorrelated Poisson noise [30], to simulate training data

may be acceptable at HD levels, whereas for the denoising of LD X-ray fluoroscopic images such models may not be suitable. As we have found that the outputs of our denoising networks are very sensitive to the noise models for LD inputs, we infer that a network trained for a particular imaging system using the proposed approach, may not be suitable for other systems with significantly different noise characteristics, particularly not for LD inputs.

To sum up, the proposed method is a robust learning-based denoising approach suitable for a range of X-ray dose levels, in particular very LD levels. For higher dose levels, the proposed method gives good results even when there is a mismatch in the noise characteristics used for training the denoising network and those present in the image on which denoising is performed. However, for lower dose levels, where only a limited number of X-ray photons are available for image generation, it is crucial to use an accurate noise model during the training phase of the proposed method to retain the underlying image fidelity without introducing any artifacts in the denoised images that might mimic anatomical structures. This indicates that networks trained for different imaging systems and imaging modes may be required during the inference phase to get the best results. On the contrary, the other learning-based denoising methods, Noise2Self [26] and Low2High [29] have under-performed. Since Noise2Self has been designed for uncorrelated AWGN, it is not surprising that it has problems with a situation it has not been developed for. On the other hand, as there is no noise variance stabilization transform during the application phase of Low2High, the method can introduce severe artifacts (as shown in figure 1). This happens when the imaging conditions (e.g., the system gain) used for training differ from those encountered during the application of the neural network. Though the results of the proposed method are promising for phantom and especially clinical images acquired at LD levels, there is a limit on the amount by which dose can be reduced. This is due to the fact that clinically important structures are likely to be missed at extreme LD levels. Hence a drastic reduction, e.g., below an X-ray dose of 10% SD, may be considered inadequate or too risky for some procedures while it may be acceptable for others. Therefore, further evaluations are needed.

Disclaimer: The concepts and results presented in this paper are based on research and are not commercially available.

Ethical Statement: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The clinical data used in the manuscript were acquired retrospectively.

References

[1] Richard Aufrichtig and David L Wilson. X-ray fluoroscopy spatio-temporal filtering with object detection. *IEEE Trans Med Imaging*, 14(4):733–746, 1995.

[2] Florian Luisier, Thierry Blu, and Michael Unser. A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Trans Image Process*, 16(3):593–606, 2007.

[3] S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, 9(9):1532–1546, 2000.

[4] Aleksandra Pizurica and Wilfried Philips. Estimating the probability of the presence of a signal of interest in multiresolution single-and multiband image denoising. *IEEE Transactions on image processing*, 15(3):654–665, 2006.

[5] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, volume 2, pages 60–65. IEEE, 2005.

[6] Pierrick Coupé, Pierre Hellier, Charles Kervrann, and Christian Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Trans Image Process*, 18(10):2221–2229, 2009.

[7] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans Image Process*, pages 2080–2095, 2007.

[8] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pages 2862–2869, 2014.

[9] Vladimir Katkovnik, Alessandro Foi, Karen Egiazarian, and Jaakko Astola. From local kernel to nonlocal multiple-model image denoising. *International journal of computer vision*, 86(1):1, 2010.

[10] Qiang Guo, Caiming Zhang, Yunfeng Zhang, and Hui Liu. An efficient svd-based method for image denoising. *IEEE Trans. Circuits Syst. Video Technol.*, 26(5):868–880, 2015.

[11] Priyam Chatterjee and Peyman Milanfar. Patch-based near-optimal image denoising. *IEEE Transactions on Image Processing*, 21(4):1635–1649, 2011.

[12] Wenzhao Zhao, Qiegen Liu, Yisong Lv, and Binjie Qin. Texture variation adaptive image denoising with nonlocal pca. *IEEE Transactions on Image Processing*, 28(11):5537–5551, 2019.

[13] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3d transform-domain collaborative filtering. In *European Signal Processing Conference*, volume 149. Tampere, Finland, 2007.

[14] Sai Gokul Hariharan, Norbert Strobel, Christian Kaethner, Markus Kowarschik, Stefanie Demirci, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. A photon recycling approach to the denoising of ultra-low dose x-ray sequences. *Int J Comput Assist Radiol Surg*, 13(6):847–854, 2018.

[15] Sai Gokul Hariharan, Christian Kaethner, Norbert Strobel, Markus Kowarschik, Julie DiNitto, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. Preliminary results of dsa denoising based on a weighted low-rank approach using an advanced neurovascular replication system. *Int J Comput Assist Radiol Surg*, pages 1–10, 2019.

[16] Curtis R Vogel and Mary E Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

[17] Joachim Weickert and Hanno Scharr. A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. *J Vis Commun Image Represent*, 13(1-2):103–118, 2002.

[18] Tommaso Cerciello, Paolo Bifulco, Mario Cesarelli, and Antonio Fratini. A comparison of denoising methods for x-ray fluoroscopic images. *Biomed Signal Process Control*, 7(6):550–559, 2012.

[19] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans Image Process*, 2017.

[20] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–

1595, 2017.

[21] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans Pattern Anal Mach Intell*, 39(6):1256–1272, 2017.

[22] Yevgen Matviychuk, Boris Mailhé, Xiao Chen, Qiu Wang, Atilla Kiraly, Norbert Strobel, and Mariappan Nadar. Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy. *Proc Int Conf Image Proc*, pages 2330–2334, 2016.

[23] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE Trans Med Imaging*, 36(12):2536–2545, 2017.

[24] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. *PMLR*, pages 2965–2974, 2018.

[25] Eunhee Kang, Hyun Jung Koo, Dong Hyun Yang, Joon Bum Seo, and Jong Chul Ye. Cycle-consistent adversarial denoising network for multiphase coronary ct angiography. *Med Phys*, 46(2):550–562, 2019.

[26] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019.

[27] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pages 2129–2137, 2019.

[28] JH Siewerdsen, LE Antonuk, Y El-Mohri, J Yorkston, W Huang, JM Boudry, and IA Cunningham. Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (amfpis) for diagnostic radiology. *Med Phys*, 24(1):71–89, 1997.

[29] Sai Gokul Hariharan, Christian Kaethner, Norbert Strobel, Markus Kowarschik, Shadi Albarqouni, Rebecca Fahrig, and Nassir Navab. Learning-based x-ray image denoising utilizing model-based image simulations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 549–557. Springer, 2019.

[30] Darwin T. Kuan, Alexander A. Sawchuk, Timothy C. Strand, and Pierre Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):165–177, 1985.

[31] Wouter JH Veldkamp, Lucia JM Kroft, Jan Pieter A van Delft, and Jacob Geleijns. A technique for simulating the effect of dose reduction on image quality in digital chest radiography. *J Digit Imaging*, 22(2):114–125, 2009.

[32] Amber J Gislason-Lee, Asli Kumcu, Stephen M Kengyelics, David S Brettle, Laura A Treadgold, Mohan Sivananthan, and Andrew G Davies. How much image noise can be added in cardiac x-ray imaging without loss in perceived image quality? *J Electron Imaging*, 24(5), 2015.

