A gentle introduction to deep learning in medical image processing

Andreas Maier[1,*], **Christopher Syben**[1], **Tobias Lasser**[2], **Christian Riess**[1]

[1] Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[2] Technical University of Munich, Germany

Received 4 October 2018; accepted 21 December 2018

Abstract

This paper tries to give a gentle introduction to deep learning in medical image processing, proceeding from theoretical foundations to applications. We first discuss general reasons for the popularity of deep learning, including several major breakthroughs in computer science. Next, we start reviewing the fundamental basics of the perceptron and neural networks, along with some fundamental theory that is often omitted. Doing so allows us to understand the reasons for the rise of deep learning in many application domains. Obviously medical image processing is one of these areas which has been largely affected by this rapid progress, in particular in image detection and recognition, image segmentation, image registration, and computer-aided diagnosis. There are also recent trends in physical simulation, modeling, and reconstruction that have led to astonishing results. Yet, some of these approaches neglect prior knowledge and hence bear the risk of producing implausible results. These apparent weaknesses highlight current limitations of deep ()learning. However, we also briefly discuss promising approaches that might be able to resolve these problems in the future.

Keywords: Introduction, Deep learning, Machine learning, Image segmentation, Image registration, Computer-aided diagnosis, Physical simulation, Image reconstruction

1 Introduction

Over the recent years, Deep Learning (DL) [1] has had a tremendous impact on various fields in science. It has lead to significant improvements in speech recognition [2] and image recognition [3], it is able to train artificial agents that beat human players in Go [4] and ATARI games [5], and it creates artistic new images [6,7] and music [8]. Many of these tasks were considered to be impossible to be solved by computers before the advent of deep learning, even in science fiction literature.

Obviously this technology is also highly relevant for medical imaging. Various introductions to the topic can be found in the literature ranging from short tutorials and reviews [9–18] over blog posts and jupyter notebooks [19–21] to entire books [22–25]. All of them serve a different purpose and offer a different view on this quickly evolving topic. A very good review paper is for example found in the work of Litjens et al. [12], as they did the incredible effort to review more than 300 papers in their article. Since then, however, many more noteworthy works have appeared – almost on a daily basis – which makes it difficult to create a review paper that matches the current pace in the field. The newest effort to summarize the entire field was attempted in [26] listing more than 350 papers. Again, since its publication several more noteworthy works appeared and others were missed. Hence, it is important to select methods of significance and describe them in high detail. Zhou et al. [22] do so for the state-of-the-art of deep learning in medical image analysis and found an excellent selection of topics. Still, deep learning is being quickly adopted in other fields of medical

* Corresponding author at: Friedrich-Alexander-University Erlangen-Nuremberg, Pattern Recognition Lab,Martensstr. 3, 91058 Erlangen, Germany.
E-mail: andreas.maier@fau.de (A. Maier).

image processing and the book misses, for example, topics such as image reconstruction. While an overview on important methods in the field is crucial, the actual implementation is as important to move the field ahead. Hence, works like the short tutorial by Breininger et al. [20] are highly relevant to introduce to the topic also on a code-level. Their jupyter notebook framework creates an interactive experience in the web browser to implement fundamental deep learning basics in Python. In summary, we observe that the topic is too complex and evolves too quickly to be summarized in a single document. Yet, over the past few months there already have been so many exciting developments in the field of medical image processing that we believe it is worthwhile to point them out and to connect them to a single introduction.

Readers of this article do not have to be closely acquainted with deep learning at its terminology. We will summarize the relevant theory and present it at a level of detail that is sufficient to follow the major concepts in deep learning. Furthermore, we connect these observations with traditional concepts in pattern recognition and machine learning. In addition, we put these foundations into the context of emerging approaches in medical image processing and analysis, including applications in physical simulation and image reconstruction. As a last aim of this introduction, we also clearly indicate potential weaknesses of the current technology and outline potential remedies.

2 Materials and methods

2.1 Introduction to machine learning and pattern recognition

Machine learning and pattern recognition essentially deal with the problem of automatically finding a decision, for example, separating apples from pears. In traditional literature [27], this process is outlined using the pattern recognition system (cf. Fig. 1). During a training phase, the so-called *training data set* is *preprocessed* and meaningful *features* are extracted. While the preprocessing is understood to remain in the original space of the data and comprised operations such as noise reduction and image rectification, feature extraction is facing the task to determine an algorithm that would be able to extract a distinctive and complete feature representation, for example, color or length of the semi-axes of a surrounding ellipse for our apples and pears example. This task is truly difficult to generalize, and it is necessary to design such features anew essentially for every new application. In the deep learning literature, this process is often also referred to as "hand-crafting" features. Based on the feature vector $\boldsymbol{x} \in \mathbb{R}^n$, the *classifier* has to predict the correct *class* y, which is typically estimated by a function $\hat{y} = \hat{f}(\boldsymbol{x})$ that directly results in the classification result $\hat{y}$. The classifier's parameter vector $\boldsymbol{\theta}$ is determined during the training phase and later evaluated on an independent *test data set*.

2.2 Neural networks

In this context, we can now follow neural networks and associated methods in their role as classifiers. The fundamental unit of a neural network is a neuron, it takes a bias w_0 and a weight vector $\boldsymbol{w} = (w_1, \ldots, w_n)$ as parameters $\boldsymbol{\theta} = (w_0, \ldots, w_n)$ to model a decision

$$\hat{f}(\boldsymbol{x}) = h(\boldsymbol{w}^\top \boldsymbol{x} + w_0) \tag{1}$$

using a non-linear activation function $h(x)$. Hence, a single neuron itself can already be interpreted as a classifier, if the activation function is chosen such that it is monotonic, bounded, and continuous. In this case, the maximum and the minimum can be interpreted as a decision for the one or the other class. Typical representatives for such activation functions in classical literature are the sign function $\text{sign}(x)$ resulting in Rosenblatt's perceptron [28], the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$, or the tangens hyperbolicus $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. (cf. Fig. 5). A major disadvantage of individual neurons is that they only allow to model linear decision boundaries, resulting in the well known fact that they are not able to solve the *XOR* problem. Fig. 2 summarizes the considerations towards the computational neuron graphically.