[33] Jean-Luc Starck, Fionn D Murtagh, and Albert Bijaoui. *Image processing and data analysis: the multiscale approach*. Cambridge University Press, 1998.

[34] Sai Gokul Hariharan, Norbert Strobel, Christian Kaethner, Markus Kowarschik, Rebecca Fahrig, and Nassir Navab. An analytical approach for the simulation of realistic low-dose fluoroscopic images. *Int J Comput Assist Radiol Surg*, pages 1–10, 2019.

[35] Kai Yang, Shih-Ying Huang, Nathan J Packard, and John M Boone. Noise variance analysis using a flat panel x-ray detector: A method for additive noise assessment with application to breast ct applications. *Med Phys*, 37(7):3527–3537, 2010.

[36] Sai Gokul Hariharan, Norbert Strobel, Christian Kaethner, Markus Kowarschik, Rebecca Fahrig, and Nassir Navab. Data-driven estimation of noise variance stabilization parameters for low-dose x-ray images. *Physics in Medicine & Biology*, 2020.

[37] Lucas Rodrigues Borges, Marcelo Andrade da Costa Vieira, and Alessandro Foi. Unbiased injection of signal-dependent noise in variance-stabilized range. *IEEE Signal Processing Letters*, 23(10):1494–1498, 2016.

[38] Andrew Kuhls-Gilcrist, Amit Jain, Daniel R Bednarek, Kenneth R Hoffmann, and Stephen Rudin.

Accurate mtf measurement in digital radiography using noise response. *Med Phys*, 37(2):724–735, 2010.

[39] Ian A Cunningham. Applied linear-systems theory. *Handbook of medical imaging*, 1:79–159, 2000.

[40] Lucas R Borges, Igor Guerrero, Predrag R Bakic, Alessandro Foi, Andrew DA Maidment, and Marcelo AC Vieira. Method for simulating dose reduction in digital breast tomosynthesis. *IEEE Trans Med Imaging*, 36(11):2331–2342, 2017.

[41] C Lanczos and B Gellai. Fourier analysis of random sequences. *Computers & Mathematics with Applications*, 1(3-4):269–276, 1975.

[42] S.G. Chang, Bin Yu, and M. Vetterli. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transactions on Image Processing*, 9(9):1522–1531, 2000.

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *MICCAI*, pages 234–241, 2015.

6.5. Summary

In this chapter, a novel X-ray imaging model-based noise simulation approach as well as a learning-based X-ray image denoising strategy have been presented. The training data for the learning-based denoising approach has been generated by simulating realistic low-dose X-ray images from their higher dose counterparts. In addition, the method uses a normalization step that takes into account imaging parameters such as the system gain and the electronic noise to stabilize the noise variance, as described in Chapter 4, to a known constant. The method delivers excellent results, especially at low-dose levels. A thorough evaluation of the method is also presented where the focus has been on understanding the cause of image degradation while utilizing such a learning-based denoising approach as well as identifying solutions to prevent the degradation.

Part III.

Summary and Outlook

X-ray dose must be kept as low as reasonably possible in order to reduce the risk of exposure to ionizing both for the patient (due to direct radiation) as well as for the clinical staff (due to scattered radiation). A consequence is that fewer X-ray photons are available for image formation resulting in lower image quality due to a reduced signal-to-noise ratio (SNR). The focus of this thesis has been on improving the image quality of low-dose X-ray images by developing advanced image processing techniques. The technical background required for the thesis is presented in Part I that contains Chapters 2 and 3. In Chapter 2, an introduction to medical X-ray imaging has been given. This includes a brief introduction to the physical principles of X-ray imaging as well as related clinical applications. In Chapter 3, an overview of the available denoising approaches has been provided. The contributions of this thesis have been presented in Chapters 4, 5 and 6.

Noise variance stabilization (NVS) is commonly used to pre-process images corrupted by signal-dependent noise prior to denoising as this allows for the application of denoising techniques designed for signals in which the noise can be considered as additive white Gaussian with a known variance. The generalized Anscombe transform (GAT) is a well-performing NVS technique to stabilize the noise variance to a known constant. To apply the GAT-based NVS, the underlying imaging parameters, i.e., the system gain and the electronic noise, are required. Unfortunately, it is difficult to derive the system gain directly from the system settings, because the X-ray spectrum received at the detector changes, e.g., because of beam hardening caused by the patient. As a solution to this problem, a data-driven method has been developed for extracting the system gain and the electronic noise. It is presented in Chapter 4. Across a range of low-dose x-ray settings, the approach has estimated both system gain and electronic noise level with an average error of only 4.2%.

Two analytical spatio-temporal X-ray image denoising methods have been presented in Chapter 5. The first method is a blind spatio-temporal denoising approach designed for ultra-low dose fluoroscopy sequences. The method includes the characteristics of noise to match patches, build a stack of temporally aligned patches and denoise them using low rank approximation. Taking into account the noise properties not only helps in reducing artifacts but also removes the necessity of computing a guide image, which is usually used in high-noise scenarios to improve patch matching. The evaluation of the approach on electrophysiology sequences has indicated that an improvement in the contrast-to-noise ratio (CNR) by a factor of around 3.5 can be achieved. In addition, X-ray image quality experts have found the improvement after denoising using the proposed method to be visually superior when compared to the well-established denoising approach video block matching 3D (V-BM3D). Most importantly, the experts have stated that the results resemble higher dose X-ray images. A minor drawback of the approach is that the denoised images may suffer from mild blurring around instrument edges due to imperfect matches caused by 3D object motion. In such cases, applying a low rank approximation will result in the averaging of pixels that may not be perfectly matched.

The second analytical spatio-temporal denoising approach presented in Chapter 5 makes use of a weighted low-rank approximation to denoise digital subtraction angiography (DSA) sequences. Since DSA sequences are inspected after acquisition, there are no real-time requirements. As a consequence, the method can utilize past and future frames when processing the current one. This approach also operates in the NVS domain and carefully estimates which pixels to include for denoising. The method results in an improvement in the CNR by a factor of 3.7, which is comparable to other state-of-the-art spatio-temporal

denoising approaches. For example, the first proposed method results in an improvement in the CNR by a factor of 3.6 when used in the context of DSA. However, a visual analysis suggests that only the use of the second proposed method yields images that share the characteristics of typical DSA images. This is because, the method makes use of constrained row-wise and column-wise low-rank approximations which do suffer from the smoothing artifacts introduced by a non-constrained low-rank approximation of misaligned patches. It should also be noted that in the case of pulsatile motion of the flowing contrast agent, temporal mismatches in the vessel regions are hard to avoid. However, since the proposed approach monitors how well patches match, it will automatically switch to in-plane denoising in such a case. As a consequence, the texture of the flowing contrast agent will be retained, albeit at the price of a somewhat higher noise level.