In combination with other neurons, modeling capabilities increase dramatically. Arranged in a *single layer*, it can already be shown that neural networks can approximate any continuous function $f(\boldsymbol{x})$ on a compact subset of $\mathbb{R}^n$ [29]. A single layer network is conveniently summarized as a linear combination of N individual neurons

$$\hat{f}(\boldsymbol{x}) = \sum_{i=0}^{N-1} v_i h(\boldsymbol{w}_i^\top \boldsymbol{x} + w_{0,i}) \tag{2}$$

using combination weights v_i. All trainable parameters of this network can be summarized as

$$\boldsymbol{\theta} = (v_0, w_{0,0}, \boldsymbol{w}_0, \ldots, v_N, w_{0,N}, \boldsymbol{w}_N)^\top.$$

The difference between the true function $f(\boldsymbol{x})$ and its approximation $\hat{f}(\boldsymbol{x})$ is bounded by

$$|f(\boldsymbol{x}) - \hat{f}(\boldsymbol{x})| < \epsilon, \tag{3}$$

where ϵ decreases with increasing N for activation functions that satisfy the criteria that we mentioned earlier (monotonicity, boundedness, continuity) [30]. Hence, given a large number of neurons, *any function can be approximated using a single layer network only*. Note that the approximation will only be valid for samples that are drawn from the same compact set on which the network was trained. As such, an additional practical requirement for an approximation is that the training set is *representative* and future observations will be similar. At first glance, this contradicts all recent

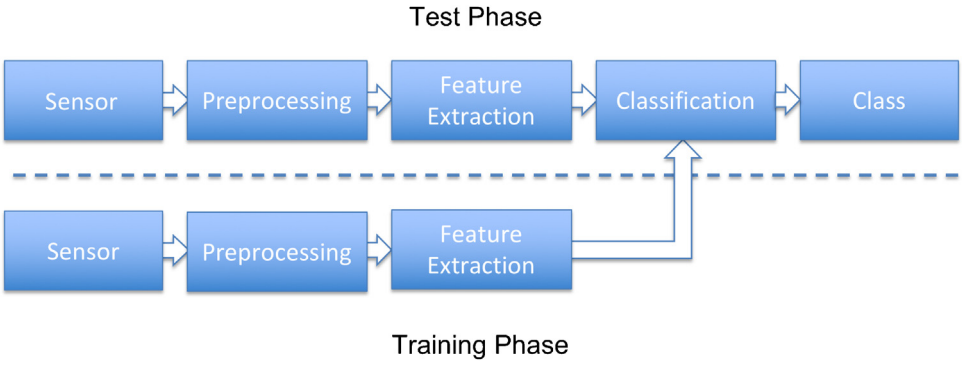

Figure 1. Schematic of the traditional pattern recognition pipeline used for automatic decision making. Sensor data is preprocessed and "hand-crafted" features are extracted in training and test phase. During training a classifier is trained that is later used in the test phase to decide the class automatically (after [27]).

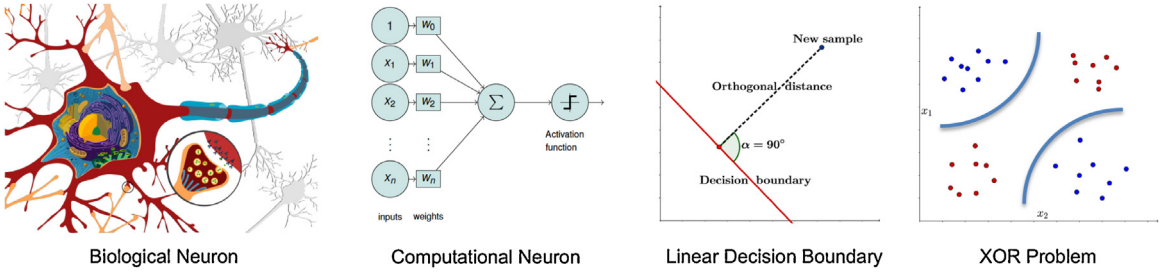

Figure 2. Neurons are inspired by biological neurons shown on the left. The resulting computational neuron computes a weighted sum of its inputs which is then processed by an activation function $h(x)$ to determine the output value (cf. Fig. 5). Doing so, we are able to model linear decision boundaries, as the weighted sum can be interpreted as a signed distance to the decision boundary, while the activation determines the actual class membership. On the right-hand side, the XOR problem is shown that cannot be solved by a single linear classifier. It typically requires either curved boundaries or multiple lines.

developments in deep learning and therefore requires additional attention.

In the literature, many arguments are found why a deep structure has benefits for feature representation, including the argument that by recombination of the weights along the different paths through the network, features may be re-used exponentially [31]. Instead of summarizing this long line of arguments, we look into a slightly simpler example that is summarized graphically in Fig. 3. Decision trees are also able to describe general decision boundaries in $\mathbb{R}^n$. A simple example is shown on the top left of the figure, and the associated partition of a two-dimensional space is shown below, where black indicates class $y=1$ and white $y=0$. According to the universal approximation theorem, we should be able to map this function into a single layer network. In the center column, we attempt to do so using the inner nodes of the tree and their inverses to construct a six neuron basis. In the bottom of the column, we show the basis functions that are constructed at every node projected into the input space, and the resulting network's approximation, also shown in the input space. Here, we chose the output weights to minimize $||y-\hat{y}||_2$. As can be seen in the result, not all areas can be recovered correctly. In fact, the maximal error ϵ is close to 0.7 for a function that is bounded by 0 and 1. In order to improve this approximation, we can choose to introduce a second layer. As shown in the right column, we can choose the strategy to map all inner nodes to a first layer and all leaf nodes of the tree to a second layer. Doing so effectively encodes every partition that is described by the respective leaf node in the second layer. This approach is able to map our tree correctly with $\epsilon=0$. In fact, this approach is general, holds for all decision trees, and was already described by Ivanova et al. in 1995 [32]. As such, we can now understand why deeper networks may have more modeling capacity.

2.3 Network training

Having gained basic insights into neural networks and their basic topology, we still need to discuss how its parameters $\boldsymbol{\theta}$ are actually determined. The answer is fairly easy: gradient descent. In order to compute a gradient, we need to define a function that measures the quality of our parameter set $\boldsymbol{\theta}$, the so-called *loss function* $L(\boldsymbol{\theta})$. In the following, we will work with simple examples for loss functions to introduce the concept of *back-propagation*, which is the algorithm that