The proposed analytical denoising methods have the potential to improve significantly the image quality associated with low-dose X-ray images. The denoising methods have also been designed such that they can be deployed on massively parallel processing graphic processing units. Nevertheless, applying them on high-resolution images in real-time would require cost-performance trade-offs. As a solution to this problem, learning-based denoising algorithms based on convolutional neural networks CNNs are available that are significantly less expensive computationally during the application phase. However, such methods require pairs of associated high- and low-dose X-ray images during the training phase. Unfortunately, it is neither acceptable nor possible to acquire such image pairs during clinical cases due to the associated increase in ionizing radiation and patient motion. Hence, a learning-based spatial denoising strategy that involves the use of model-based simulations of realistic low-dose X-ray images during the training phase is introduced in Chapter. 6.

The noise simulation method introduced in Chapter 6 takes into account an X-ray image formation model and the corresponding system parameters used for imaging to derive a low-dose X-ray image from its higher-dose counterpart. Signal-dependent noise has been simulated using Gaussian noise in a noise variance stabilized domain and signal-independent noise has been injected in the image domain. The use of an image formation model along with the system parameters helps in reducing the extensive calibrations otherwise needed to construct look-up tables relating mean intensity values and their associated noise variance. Moreover, the method is easily adaptable to different imaging systems. The results of the Kolmogorov-Smirnov test at 5% significance level performed for 80 corresponding regions of the real and simulated low-dose images has suggested that the proposed method generates statistically similar regions at an average rate of 81.43%. The method has been further improved by taking into consideration the influence of detector blur when simulating the signal-dependent noise. This has been achieved by adding correlated (colored) Gaussian noise in an NVS domain (based on the GAT). The simulated low-dose images have then been used for developing a learning-based denoising algorithm.

The learning-based denoising algorithm introduced in Chapter 6 is based on transforming noise-corrupted instances of an image into another noise-corrupted instance of the same image, where the different instances have the same underlying noise characteristics. It is to be noted that the imaging model has not only been used for generating training data but also for normalizing the input to the network using the GAT so that the input to the network has noise with known variance, e.g., unit variance. This step ensures that the

network always receives inputs with specific noise characteristics, even if there is a change in the system gain. Due to the use of a large amount of simulated instances during training, the random nature of noise and the dependency of the output of the network only on the input, the network is forced to generate a noise-free instance in order to minimize the dissimilarity of the generated output to the target. A quantitative and a qualitative evaluation have indicated that the proposed learning-based denoising algorithm is robust across dose levels, and it has outperformed well-established analytical and learning-based spatial denoising approaches. In addition, the method has resulted in an image quality that is superior to that of the standard-dose images used for training. This indicates that the performance of the method is not limited by the quality of the data used during the training phase. Using this method, an X-ray image of size 1024×1024 can be processed using a mid range GPU NVIDIA RTX A4000 in around 8 milliseconds.

In the final part of Chapter 6, the potential pitfalls associated with such a learning-based denoising approach have been thoroughly investigated. The analysis has confirmed the assumption that it is important to rely on an accurate noise model during the training phase. A mismatch in the characteristics of noise present in the training data and the application data can yield images with blurring, noise or artifacts. It is important to understand the consequences of a mismatch as such a situation could arise when there is an unexpected change in the operating conditions of the imaging system that has not been accounted for. By drawing an analogy to sub-band thresholding approaches, it has been possible to get a better understanding of the denoising network and also predict the behavior of the network when there is a mismatch in the noise characteristics.

To sum up, three denoising approaches (two analytical and one learning-based) that are suitable for a range of dose levels, in particular very low-dose levels, have been proposed in this thesis. The analytical patch-based approaches, designed by taking into account imaging and noise characteristics, make use of spatio-temporal information. The methods reuse information, belonging to associated regions of interest, from past frames, to improve the current frame. This type of processing yields images that resemble higher dose X-ray images. These approaches, however, involve parameters that need to be carefully set in order to get suitable results. Moreover, as mentioned above, cost-performance trade-offs need to be made to satisfy real-time requirements. This is why a computationally inexpensive learning-based spatial denoising approach has been designed by accurately taking into account the noise characteristics associated with X-ray images. Unlike the analytical approaches, the learning-based approach performs spatial denoising of one image at a time. This means that if information is lost due to noise in the input images, it cannot be recovered after denoising. Nevertheless, the results have been found to be comparable to those of the analytical spatio-temporal denoising approaches, in spite of not utilizing temporal information.

A loss of information in the input images is a consequence of acquiring X-ray images at very low radiation dose levels, as only a limited number of X-ray photons are available for image generation. Even then the proposed methods, especially the learning-based approach, have retained the underlying spatial resolution without introducing any artifacts. However, at extreme low-dose levels, clinically important structures may be missed. Hence a drastic reduction, e.g., below an X-ray dose of 10% standard-dose, may be considered inadequate or too risky for some procedures while it may be acceptable for others. Therefore, further evaluations are needed to identify the dose saving potential of the pro-

posed algorithms for different clinical applications. The proposed denoising techniques not only protect patients and staff, but they also reduce the power requirements for the X-ray tubes. Hence it is possible to get better image quality under challenging circumstances which may arise when obese patients are imaged or when a steep angulation is used for imaging.

Furthermore, the proposed denoising methods have the potential to enable the clinical use of X-ray imaging applications that are considered noise-sensitive and dose-intensive. 2D dual energy DSA is such an application. Here, subsequent frames are acquired with different energy spectra, typically with tube voltages of 70 kVp and 120 kVp. The frames are then registered and a weighted subtraction is performed. The weights are chosen depending on the anatomy that needs to be visualized. For the case where a lower energy is used, the detector entrance dose may be low as low-energy photons are better absorbed by the body. As a result, the SNR in these images will be lower and the weighted subtraction will boost the noise. Therefore, future work will focus on evaluating the approach for such applications that would benefit from (and are limited by) low-dose X-ray imaging. In addition, clinical evaluations will be performed to determine the amount of dose saving possible with learning-based methods for existing applications. This would also give an insight into whether low-power and less expensive X-ray tubes that do not require sophisticated cooling solutions may be sufficient for certain applications. This will facilitate lighter and more mobile X-ray imaging devices making X-ray imaging even more widely available.