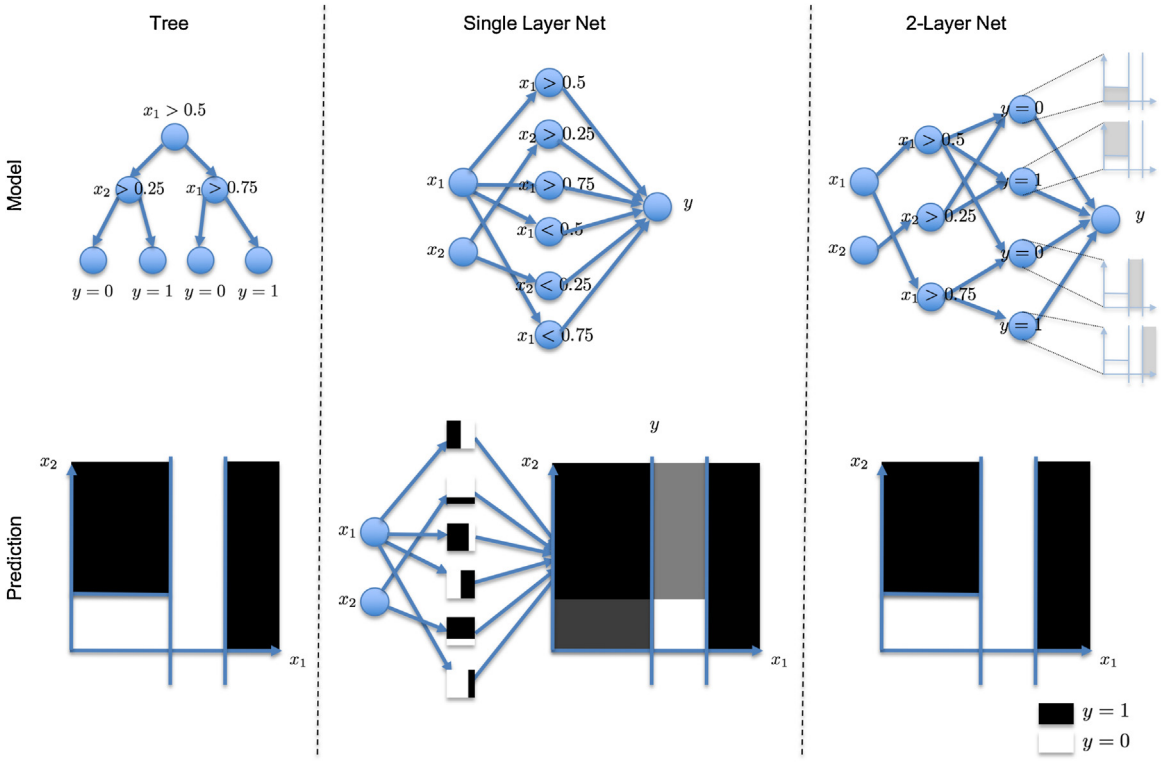

Figure 3. A decision tree allows to describe any partition of space and can thus model any decision boundary. Mapping the tree into a one-layer network is possible. Yet, there still is significant residual error in the resulting function. In the center example, $\epsilon \approx 0.7$. In order to reduce this error further, a higher number of neurons would be required. If we construct a network with one node for every inner node in the first layer and one node for every leaf node in the second layer, we are able to construct a network that results in $\epsilon = 0$.

is commonly used to efficiently compute gradients for neural network training.

We can represent a single-layer fully connected network with linear activations simply as $\hat{\boldsymbol{y}} = \hat{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x}$, i.e., a matrix multiplication. Note that the network's output is now multidimensional with $\hat{\boldsymbol{y}}$, $\boldsymbol{y} \in \mathbb{R}^m$. Using an L2-loss, we end up with the following objective function:

$$L(\boldsymbol{\theta}) = \frac{1}{2}||\hat{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{y}||_2^2 = \frac{1}{2}||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y}||_2^2. \tag{4}$$

In order to update the parameters $\boldsymbol{\theta} = \boldsymbol{W}$ in this example, we need to compute

$$\frac{\partial L}{\partial \boldsymbol{W}} = \underbrace{\frac{\partial L}{\partial \hat{\boldsymbol{f}}}}_{(\boldsymbol{W}\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{x}^\top)} \underbrace{\frac{\partial \hat{\boldsymbol{f}}}{\partial \boldsymbol{W}}} = (\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y})(\boldsymbol{x}^\top) \tag{5}$$

using the chain rule. Note that $\cdot$ indicates the operator's side, as matrix vector multiplications generally do not commute. The final weight update is then obtained as

$$\boldsymbol{W}^{j+1} = \boldsymbol{W}^j + \eta(\boldsymbol{W}^j\boldsymbol{x} - \boldsymbol{y})\boldsymbol{x}^\top, \tag{6}$$

where η is the so-called *learning rate* and j is used to index the iteration number.

Now, let us consider a slightly more complicated network structure with three layers $\hat{\boldsymbol{y}} = \hat{\boldsymbol{f}}_3(\hat{\boldsymbol{f}}_2(\hat{\boldsymbol{f}}_1(\boldsymbol{x}))) = \boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}$, again using linear activations. This yields the following objective function:

$$L(\boldsymbol{\theta}) = \frac{1}{2}||\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x} - \boldsymbol{y}||_2^2. \tag{7}$$

Note that this example is academic, as $\boldsymbol{\theta} = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3\}$ could simply be collapsed to a single matrix. Yet, the concept that we use to derive this gradient is generally applicable also to non-linear functions. Computing the gradient with respect to the parameters of the last layer $\boldsymbol{W}_3$ follows the same recipe as in the previous network:

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{W}_3} &= \underbrace{\frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}}_{(\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x})^\top} \underbrace{\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \boldsymbol{W}_3}} \\ &= (\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x} - \boldsymbol{y})(\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x})^\top.\end{aligned} \tag{8}$$

For the computation of the gradient with respect to the second layer $\boldsymbol{W}_2$, we already need to apply the chain rule twice:

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{W}_2} &= \frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \boldsymbol{W}_2} = \underbrace{\frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}}_{(\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}-\boldsymbol{y})}\underbrace{\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \hat{\boldsymbol{f}}_2}}_{(\boldsymbol{W}_3)^\top}\cdot\underbrace{\frac{\partial \hat{\boldsymbol{f}}_2}{\partial \boldsymbol{W}_2}}_{\cdot\,(\boldsymbol{W}_1\boldsymbol{x})^\top} \\ &= \boldsymbol{W}_3^\top(\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{W}_1\boldsymbol{x})^\top.\end{aligned} \tag{9}$$

Which leads us to the input layer gradient that is determined as

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{W}_1} &= \frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \boldsymbol{W}_1} = \frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \hat{\boldsymbol{f}}_2}\frac{\partial \hat{\boldsymbol{f}}_2}{\partial \boldsymbol{W}_1} \\ &= \underbrace{\frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}}_{(\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}-\boldsymbol{y})}\underbrace{\frac{\partial \hat{\boldsymbol{f}}_3}{\partial \hat{\boldsymbol{f}}_2}}_{(\boldsymbol{W}_3)^\top}\cdot\underbrace{\frac{\partial \hat{\boldsymbol{f}}_2}{\partial \hat{\boldsymbol{f}}_1}}_{(\boldsymbol{W}_2)^\top}\cdot\underbrace{\frac{\partial \hat{\boldsymbol{f}}_1}{\partial \boldsymbol{W}_1}}_{\cdot\,(\boldsymbol{x})^\top} \\ &= \boldsymbol{W}_2^\top\boldsymbol{W}_3^\top(\boldsymbol{W}_3\boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{x})^\top.\end{aligned} \tag{10}$$

The matrix derivatives above are also visualized graphically in Fig. 4. Note that many intermediate results can be reused during the computation of the gradient, which is one of the reasons why back-propagation is efficient in computing updates. Also note that the forward pass through the net is part of $\frac{\partial L}{\partial \hat{\boldsymbol{f}}_3}$, which is contained in all gradients of the net. The other partial derivatives are only partial derivatives either with respect to the input or the parameters of the respective layer. Hence, back-propagation can be used if both operations are known for every layer in the net. Having determined the gradients, each parameter can now be updated analogous to Eq. (6).