Acronyms

a-Se Amorphous selenium. 21

a-Si Amorphous silicon. 21

a-Si:H Hydrogenated amorphous silicon. 20

ALARA As low as reasonably achievable. 107

AWGN Additive white Gaussian noise. 28, 30

BM3D Block matching 3D. 31, 32

BN Batch normalization. 39–41

CMOS Complementary metal oxide semiconductor. 21

CNN Convolutional neural network. 35–37, 39–41

CNNs Convolutional neural networks. xi, 35–37, 40, 41, 46, 164

CNR Contrast-to-noise ratio. xii, 44, 46, 75, 90, 163, 164

CR Computed radiography. 19

CsI Caesium iodide. 19, 20

CT Computed tomography. 10, 14, 17, 18

DCT Discrete cosine transform. 30, 49, 50

DQE Detective quantum efficiency. 20, 23

DR Digital radiography. 19

DSA Digital subtraction angiography. 24, 75, 90, 163, 164, 166

ELU Exponential linear unit. 38

GANs Generative adversarial networks. 42

GAT Generalized Anscombe transform. ix, 49, 50, 163, 164

GELU Gaussian Error Linear Unit. 38

GOS Gadolinium oxiSulfate. 19

GPU Graphics processing unit. 35, 46, 165

Gy Gray. 8

keV Kilo electron volt. 11, 13, 15, 17, 18, 21

kVp Kilo volt peak. 11–14, 18

MAP Maximum a posterior. 28, 30

MTF Modulation transfer function. 23

MTFs Modulation transfer functions. 22

NL-means Non-local means. 31, 32

NLF Noise level function. 22, 23, 49, 50

NNM Nuclear norm minimization. 33, 34

NPS Noise power spectrum. 23, 24, 28

NVS Noise variance stabilization. 49, 73, 163, 164

PReLU Parametric rectified linear unit. 38

PSF Point spread function. 23, 36

PSNR Peak signal-to-noise ratio. xii, 36, 44–46

QE Quantum efficiency. 23

R Roentgen. 8

ReLU Rectified linear unit. 38, 40–43

SiLU Sigmoid linear unit. 38

SNR Signal-to-noise ratio. ix, 23, 26, 27, 30, 163, 166

SSIM Structural similarity index. xi, 41, 44, 46

Sv Sievert. 8

SVD Singular value decomposition. 33, 34

TFT Thin film transistor. 21

TFTs Thin film transistors. 20

Tl Thallium. 20

TV Total variation. 29

V-BM3D Video block matching 3D. 31, 32, 163

WNNM Weighted nuclear norm minimization. 34

Contributions

[1] S. G. Hariharan, N. Strobel, C. Kaethner, M. Kowarschik, R. Fahrig, and N. Navab, "Data-driven estimation of noise variance stabilization parameters for low-dose x-ray images," *Physics in Medicine & Biology*, vol. 65, no. 22, p. 225027, 2020. [Online]. Available: https://doi.org/10.1088/1361-6560/abbc82

[2] S. G. Hariharan, N. Strobel, C. Kaethner, M. Kowarschik, S. Demirci, S. Albarqouni, R. Fahrig, and N. Navab, "A photon recycling approach to the denoising of ultra-low dose X-ray sequences," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 6, pp. 847–854, 2018. [Online]. Available: https://doi.org/10.1007/s11548-018-1746-2

[3] S. G. Hariharan, C. Kaethner, N. Strobel, M. Kowarschik, J. DiNitto, S. Albarqouni, R. Fahrig, and N. Navab, "Preliminary results of DSA denoising based on a weighted low-rank approach using an advanced neurovascular replication system," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 7, pp. 1117–1126, 2019. [Online]. Available: https://doi.org/10.1007/s11548-019-01968-4

[4] S. G. Hariharan, N. Strobel, C. Kaethner, M. Kowarschik, R. Fahrig, and N. Navab, "An analytical approach for the simulation of realistic low-dose fluoroscopic images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 4, pp. 601–610, 2019. [Online]. Available: https://doi.org/10.1007/s11548-019-01912-6

[5] S. G. Hariharan, C. Kaethner, N. Strobel, M. Kowarschik, S. Albarqouni, R. Fahrig, and N. Navab, "Learning-based x-ray image denoising utilizing model-based image simulations," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Springer International Publishing, 2019, pp. 549–557.

[6] S. G. Hariharan, C. Kaethner, N. Strobel, M. Kowarschik, R. Fahrig, and N. Navab, "Robust learning-based x-ray image denoising—potential pitfalls, their analysis and solutions," *Biomedical Physics & Engineering Express*, vol. 8, no. 3, p. 035013, 2022. [Online]. Available: https://doi.org/10.1088/2057-1976/ac3489

[7] S. G. Hariharan, N. Strobel, M. Kowarschik, R. Fahrig, and N. Navab, "Simulation of realistic low dose fluoroscopic images from their high dose counterparts," in *Bildverarbeitung für die Medizin 2018*, A. Maier, T. M. Deserno, H. Handels, K. H. Maier-Hein, C. Palm, and T. Tolxdorff, Eds. Springer Berlin Heidelberg, 2018, pp. 80–85.

[8] S. G. Hariharan, C. Kaethner, N. Strobel, M. Kowarschik, J. DiNitto, R. Fahrig, and N. Navab, "Model-based motion artifact correction in digital subtraction angiography using optical-flow," in *Bildverarbeitung für die Medizin 2019*, H. Handels, T. M. Deserno,

A. Maier, K. H. Maier-Hein, C. Palm, and T. Tolxdorff, Eds. Springer Fachmedien Wiesbaden, 2019, pp. 146–151.

References

[9] N. R. Council, *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*. Washington, DC: The National Academies Press, 2006. [Online]. Available: https://nap.nationalacademies.org/catalog/11340/health-risks-from-exposure-to-low-levels-of-ionizing-radiation

[10] M. Andreucci, R. Solomon, and A. Tasanarong, "Side Effects of Radiographic Contrast Media: Pathogenesis, Risk Factors, and Prevention," *Biomed Res. Int.*, vol. 2014, p. 741018, 2014.

[11] M. Hensel, *Real-Time Noise Reduction of Medical X-Ray Image Sequences*, 1st ed. Books on Demand, 2008.

[12] M. Spahn, "Flat detectors and their clinical applications," *European Radiology*, vol. 15, no. 9, pp. 1934–1947, 2005.

[13] ——, "X-ray detectors in medical imaging," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 731, pp. 57–63, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900213007961

[14] S. L. Fritz, S. E. Mirvis, S. Osher Pais, and S. Roys, "Phantom evaluation of angiographer performance using low frame rate acquisition fluoroscopy," *Medical Physics*, vol. 15, no. 4, pp. 600–603, 1988. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.596211

[15] D. L. Parker, P. D. Clayton, L. R. Tarbox, and P. L. VonBehren, "Optimal dose utilization with variable x-ray intensity in digital radiography," in *Application of optical instrumentation in medicine XI*, vol. 419. International Society for Optics and Photonics, 1983, pp. 102–110.

[16] T. S. Curry, J. E. Dowdey, and R. C. Murry, *Christensen's physics of diagnostic radiology*, 4th ed. Lippincott Williams & Wilkins, 1990.

[17] O. Dössel, *Die bildgebenden Verfahren in der Medizin*, 2nd ed. Springer Vieweg Berlin, Heidelberg, 2014.

[18] R. L. Van Metter, H. L. Kundel, and J. Beutel, *Handbook of medical imaging*. SPIE Press, 2000, vol. 1.

[19] A. Oppelt, *Imaging Systems for Medical Diagnostics: Fundamentals, Technical Solutions and Applications for Systems Applying Ionizing Radiation, Nuclear Magnetic Resonance and Ultrasound*, 2nd ed. Publicis Corporate Publishing, 2005.

[20] W. C. Röntgen, "Ueber eine neue art von strahlen," *Annalen der Physik*, vol. 300, no. 1, pp. 12–17, 1898. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18983000103

[21] Wilhelm Conrad Roentgen - Facts. Nobel Media AB 2014. [Online]. Available: http://www.nobelprize.org/nobel_prizes/physics/laureates/1901/rontgen-facts.html

[22] Charles Glover Barkla - Facts. Nobel Media AB 2014. [Online]. Available: http://www.nobelprize.org/nobel_prizes/physics/laureates/1917/barkla-facts.html

[23] A. Assmus, "Early history of x rays," *Beam Line*, vol. 25, no. 2, pp. 10–24, 1995.

[24] G. Meggitt, *Taming the Rays-A History of Radiation and Protection*. Pitchpole Books, 2018.

[25] G. F. Jost, K. S. Yonemura, and R. A. von Jako, *Image Guidance in Spine Surgery*. New York, NY: Springer New York, 2014, pp. 613–628. [Online]. Available: https://doi.org/10.1007/978-1-4614-7657-3_47

[26] S. Vavilov, G. Koshkin, V. Udod, and O. Fofanof, "Analytical approximation of spectrum for pulse x-ray tubes," *Journal of Physics: Conference Series*, vol. 671, p. 012061, 2016. [Online]. Available: https://doi.org/10.1088/1742-6596/671/1/012061

[27] M. G. Tsagareli, "Ivane Tarkhnishvili (Tarchanoff): A Major Georgian Figure from the Russian Physiological School," *Journal of the History of the Neurosciences*, vol. 21, no. 4, pp. 393–408, 2012.

[28] A. Pusey, "Roentgen-rays in the treatment of skin diseases and for the removal of hair," *Archives of Dermatology*, vol. 119, no. 2, pp. 162–175, 1983. [Online]. Available: https://doi.org/10.1001/archderm.1983.01650260070022

[29] R. L. Kathren and P. L. Ziemer, *Health Physics, a backward glance*. Pergamon Press, 1980.

[30] The Nobel Prize in Physiology or Medicine 1946. Nobel Media AB 2014. [Online]. Available: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1946/

[31] O. W. Richardson, "The electrical conductivity imparted to a vacuum by hot conductors," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 201, pp. 497–549, 1903. [Online]. Available: http://www.jstor.org/stable/90907

[32] After Class: X-Rays October 1960 Popular Electronics. RF Cafe 2014. [Online]. Available: http://www.rfcafe.com/references/popular-electronics/x-rays-october-1960-popular-electronics.htm

[33] B. R. David, H. Barschdorf, V. Doormann, R. Eckart, G. Harding, J.-P. Schlomka, A. Thran, P. Bachmann, and P. Flisikowski, "Liquid-metal anode x-ray tube," in *Laser-Generated and Other Laboratory X-Ray and EUV Sources, Optics, and Applications*,

G. A. Kyrala, J.-C. J. Gauthier, C. A. MacDonald, and A. M. Khounsary, Eds., vol. 5196, International Society for Optics and Photonics. SPIE, 2004, pp. 432–443. [Online]. Available: https://doi.org/10.1117/12.504503

[34] H. Hahn, D. Färber, H. Allmendinger, and J. Brendler, "Grid-controlled fluoroscopy in paediatric radiology," *medicamundi*, vol. 41, pp. 12–17, 1997.

[35] D. Stueve, "Management of pediatric radiation dose using Philips fluoroscopy systems DoseWise: perfect image, perfect sense," *Pediatric radiology*, vol. 36, no. 2, pp. 216–220, 2006.

[36] P. Allisy-Roberts and J. Williams, "Chapter 6 - fluoroscopy," in *Farr's Physics for Medical Imaging*, 2nd ed. W.B. Saunders, 2008, pp. 91–102. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780702028441500107

[37] C. Martin, "The importance of radiation quality for optimisation in radiology," *Biomedical imaging and intervention journal*, vol. 3, no. 2, 2007. [Online]. Available: https://doi.org/10.2349/biij.3.2.e38

[38] E. Haschek and O. T. Lindenthal, "Ein Beitrag zur praktischen Verwerthung der Photographie nach Röntgen," *Wiener Klinische Wochenschrift*, vol. 9, pp. 63–64, 1896.

[39] E. D. Osborne, C. G. Sutherland, A. J. Scholl, and L. G. Rowntree, "Roentgenography of urinary tract during excretion of sodium iodid," *Journal of the American Medical Association*, vol. 80, no. 6, pp. 368–373, 1923. [Online]. Available: https://doi.org/10.1001/jama.1923.02640330004002

[40] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. Siewerdsen, "Technical note: Spektr 3.0 - a computational tool for x-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8, pp. 4711–4717, 2016.

[41] D. Pacella, "Energy-resolved x-ray detectors: the future of diagnostic imaging," *Reports in Medical Imaging*, vol. 8, pp. 1–13, 2015. [Online]. Available: https://doi.org/10.2147/RMI.S50045

[42] H. Aichinger, J. Dierker, S. Joite-Barfuß, and M. Säbel, *Radiation exposure and image quality in X-ray diagnostic radiology: physical principles and clinical applications*, 2nd ed. Springer, 2004. [Online]. Available: https://doi.org/10.1007/978-3-642-11241-6

[43] C. E. Floyd, R. J. Warp, J. T. Dobbins, H. G. Chotas, A. H. Baydush, R. Vargas-Voracek, and C. E. Ravin, "Imaging characteristics of an amorphous silicon flat-panel detector for digital chest radiography," *Radiology*, vol. 218, no. 3, pp. 683–688, 2001. [Online]. Available: https://doi.org/10.1148/radiology.218.3.r01fe45683

[44] J. Siewerdsen, L. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, J. Boudry, and I. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (amfpis) for diagnostic radiology," *Medical physics*, vol. 24, no. 1, pp. 71–89, 1997.

[45] J. Chabbal, C. Chaussat, T. Ducourant, L. Fritsch, J. Michailos, V. Spinnler, G. Vieux, M. Arques, G. Hahm, M. Hoheisel, H. Horbaschek, R. F. Schulz, and M. F. Spahn, "Amorphous silicon x-ray image sensor," in *Medical Imaging 1996: Physics of Medical Imaging*, R. L. V. Metter and J. Beutel, Eds., vol. 2708, International Society for Optics and Photonics. SPIE, 1996, pp. 499–510. [Online]. Available: https://doi.org/10.1117/12.237812

[46] M. Nikl, "Scintillation detectors for x-rays," *Measurement Science and Technology*, vol. 17, no. 4, pp. R37–R54, 2006. [Online]. Available: https://doi.org/10.1088/0957-0233/17/4/r01

[47] T. Yamazaki, T. Tamura, M. Nokita, S. Okada, S. Hayashida, and Y. Ogawa, "Performance of a novel 43-cm x 43-cm flat-panel detector with CsI:Tl scintillator," in *Medical Imaging 2004: Physics of Medical Imaging*, M. J. Yaffe and M. J. Flynn, Eds., vol. 5368, International Society for Optics and Photonics. SPIE, 2004, pp. 379–385. [Online]. Available: https://doi.org/10.1117/12.534010

[48] A. Konstantinidis, "Evaluation of digital x-ray detectors for medical imaging applications," Ph.D. dissertation, UCL (University College London), 2011.