2.4 Deep learning

With the knowledge summarized in the previous sections, networks can be constructed and trained. However, deep learning is not possible. One important element was the establishment of additional activation functions that are displayed in Fig. 5. In contrast to classical bounded activations like $\text{sign}(x)$, $\sigma(x)$, and $\tanh(x)$, the new functions such as the *Rectified Linear Unit*

$$\text{ReLU}(x) = \begin{cases} x & \text{if } \; x \geq 0 \\ 0 & \text{else,} \end{cases}$$

and many others, of which we only mention the *Leaky ReLU*

$$\text{LReLU}(x) = \begin{cases} x & \text{if } \; x \geq 0 \\ \alpha x & \text{else,} \end{cases}$$

were identified to be useful to train deeper networks. Contrary to the classical activation functions, many of the new activation functions are *convex* and have large areas with non-zero derivatives. As can be seen in Eq. (10), the computation of the gradient of deeper layers using the chain rule requires several multiplications of partial derivatives. The deeper the net, the more multiplications are required. If several elements along this chain are smaller than 1, the entire gradient decays exponentially with the number of layers. Hence, non-saturating derivatives are important to solve numerical issues, which were the reasons why *vanishing gradients* did not allow training of networks that were much deeper than about three layers. Also note that each neuron does not loose its interpretation as a classifier, if we consider 0 as the classification boundary. Furthermore, the universal approximation theorem still holds for a single-layer network with ReLUs [33]. Hence, several useful and desirable properties are attained using such modern activation functions.

One disadvantage is, of course, that the ReLU is not differentiable over the entire domain of x. At $x = 0$ a kink is found that does not allow to determine a unique gradient. For optimization, an important property of the gradient of a function is that it will point towards the direction of the steepest ascent. Hence, following the negative direction will allow minimization of the function. For a differentiable function, this direction is unique. If this constraint is relaxed to allow multiple directions that lead to an extremum, we arrive at sub-gradient theory [34]. It allows us to still use gradient descent algorithms to optimize such problems, if it is possible to determine a *sub-gradient*, i.e., at least one instance of a valid direction towards the optimum. For the ReLU, any value between 0 and -1 would be acceptable at $x = 0$ for the descent operation. If such a direction can be obtained, convergence is guaranteed for convex problems by application of specific optimization programs, such as using a fixed step size in the gradient descent [35]. This allows us to remain with back-propagation for optimization, while using non-differentiable activation functions.

Another significant advance towards deep learning is the use of specialized layers. In particular, the so-called *convolution* and *pooling layers* enable to model locality and abstraction (cf. Fig. 6). The major advantage of the convolution layers is that they only consider a local neighborhood for each neuron, and that all neurons of the same layer share the same weights, which dramatically reduces the amount of parameters and therefore memory required to store such a layer. These restrictions are identical to limiting the matrix multiplication to a matrix with circulant structure, which exactly models the operation of convolution. As the operation is generally of the form of a matrix multiplication, the gradients introduced in Section 2.3 still apply. *Pooling* is an operation that is used to reduce the scale of the input. For images, typically areas of 2×2 or 3×3 are analyzed and summarized to a single value. The average operation can again be expressed as a matrix with hard-coded weights, and gradient computation follows essentially the previous section. Non-linear operations, such as maximum or median, however, require more attention. Again, we can exploit the sub-gradient approach. During the forward pass through the net, the maximum or median can easily be determined. Once this is known,

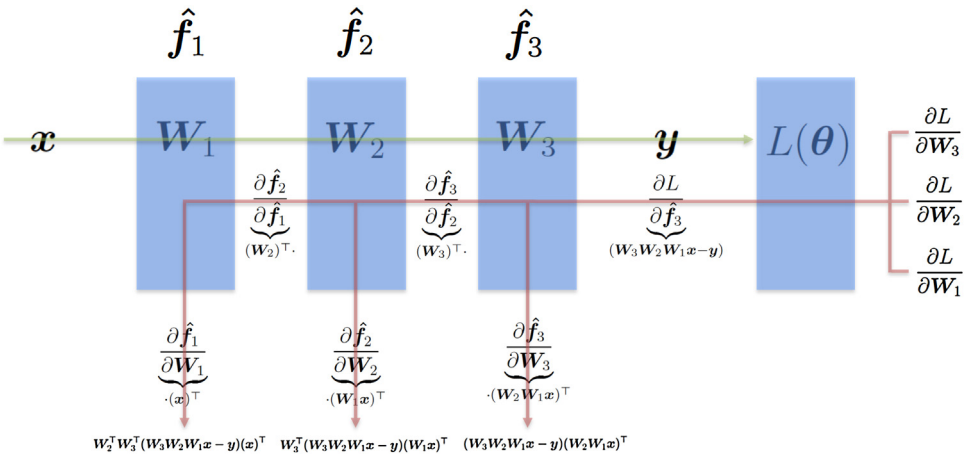

Figure 4. Graphical overview of back-propagation using layer derivatives. During the forward pass, the network is evaluated once and compared to the desired output using the loss function. The back-propagation algorithm follows different paths through the layer graph in order to compute the matrix derivatives efficiently.