[49] L. E. Antonuk, Y. El-Mohri, J. H. Siewerdsen, J. Yorkston, W. Huang, V. E. Scarpine, and R. A. Street, "Empirical investigation of the signal performance of a high-resolution, indirect detection, active matrix flat-panel imager (amfpi) for fluoroscopic and radiographic operation," *Medical Physics*, vol. 24, no. 1, pp. 51–70, 1997. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.597918

[50] J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, and I. A. Cunningham, "Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology," *Medical Physics*, vol. 25, no. 5, pp. 614–628, 1998. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.598243

[51] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 1997.

[52] I. A. Cunningham, *Handbook of medical imaging*. SPIE Press, 2000, vol. 1, ch. Applied linear-systems theory, pp. 79–159.

[53] P. P. Kawathekar and K. J. Karande, "Severity analysis of osteoarthritis of knee joint from x-ray images: A literature review," in *2014 International Conference on Signal Propagation and Computer Technology*, 2014, pp. 648–652.

[54] C. Pineda, R. Espinosa, and A. Pena, "Radiographic imaging in osteomyelitis: The role of plain radiography, computed tomography, ultrasonography, magnetic resonance imaging, and scintigraphy," *Seminars in Plastic Surgery*, vol. 23, no. 2, pp. 80–89, 2009.

[55] Z. Kundu, "Classification, imaging, biopsy and staging of osteosarcoma," *Indian journal of orthopaedics*, vol. 48, pp. 238–46, 2014.

[56] D. Krakow and D. L. Rimoin, "The skeletal dysplasias," *Genetics in Medicine*, vol. 12, pp. 327–341, 2010. [Online]. Available: https://doi.org/10.1097/GIM.0b013e3181daae9b

[57] L. Cardinale, G. Volpicelli, A. Lamorte, J. Martino, and A. Veltri, "Revisiting signs, strengths and weaknesses of standard chest radiography in patients of acute dyspnea in the emergency department," *Journal of Thoracic Disease*, vol. 4, no. 4, 2012.

[58] M. Hamon, *Catheter-Based Cardiovascular Interventions: A Knowledge-Based Approach*. Springer Berlin Heidelberg, 2013, ch. Transradial Access for Percutaneous Coronary Interventions, pp. 393–402.

[59] P. Lanzer, *Catheter-Based Cardiovascular Interventions: A Knowledge-Based Approach*. Springer Berlin Heidelberg, 2013, ch. Interventional X-Ray Coronary Angiography, pp. 279–301.

[60] V. Z. Taenzer, *Contrast media in urography, angiography and computerized tomography*. Thieme-Stratton Corp, 1983.

[61] J. Racadio, K. Strauss, T. Abruzzo, M. Patel, K. Kukreja, N. Johnson, M. den Hartog, B. Hoornaert, and R. Nachabe, "Significant dose reduction for pediatric digital subtraction angiography without impairing image quality: Preclinical study in a piglet model," *American Journal of Roentgenology*, vol. 203, no. 4, pp. 904–908, 2014. [Online]. Available: https://doi.org/10.2214/AJR.13.12170

[62] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale modeling & simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[63] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Networks*, vol. 131, pp. 251–275, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608020302665

[64] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, 2019. [Online]. Available: https://doi.org/10.1186/s42492-019-0016-7

[65] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice Hall, 2007.

[66] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, Ltd, 2011.

[67] I. Pitas and A. N. Venetsanopoulos, "Nonlinear digital filters: principles and applications," in *The Springer International Series in Engineering and Computer Science*. Springer Science & Business Media, 2013. [Online]. Available: https://doi.org/10.1007/978-1-4757-6017-0

[68] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 591–604, 1995.

[69] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*, 1998, pp. 839–846.

[70] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," in *Doklady Akademii Nauk*, vol. 151, no. 3, 1963, pp. 501–504. [Online]. Available: https://doi.org/10.1007/978-0-387-92920-0_3

[71] A. N. Tikhonov, A. Goncharsky, V. Stepanov, and A. G. Yagola, *Numerical methods for the solution of ill-posed problems*. Springer Dordrecht, 1995, vol. 328. [Online]. Available: https://doi.org/10.1007/978-94-015-8480-7

[72] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[73] J. Weickert, *Anisotropic diffusion in image processing*. Teubner Stuttgart, 1998, vol. 1.

[74] C. Tsiotsios and M. Petrou, "On the choice of the parameters for anisotropic diffusion in image processing," *Pattern Recognition*, vol. 46, no. 5, pp. 1369–1381, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320312004748

[75] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM Journal on Numerical analysis*, vol. 29, no. 1, pp. 182–193, 1992. [Online]. Available: http://www.jstor.org/stable/2158083

[76] S. Esedoglu and S. J. Osher, "Decomposition of images by the anisotropic rudin-osher-fatemi model," *Communications on Pure and Applied Mathematics*, vol. 57, no. 12, pp. 1609–1626, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20045

[77] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016727899290242F

[78] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *2006 International Conference on Image Processing*, 2006, pp. 2633–2636.

[79] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[80] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proceedings of 1st International Conference on Image Processing*, vol. 1, 1994, pp. 31–35.

[81] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996. [Online]. Available: https://www.osti.gov/biblio/218524

[82] Y. Lou, T. Zeng, S. Osher, and J. Xin, "A weighted difference of anisotropic and isotropic total variation model for image processing," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1798–1823, 2015. [Online]. Available: https://doi.org/10.1137/14098435X

[83] M. Zibulevsky and M. Elad, "L1-l2 optimization in signal and image processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 76–88, 2010.

[84] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[85] L. Zhang, P. Bao, and X. Wu, "Multiscale LMMSE-based image denoising with optimal wavelet selection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 4, pp. 469–481, 2005.

[86] T. Blu and F. Luisier, "The SURE-LET Approach to Image Denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, 2007.

[87] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[88] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 654–665, 2006.

[89] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994. [Online]. Available: http://www.jstor.org/stable/2337118

[90] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[91] S. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical wiener filtering," in *SPIE Technical Conference on Wavelet Applications in Signal Processing*, 1997.

[92] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1322–1331, 2001.

[93] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[94] K. Siedenburg and M. Doerfler, "Persistent time-frequency shrinkage for audio denoising," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 29–38, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16665

[95] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, 2005, pp. 60–65.

[96] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Processing On Line*, vol. 2, pp. 175–213, 2012. [Online]. Available: https://doi.org/10.5201/ipol.2012.l-bm3d

[97] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *2007 15th European Signal Processing Conference*, 2007, pp. 145–149.

[98] I. Rish and G. Grabarnik, *Sparse modeling: theory, algorithms, and applications*. CRC Press, 2014.