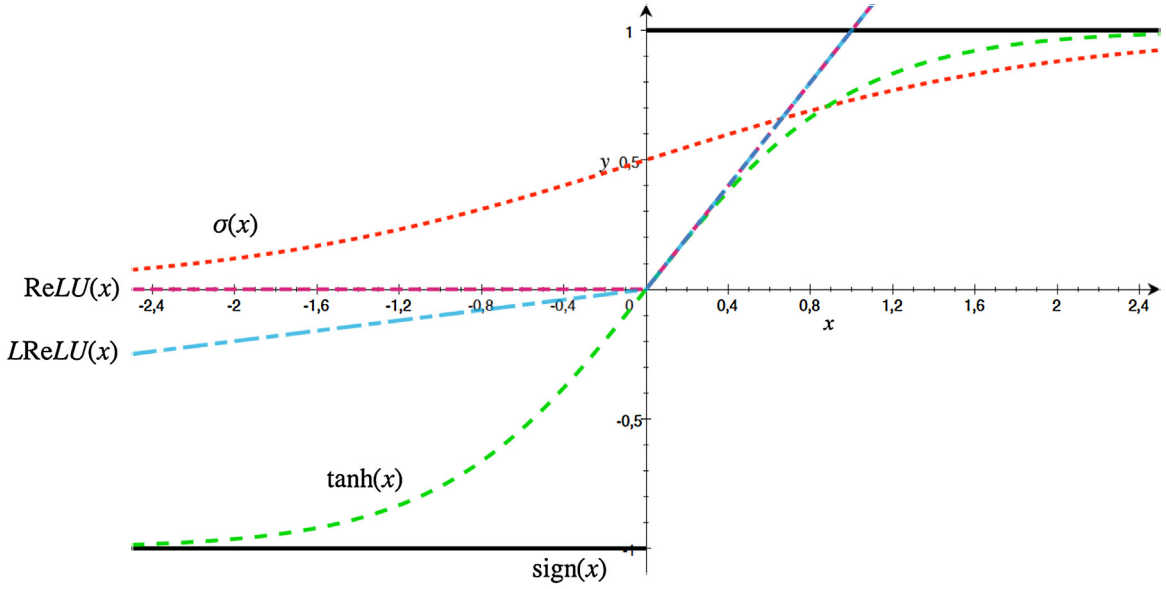

Figure 5. Overview of classical (sign(x), $\sigma(x)$, and tanh(x)) and modern activation functions, like the Rectified Linear Unit ReLU(x) and the leaky ReLU LReLU(x).

a matrix is constructed that simply selects the correct elements that would also have been selected by the non-linear methods. The transpose of the same matrix is then employed during the backward pass to determine an appropriate sub-gradient [36]. Fig. 6 shows both operations graphically and highlights an example for a convolutional neural network (CNN). If we now compare this network with Fig. 1, we see that the original interpretation as only a classifier is no longer valid. Instead, the deep network now models all steps directly from the signal up to the classification stage. Hence, many authors claim that feature "hand-crafting" is no longer required because everything is learned by the network in a data-driven manner.

So far, deep learning seems quite easy. However, there are also important practical issues that all users of deep learning need to be aware of. In particular, a look at the loss over the training iterations is very important. If the loss increases quickly after the beginning, a typical problem is that the learning rate η is set too high. This is typically referred to as *exploding gradient*. Setting η too low, however, can also result in a stagnation of the loss over iterations. In this case, we observe again vanishing gradients. Hence, correct choice of η and other training hyper-parameters is crucial for successful training [37].

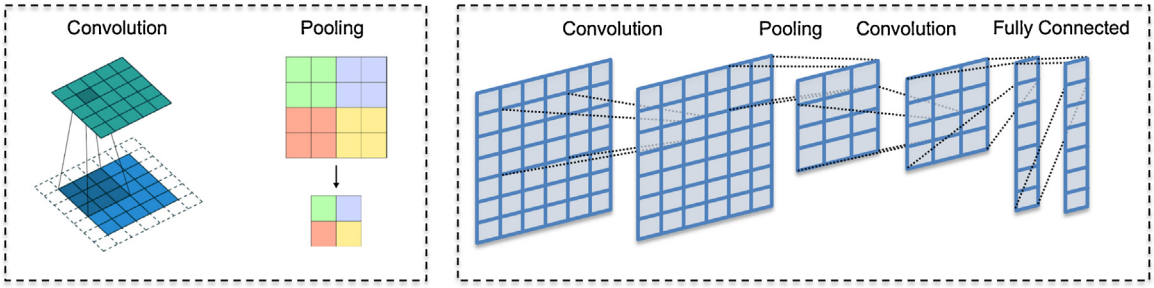

Figure 6. Convolutional layers only face a limited preceptive field and all neurons share the same weights (cf. left side of the figure; adopted from [40]). Pooling layers reduce the total input size. Both are typically combined in an alternating manner to construct convolutional neural networks (CNNs). An example is shown on the right.

In addition to the training set, a validation set is used to determine over-fitting. In contrast to the training set, the validation set is never used to actually update the parameter weights. Hence, the loss of the validation set allows an estimate for the error on unseen data. During optimization, the loss on the training set will continuously fall. However, as the validation set is independent, the loss on the validation set will increase at some point in training. This is typically a good point to stop updating the model before it over-fits to the training data.

Another common mistake is bias in training or test data. First of all, hyper-parameter tuning has to be done on validation data before actual test data is employed. In principle, test data should only be looked at once architecture, parameters, and all other factors of influence are set. Only then the test data is to be used. Otherwise, repeated testing will lead to optimistic results [37] and the system's performance will be over-estimated. This is as forbidden as including the test data in the training set. Furthermore, confounding factors may influence the classification results. If, for example, all pathological data was collected with Scanner A and all control data was collected with Scanner B, then the network may simply learn to differentiate the two scanners instead of the identifying the disease [38].

Due to the nature of gradient descent, training will stop once a minimum is reached. However, due to the general non-convexity of the loss function, this minimum is likely to be only a local minimum. Hence, it is advisable to perform multiple training runs with different initialization techniques in order to estimate a mean and a standard deviation for the model performance. Single training runs may be biased towards a single more or less random initialization.

Furthermore, it is very common to use typical regularization terms on parameters, as it is commonly done in other fields of medical imaging. Here, L2- and L1-norms are common choices. In addition, regularization can also be enforced by other techniques such as *dropout*, *weight-sharing*, and *multi-task learning*. An excellent overview is given in [37].

Also note that the output of a neural network does not equal to confidence, even if they are scaled between 0 and 1 and appear like probabilities, e.g. when using the so-called *softmax* function. In order to get realistic estimates of confidence other techniques have to be employed [39].

The last missing remark towards deep learning is the role of availability of large amounts of data and labels or annotations that could be gathered over the internet, the immense compute power that became available by using graphics cards for general purpose computations, and, last but not least, the positive trend towards open source software that enables users world-wide to download and extend deep learning methods very quickly. All three elements were crucial to enable this extremely fast rise of deep learning.

2.5 Important architectures in deep learning

With the developments of the previous section, much progress was made towards improved signal, image, video, and audio processing, as already detailed earlier. In this introduction, we are not able to highlight all developments, because this would go well beyond the scope of this document, and there are other sources that are more suited for this purpose [31,37,12]. Instead, we will only shortly discuss some advanced network architectures that we believe had, or will have, an impact on medical image processing.

Autoencoders use a contracting and an expanding branch to find representations of the input of a lower dimensionality [41]. They do not require annotations, as the network is trained to predict the original input using loss functions such as $L(\boldsymbol{\theta}) = ||\hat{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{x}||_2^2$. Variants use convolutional networks [42], add noise to the input [43], or aim at finding sparse representations [44].

Generative adversarial networks (GANs) employ two networks to learn a representative distribution from the training data [45]. A *generator network* creates new images from a noise input, while a *discriminator network* tries to differentiate real images from generated images. Both are trained in an alternating manner such that both gradually improve for their respective tasks. GANs are known to generate plausible and realistically looking images. So-called Wasserstein GANs can reduce instability in training [46]. Conditional GANs [47]

allow to encode states in the process such that images with desired properties can be generated. CycleGANs [48] drive this even further as they allow to convert one image from one domain to another, for example from day to night, without directly corresponding images in the training data.