[99] M. Lebrun and A. Leclaire, "An implementation and detailed analysis of the K-SVD image denoising algorithm," *Image Processing On Line*, vol. 2, pp. 96–133, 2012. [Online]. Available: https://doi.org/10.5201/ipol.2012.llm-ksvd

[100] X. Deng and Z. Liu, "Image denoising based on steepest descent OMP and K-SVD," in *2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2015, pp. 1–5.

[101] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[102] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[103] L. Zhang and W. Zuo, "Image restoration: From sparse and low-rank priors to deep priors [lecture notes]," *IEEE Signal Processing Magazine*, vol. 34, no. 5, pp. 172–179, 2017.

[104] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320309003677

[105] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 2272–2279.

[106] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally Centralized Sparse Representation for Image Restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[107] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1791–1798.

[108] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust Video Restoration by Joint Sparse and Low Rank Matrix Approximation," *SIAM Journal on Imaging Sciences*, vol. 4, no. 4, pp. 1122–1142, 2011. [Online]. Available: https://doi.org/10.1137/100817206

[109] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

[110] A. Eriksson and A. van den Hengel, "Efficient computation of robust weighted low-rank matrix approximations using the l˙1 norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1681–1690, 2012.

[111] M. Bertalmío, *Denoising of photographic images and video: Fundamentals, open challenges and new trends*. Springer Cham, 2018.

[112] Q. Guo, C. Zhang, Y. Zhang, and H. Liu, "An Efficient SVD-Based Method for Image Denoising," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2016.

[113] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *International Journal of Computer Vision*, vol. 121, no. 2, pp. 183–208, 2017.

[114] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted Nuclear Norm Minimization with Application to Image Denoising," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[115] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust Recovery of Subspace Structures by Low-Rank Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[116] J.-F. Cai, E. J. Candès, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010. [Online]. Available: https://doi.org/10.1137/080738970

[117] Q. Guo, C. Zhang, Y. Zhang, and H. Liu, "An efficient svd-based method for image denoising," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2016.

[118] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[119] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran

Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[120] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[121] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary CT angiography," *Medical Physics*, vol. 46, no. 2, pp. 550–562, 2019. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13284

[122] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2774–2781.

[123] S. Roth and M. J. Black, "Fields of Experts," *International Journal of Computer Vision*, vol. 82, no. 2, p. 205, 2009.

[124] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[125] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[126] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018. [Online]. Available: https://doi.org/10.1007/s13244-018-0639-9

[127] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, pp. 1–21, 2015. [Online]. Available: https://doi.org/10.1186/s40537-014-0007-7

[128] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[129] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Springer International Publishing, 2015, pp. 234–241.

[130] Q. Yan and W. Wang, "DCGANs for image super-resolution, denoising and deblurring," *Advances in Neural Information Processing Systems*, pp. 487–495, 2017.

[131] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986. [Online]. Available: https://doi.org/10.1038/323533a0

[132] U. Michelucci, *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*, 1st ed. Apress Berkeley, CA, 2019. [Online]. Available: https://doi.org/10.1007/978-1-4842-4976-5

[133] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016. [Online]. Available: http://www.deeplearningbook.org

[134] A. W. Trask, *Grokking Deep Learning*. Manning Publications, 2019.

[135] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," *arXiv preprint arXiv:1606.08415*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1606.08415

[136] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv preprint arXiv:1511.07289*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1511.07289

[137] D. Hendrycks and K. Gimpel, "Bridging nonlinearities and stochastic regularizers with gaussian error linear units," *Computing Research Repository*, vol. abs/1606.08415, 2016. [Online]. Available: http://arxiv.org/abs/1606.08415

[138] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *Computing Research Repository*, vol. abs/1603.07285, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1603.07285

[139] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper/2018/file/905056c1ac1dad141560467e0a99e1cf-Paper.pdf

[140] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. PMLR, 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html

[141] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger, "Understanding Batch Normalization," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[142] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010. [Online]. Available: http://jmlr.org/papers/v11/vincent10a.html

[143] M. Prakash, M. Lalit, P. Tomancak, A. Krul, and F. Jug, "Fully unsupervised probabilistic noise2void," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 154–158.

[144] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 2019, pp. 524–533. [Online]. Available: https://proceedings.mlr.press/v97/batson19a.html

[145] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void - Learning Denoising From Single Noisy Images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2124–2132.

[146] Y. Matviychuk, B. Mailhé, X. Chen, Q. Wang, A. Kiraly, N. Strobel, and M. Nadar, "Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2330–2334.

[147] T. Wang, M. Sun, and K. Hu, "Dilated Deep Residual Network for Image Denoising," in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017, pp. 1272–1279.

[148] W. Bae, J. Yoo, and J. C. Ye, "Beyond Deep Residual Learning for Image Restoration: Persistent Homology-Guided Manifold Simplification," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1141–1149.

[149] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, "Medical image denoising using convolutional neural network: a residual learning approach," *The Journal of Supercomputing*, vol. 75, no. 2, pp. 704–718, 2019. [Online]. Available: https://doi.org/10.1007/s11227-017-2080-0

[150] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A Persistent Memory Network for Image Restoration," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4549–4557.

[151] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral Image Restoration via Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 667–682, 2019.

[152] M. P. Heinrich, M. Stille, and T. M. Buzug, "Residual U-Net Convolutional Neural Network Architecture for Low-Dose CT Denoising," *Current Directions in Biomedical Engineering*, vol. 4, no. 1, pp. 297–300, 2018. [Online]. Available: https://doi.org/10.1515/cdbme-2018-0072

[153] W. Liu and J. Lee, "A 3-d atrous convolution neural network for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5701–5715, 2019.

[154] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral Image Denoising Employing a Spatial–Spectral Deep Residual Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205–1218, 2019.

[155] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Low-dose CT Denoising with Dilated Residual Network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5117–5120.

[156] Y. Wang, G. Wang, C. Chen, and Z. Pan, "Multi-scale dilated convolution of convolutional neural network for image denoising," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19 945–19 960, 2019. [Online]. Available: https://doi.org/10.1007/s11042-019-7377-y

[157] Y. Su, Q. Lian, X. Zhang, B. Shi, and X. Fan, "Multi-scale cross-path concatenation residual network for Poisson denoising," *IET Image Processing*, vol. 13, no. 8, pp. 1295–1303, 2019.

[158] O. Sheremet, K. Sheremet, O. Sadovoi, and Y. Sokhina, "Convolutional Neural Networks for Image Denoising in Infocommunication Systems," in *2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, 2018, pp. 429–432.

[159] K. Uchida, M. Tanaka, and M. Okutomi, "Non-blind image restoration based on convolutional neural network," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, 2018, pp. 40–44.

[160] A. Zarshenas and K. Suzuki, "Deep Neural Network Convolution for Natural Image Denoising," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 2534–2539.