Google's inception network is an advanced and deep architecture that was applied successfully for several tasks [49]. Its main highlight is the introduction of the so-called *inception block* that essentially allows to compute convolutions and pooling operations in parallel. By repeating this block in a network, the network can select by itself in which sequence convolution and pooling layers should be combined in order to solve the task at hand effectively.

Ronneberger's U-net is a breakthrough towards automatic image segmentation [50] and has been applied successfully in many tasks that require image-to-image transforms, for example, images to segmentation masks. Like the autoencoder, it consists of a contracting and an expanding branch, and it enables multi-resolution analysis. In addition, U-net features skip connections that connect the matching resolution levels of the encoder and the decoder stage. Doing so, the architecture is able to model general high-resolution multi-scale image-to-image transforms. Originally proposed in 2-D, many extensions, such as 3-D versions, exist [51,52].

ResNets have been designed to enable training of very deep networks [53]. Even with the methods described earlier in this paper, networks will not benefit from more than 30 to 50 layers, as the gradient flow becomes numerically unstable in such deep networks. In order to alleviate the problem, a so-called *residual block* is introduced, and layers take the form $\hat{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{x} + \hat{\boldsymbol{f}}'(\boldsymbol{x})$, where $\hat{\boldsymbol{f}}'(\boldsymbol{x})$ contains the actual network layer. Doing so has the advantage that the addition introduces a second parallel branch into the network that lets the gradient flow from end to end. ResNets also have other interesting properties, e.g., their residual blocks behave like ensembles of classifiers [54].

Variational networks enable the conversion of an energy minimization problem into a neural network structure [55]. We consider this type of network as particular interesting, as many problems in traditional medical image processing are expressed as energy minimization problems. The main idea is as follows: The energy function is typically minimized by optimization programs such as gradient descent. Thus, we are able to use the gradient of the original problem to construct a so-called *variational unit* that describes exactly one update step of the optimization program. Succession of such units then describe the complete variational network. Two observations are noteworthy: First, this type of framework allows to learn operators within one variational unit, such as a sparsifying transform for compressed sensing problems. Second, the variational units generally form residual blocks, and thus variational networks are always ResNets as well.

Recurrent neural networks (RNNs) enable the processing of sequences with long term dependencies [56]. Furthermore, recurrent nets introduce state variables that allow the cells to carry memory and essentially model any finite state machine. Extensions are long-short-term memory (LSTM) networks [57] and gated recurrent units (GRU) [58] that can model explicit read and write memory transactions similar to a computer.

2.6 Advanced deep learning concepts

In addition to the above mentioned architectures, there are also useful concepts that allow building more robust and versatile networks. Again, the here listed methods are incomplete. Still, we aimed at including the most useful ones.

Data augmentation In data augmentation, common sources of variation are explicitly added to training samples. These models of variation typically include noise, changes in contrast, and rotations and translations. In biased data, it can be used to improve the numbers of infrequent observations. In particular, the success of U-net is also related to very powerful augmentation techniques that include, for example, non-rigid deformations of input images and the desired segmentation [50]. In most recent literature, reports are found that also GANs are useful for data augmentation [59].

Precision learning is a strategy to include known operators into the learning process [60]. While this idea is counter-intuitive for most recognition tasks, where we want to learn the optimal representation, the approach is actually very useful for signal processing tasks in which we know *a priori* that a certain operator must be present in the processing chain. Embedding the operator in the network reduces the maximal training error, reduces the number of unknowns and therefore the number of required training samples, and enables mixing of most signal processing methods with deep learning. The approach is applicable to a broad range of operators. The main requirement is that a gradient or sub-gradient must exist.

Adversarial examples consider the input to a neural network as a possible weak spot that could be exploited by an attacker [61]. Generally, attacks try to find a perturbation $\boldsymbol{e}$ such that $\hat{f}(\boldsymbol{x} + \boldsymbol{e})$ indicates a different class than the true y, while keeping the magnitude of $\boldsymbol{e}$ low, for example, by minimizing $||\boldsymbol{e}||_2^2$. Using different objective functions allows to form different types of attacks. Attacks range from generating noise that will mislead the network, but will remain unnoticed by a human observer, to specialized patterns that will even mislead networks after printing and re-digitization [62].

Deep reinforcement learning is a technique that allows to train an artificial agent to perform actions given inputs from an environment and expands on traditional reinforcement learning theory [63]. In this context, deep networks are often used as flexible function approximators representing value functions and/or policies [4]. In order to enable time-series processing, sequences of environmental observations can be employed [5].

3 Results

As can be seen in the last few paragraphs, deep learning now offers a large set of new tools that are applicable to many problems in the world of medical image processing. In fact, these tools have already been widely employed. In particular, perceptual tasks are well suited for deep learning. We present some highlights that are discussed later in this section in Fig. 7. On the international conference of *Medical Image Computing and Computer-Assisted Intervention* (MICCAI) in 2018, approximately 70% of all accepted publications were related to the topic of deep learning. Given this fast pace of progress, we are not able to describe all relevant publications here. Hence, this overview is far from being complete. Still we want to highlight some publications that are representative for the current developments in the field. In terms of structure and organization, we follow [22] here, but add recent developments in physical simulation and image reconstruction.

3.1 Image detection and recognition

Image detection and recognition deals with the problem of detecting a certain element in a medical image. In many cases, the images are volumetric. Therefore efficient parsing is a must. A popular strategy to do so is marginal space learning [64], as it is efficient and allows to detect organs robustly. Its deep learning counter-part [65] is even more efficient, as its probabilistic boosting trees are replaced using a neural network-based boosting cascade. Still, the entire volume has to be processed to detect anatomical structures reliably. [65] drives efficiency even further by replacing the search process by an artificial agent that follows anatomy to detect anatomical landmarks using deep reinforcement learning. The method is able to detect hundreds of landmarks in a complete CT volume in few seconds.

Bier et al. proposed an interesting method in which they detect anatomical landmarks in 2-D X-ray projection images [66]. In their method, they train projection-invariant feature descriptors from 3-D annotated landmarks using a deep network. Yet another popular method for detection are the so-called region proposal convolutional neural networks. In [67] they are applied to robustly detect tumors in mammographic images.

Detection and recognition are obviously also applied in many other modalities and a great body of literature exists. Here, we only report two more applications. In histology, cell detection and classification is an important task, which is tackled by Aubreville et al. using guided spatial transformer networks [68] that allow refinement of the detection before the actual classification is done. The task of mitosis classification benefits from this procedure. Convolutional neural networks are also very effective for other image classification tasks. In [69] they are employed to automatically detect images containing motion artifacts in confocal laser-endoscopy images.