[161] C. Chen, Z. Xiong, X. Tian, and F. Wu, "Deep Boosting for Image Denoising," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[162] R. Couturier, G. Perrot, and M. Salomon, "Image Denoising Using a Deep Encoder-Decoder Network with Skip Connections," in *International Conference on Neural Information Processing*, vol. 11306. Springer, 2018, pp. 554–565. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02182820

[163] Y. Chen, M. Yu, G. Jiang, Z. Peng, and F. Chen, "End-to-end single image enhancement based on a dual network cascade model," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 284–295, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320319301415

[164] P. Xiao, Y. Guo, and P. Zhuang, "Removing stripe noise from infrared cloud images via deep convolutional networks," *IEEE Photonics Journal*, vol. 10, no. 4, pp. 1–14, 2018.

[165] A. Abbasi, A. Monadjemi, L. Fang, H. Rabbani, and Y. Zhang, "Three-Dimensional Optical Coherence Tomography Image Denoising through Multi-Input Fully-Convolutional Networks," *Computers in biology and medicine*, vol. 108, pp. 1–8, 2019.

[166] C. Chen and Z. Xu, "Aerial-Image Denoising Based on Convolutional Neural Network with Multi-Scale Residual Learning Approach," *Information*, vol. 9, no. 7, p. 169, 2018. [Online]. Available: https://www.mdpi.com/2078-2489/9/7/169

[167] W. Jeon, W. Jeong, K. Son, and H. Yang, "Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks," *Optics letters*, vol. 43, no. 17, pp. 4240–4243, 2018. [Online]. Available: http://opg.optica.org/ol/abstract.cfm?URI=ol-43-17-4240

[168] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," *Japanese Journal of Radiology*, vol. 36, no. 9, pp. 566–574, 2018.

[169] A. Panda, R. Naskar, and S. Pal, "Exponential linear unit dilated residual network for digital image denoising," *Journal of Electronic Imaging*, vol. 27, no. 5, p. 053024, 2018.

[170] S. A. Priyanka and Y.-K. Wang, "Fully Symmetric Convolutional Network for Effective Image Denoising," *Applied Sciences*, vol. 9, no. 4, p. 778, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/4/778

[171] R. Aljadaany, D. K. Pal, and M. Savvides, "Proximal Splitting Networks for Image Restoration," in *Image Analysis and Recognition*, Springer. Springer International Publishing, 2019, pp. 3–17.

[172] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, 2003, pp. 1398–1402.

[173] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[174] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[175] J. Chen, J. Chen, H. Chao, and M. Yang, "Image Blind Denoising with Generative Adversarial Network Based Noise Modeling," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3155–3164.

[176] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," vol. 80, pp. 2965–2974, 2018. [Online]. Available: https://proceedings.mlr.press/v80/lehtinen18a.html

[177] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-Quality Self-Supervised Deep Image Denoising," *Advances in Neural Information Processing Systems*, vol. 32, pp. 6970–6980, 2019.

[178] H. S. Park, J. Baek, S. K. You, J. K. Choi, and J. K. Seo, "Unpaired Image Denoising Using a Generative Adversarial Network in X-Ray CT," *IEEE Access*, vol. 7, pp. 110 414–110 425, 2019.

[179] D.-W. Kim, J. R. Chung, and S.-W. Jung, "GRDN:Grouped Residual Dense Network for Real Image Denoising and GAN-Based Real-World Noise Modeling," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2086–2094.

[180] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2016.

[181] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[182] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[183] J.-L. Starck, F. D. Murtagh, and A. Bijaoui, *Image processing and data analysis: the multiscale approach*. Cambridge University Press, 1998.

[184] N. Ponomarenko, V. Lukin, M. Zriakhov, A. Kaarna, and J. Astola, "An automatic approach to lossy compression of aviris images," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 472–475.

[185] M. Colom and A. Buades, "Analysis and Extension of the Ponomarenko et al. Method, Estimating a Noise Curve from a Single Image," *Image Processing On Line*, vol. 3, pp. 173–197, 2013. [Online]. Available: https://doi.org/10.5201/ipol.2013.45

[186] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008.

[187] S. Pyatykh and J. Hesser, "Image Sensor Noise Parameter Estimation by Variance Stabilization and Normality Assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3990–3998, 2014.

[188] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D Image Denoising with Shape-Adaptive Principal Component Analysis," in *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, 2009. [Online]. Available: https://hal.inria.fr/inria-00369582

[189] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3d transform-domain collaborative filtering," in *2007 15th European Signal Processing Conference*, 2007, pp. 145–149.

[190] H. Bogunović and S. Lončarić, *Denoising of Time-Density Data in Digital Subtraction Angiography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1157–1166.

[191] K. Niu, Y. Li, S. Schafer, K. Royalty, Y. Wu, C. Strother, and G.-H. Chen, "Ultra low radiation dose digital subtraction angiography (DSA) imaging using low rank constraint," in *Medical Imaging 2015: Physics of Medical Imaging*, C. Hoeschen and D. Kontos, Eds., vol. 9412. SPIE, 2015, pp. 419–424. [Online]. Available: https://doi.org/10.1117/12.2082070

[192] N. Menger, T. Elsässer, G.-H. Chen, and M. Manhart, "Noise reduction in low dose dsa imaging using pixel adaptive svd-based approach," in *Bildverarbeitung für die Medizin 2017*, K. H. Maier-Hein, geb. Fritzsche, T. M. Deserno, geb. Lehmann, H. Handels, and T. Tolxdorff, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 37–42.

[193] Y. Matviychuk, B. Mailhé, X. Chen, Q. Wang, A. Kiraly, N. Strobel, and M. Nadar, "Learning a multiscale patch-based representation for image denoising in x-ray fluoroscopy," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2330–2334.

[194] C. Amiot, C. Girard, J. Chanussot, J. Pescatore, and M. Desvignes, "Spatio-temporal multiscale denoising of fluoroscopic sequence," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1565–1574, 2016.

[195] C. L. Chan, B. J. Sullivan, A. V. Sahakian, A. K. Katsaggelos, S. Swiryn, D. C. Hueter, and T. Frohlich, "Simulation of quantum mottle in digital angiographic images," in *Biomedical Image Processing*, vol. 1245. SPIE, 1990, pp. 104–111.

[196] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 165–177, 1985.

[197] W. J. Veldkamp, L. J. Kroft, J. P. A. van Delft, and J. Geleijns, "A technique for simulating the effect of dose reduction on image quality in digital chest radiography," *Journal of Digital Imaging*, vol. 22, no. 2, pp. 114–125, 2009.

[198] A. J. Gislason-Lee, A. Kumcu, S. M. Kengyelics, D. S. Brettle, L. A. Treadgold, M. Sivananthan, and A. G. Davies, "How much image noise can be added in cardiac x-ray imaging without loss in perceived image quality?" *Journal of Electronic Imaging*, vol. 24, no. 5, pp. 1–14, 2015. [Online]. Available: https://doi.org/10.1117/1.JEI.24.5.051006

[199] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.