3.2 Image segmentation

Also image segmentation greatly benefited from the recent developments in deep learning. In image segmentation, we aim to determine the outline of an organ or anatomical structure as accurately as possible. Again, approaches based on convolutional neural networks seem to dominate. Here, we only report Holger Roth's Deeporgan [72], the brain MR segmentation using CNN by Moeskops et al. [73], a fully convolutional multi-energy 3-D U-net presented by Chen et al. [74], and a U-net-based stent segmentation in X-ray projection domain by Breininger et al. [71] as representative examples. Obviously segmentation using deep convolutional networks also works in 2-D as shown by Nirschl et al. for histopathologic images [75].

Middleton et al. already experimented with the fusion of neural networks and active contour models in 2004 well before the advent of deep learning [76]. Yet, their approach is neither using deep nets nor end-to-end training, which would be desirable for a state-of-the-art method. Hence, revisiting traditional segmentation approaches and fusing them with deep learning in an end-to-end fashion seems a promising scope of research. Fu et al. follow a similar idea by mapping Frangi's vesselness into a neural network [77]. They demonstrate that they are able to adjust the convolution kernels in the first step of the algorithm towards the specific task of vessel segmentation in ophthalmic fundus imaging.

Yet another interesting class of segmentation algorithms is the use of recurrent networks for medical image segmentation. Poudel et al. demonstrate this for a recurrent fully convolutional neural network on multi-slice MRI cardiac data [78], while Andermatt et al. show effectiveness of GRUs for brain segmentation [79].

3.3 Image registration

While the perceptual tasks of image detection and classification have been receiving a lot of attention with respect to applications of deep learning, image registration has not seen this large boost yet. However, there are several promising works found in the literature that clearly indicate that there are also a lot of opportunities.

One typical problem in point-based registration is to find good feature descriptors that allow correct identification of corresponding points. Wu et al. propose to do so using autoencoders to mine good features in an unsupervised way [80]. Schaffert et al. drive this even further and use the registration metric itself as loss function for learning good feature representations [81]. Another option to solve 2-D/3-D registration problems is to estimate the 3-D pose directly from the 2-D point features [82].

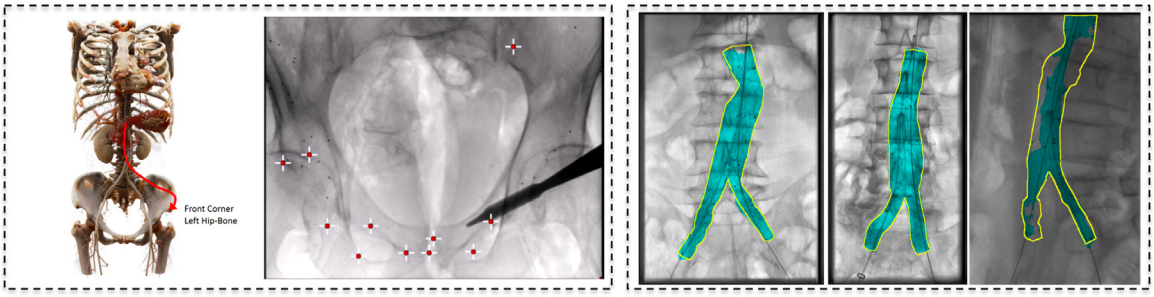

Figure 7. Deep learning excels in perceptual tasks such as detection and segmentation. The left hand side shows the artificial agent-based landmark detection after Ghesu et al. [70] and the X-ray transform-invariant landmark detection by Bier et al. [66] (projection image courtesy of Dr. Unberath). The right hand side shows a U-net-based stent segmentation after Breininger et al. [71]. Images are reproduced with permission by the authors.

For full volumetric registration, examples of deep learning-based approaches are also found. The quicksilver algorithm is able to model a deformable registration and uses a patch-wise prediction directly from the image appearance [83]. Another approach is to model the registration problem as a control problem that is dealt with using an agent and reinforcement learning. Liao et al. propose to do so for rigid registration predicting the next optimal movement in order to align both volumes [84]. This approach can also be applied to non-rigid registration using a statistical deformation model [85]. In this case, the actions are movements in the vector space of the deformation model. Obviously, agent-based approaches are also applicable for point-based registration problems. Zhong et al. demonstrate this for intra-operative brain shift using imitation learning [86].

3.4 Computer-aided diagnosis

Computer-aided diagnosis is regarded as one of the most challenging problems in the field of medical image processing. Here, we are not only acting in a supportive role quantifying evidence towards the diagnosis. Instead the diagnosis itself is to be predicted. Hence, decisions have to be done with utmost care and decisions have to be reliable.

The analysis of chest radiographs comprises a significant amount of work for radiologistic and is performed routinely. Hence, reliable support to prevent human error is highly desirable. An example to do so is given in [87] by Diamant et al. using transfer learning techniques.

A similar workload is imposed on ophthalmologists in the reading of volumetric optical coherence tomography data. Google's Deep Mind just recently proposed to support this process in terms of referral decision support [88].

There are many other studies found in this line, for example, automatic cancer assessment in confocal laser endoscopy in different tissues of the head and neck [89], deep learning for mammogram analysis [90], and classification of skin cancer [91].

3.5 Physical simulation

A new field of deep learning is the support of physical modeling. So far this has been exploited in the gaming industry to compute realistically appearing physics engines [92], or for smoke simulation [93] in real-time. A first attempt to bring deep learning to bio-medical modeling was done by Meister et al. [94].

Based on such observations, researchers started to bring such methods into the field of medical imaging. One example to do so is the deep scatter estimation by Maier et al. [95]. Unberath et al. drive this even further to emulate the complete X-ray formation process in their DeepDRR [96]. In [97] Horger et al. demonstrate that even noise of unknown distributions can be learned, leading to an efficient generative noise model for realistic physical simulations.

Also other physical processes have been investigated using deep learning. In [60] a material decomposition using deep learning embedding prior physical operators using precision learning is proposed. Also physically less plausible interrelations are attempted. In [98], Han et al. attempt to convert MR volumes to CT volumes. Stimpel et al. drive this even further predicting X-ray projections from MR projection images [99]. While these observations seem promising, one has to follow such endeavors with care. Schiffers et al. demonstrate that cycleGANs may create correctly appearing fluorescence images from fundus images in ophthalmology [100]. Yet, undesired effects appear, as occasionally drusen are mapped onto micro aneurysms in this process. Cohen et al. demonstrate even worse effects [101]. In their study, cancers disappeared or were created during the modality-to-modality mapping. Hence, such approaches have to be handled with care.

3.6 Image reconstruction

Also the field of medical image reconstruction has been affected by deep learning and was just recently the topic of a special issue in the IEEE Transactions on Medical Imaging.

The editorial actually gives an excellent overview on the latest developments [102] that we will summarize in the next few lines.

One group of deep learning algorithms omit the actual problem of reconstruction and formulate the inverse as image-to-image transforms with different initialization techniques before processing with a neural network. Recent developments in this *image-to-image reconstruction* are summarized in [103]. Still, there is continuous progress in the field, e.g. by application of the latest network architectures [104] or cascading of U-nets [105].

A recent paper by Zhu et al. proposes to learn the entire reconstruction operation only from raw data and corresponding images [106]. The basic idea is to model an autoencoder-like dimensionality reduction in raw data and reconstruction domain. Then both are linked using a non-linear correlation model. The entire model can then be converted into a single network and trained in an end-to-end manner. In the paper, they show that this is possible for 2-D MR and PET imaging and largely outperforms traditional approaches.

Learning operators completely data-driven carries the risk that undesired effects may occur [107], as is shown in Fig. 8. Hence integration of prior knowledge and the structure of the operators seems beneficial, as already described in the concept of precision learning in the previous section. Ye et al. embed a multi-scale transform into the encoder and decoder of a U-net-like network, which gives rise to the concept of deep convolutional framelets [108]. Using wavelets for the multi-scale transform has been successfully applied in many applications ranging from denoising [109] to sparse view computed tomography [110].

If we design a neural network inspired by iterative algorithms that minimize an energy function step by step, the concept of variational networks is useful. Doing so allows to map virtually all iterative reconstruction algorithms onto deep networks, e.g., by using a fixed number of iterations. There are several impressive works found in the literature, of which we only name the MRI reconstruction by Hammernik et al. [111] and the sound speed reconstruction by Vishnevskiy et al. [112] at this point. The concept can be expanded even further, as Adler et al. demonstrate by learning an entire primal-dual reconstruction [113].

Würfl et al. also follow the idea of using prior operators [114,115]. Their network is inspired by the classical filtered back-projection that can be retrained to better approximate limited angle geometries that typically cannot be solved by classical analytic inversion models. Interestingly, as the approach is described in an end-to-end fashion, errors in the discretization or initialization of the filtering steps are intrinsically corrected by the learning process [116]. They also show that their method is compatible with other approaches, such as variational networks that are able to learn an additional de-streaking sparsifying transform [117]. Syben et al. drive these efforts even further and demonstrate that the concept of precision learning is able to mathematically derive a neural network structure [118]. In their work, they demonstrate that they are able to postulate that an expensive matrix inverse is a circulant matrix and hence can be replaced by a convolution operation. Doing so leads to the derivation of a previously unknown filtering, back-projection, re-projection-style rebinning algorithm that intrinsically suffers less from resolution loss than traditional interpolation-based rebinning methods.

As noted earlier, all networks are prone to adversarial attacks. Huang et al. demonstrate this [107] in their work, showing that already incorrect noise modeling may distort the entire image. Yet, the networks reconstruct visually pleasing results and artifacts cannot be as easily identified as in classical methods. One possible remedy is to follow the precision learning paradigm and fix as much of the network as possible, such that it can be analyzed with classical methods as demonstrated in [115]. Another promising approach is Bayesian deep learning [39]. Here the network output is two-fold: the reconstructed image plus a confidence map on how accurate the content of the reconstructed image was actually measured.

Obviously, deep learning also plays a role in suppression of artifacts. In [119], Zhang et al. demonstrate this effectively for metal artifacts. As a last example, we list Bier et al. here, as they show that deep learning-based motion tracking is also feasible for motion compensated reconstruction [120].

4 Discussion

In this introduction, we reviewed the latest developments in deep learning for medical imaging. In particular detection, recognition, and segmentation tasks are well solved by the deep learning algorithms. Those tasks are clearly linked to perception and there is essentially no prior knowledge present. Hence, state-of-the-art architectures from other fields, such as computer vision, can often be easily adopted to medical tasks. In order to gain better understanding of the black box, reinforcement learning and modeling of artificial agents seem well suited.

In image registration, deep learning is not that broadly used. Yet, interesting approaches already exist that are able to either predict deformations directly from the image input, or take advantage of reinforcement learning-based techniques that model registration as on optimal control problem. Further benefits are obtained using deep networks for learning representations, which are either done in an unsupervised fashion or using the registration metric itself.

Computer-aided diagnosis is a hot topic with many recent publications address. We expect that simpler standard tasks that typically result in a high workload for medical doctors will be solved first. For more complex diagnoses, the current deep nets that immediately result in a decision are not that well suited, as it is difficult to understand the evidence. Hence, approaches are needed that link observations to evidence to construct a line of argument towards a decision. It

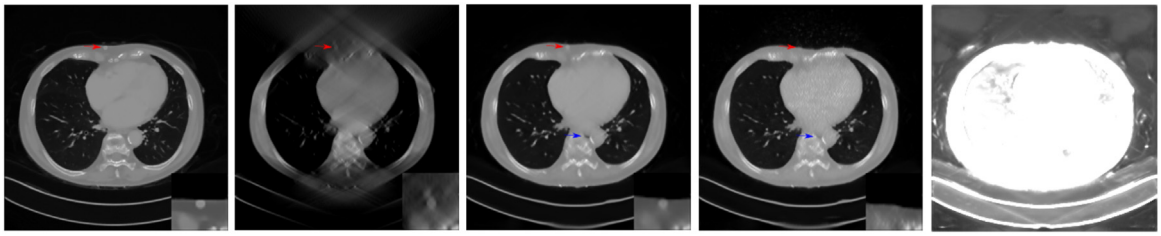

Figure 8. Results from a deep learning image-to-image reconstruction based on U-net. The reference image with a lesion embedded is shown on the left followed by the analytic reconstruction result that is used as input to U-net. U-net does an excellent job when trained and tested without noise. If unmatched noise is provided as input, an image is created that appears artifact-free, yet not just the lesion is gone, but also the chest surface is shifted by approximately 1 cm. On the right hand side, an undesirable result is shown that emerged at some point during training of several different versions of U-net which shows organ-shaped clouds in the air in the background of the image. Note that we omitted displaying multiple versions of "Limited Angle" as all three inputs to the U-Nets would appear identically given the display window of the figure of [−1000, 1000] HU.

is the strong belief of the authors that only if such evidence-based decision making is achieved, the new methodology will make a significant impact to computer-aided diagnosis.

Physical simulation can be accelerated dramatically with realistic outcomes as shown in the field of computer games and graphics. Therefore, the methods are highly relevant, in particular for interventional applications, in which real-time processing is mandatory. First approaches exist, yet there is considerable room for more new developments. In particular, precision learning and variational networks seem to be well suited for such tasks, as they provide some guarantees to prediction outcomes. Hence, we believe that there are many new developments to follow, in particular in radiation therapy and real-time interventional dose tracking.

Reconstruction based on data-driven methods yield impressive results. Yet, they may suffer from a "new kind" of *deep learning artifacts*. In particular, the work by Huang et al. [107] show these effects in great detail. Both precision learning and Bayesian approaches seem well suited to tackle the problem in the future. Yet, it is unclear how to benefit best from the data-driven methods while maintaining intuitive and safe image reading.

A great advantage of all the deep learning methods is that they are inherently compatible to each other and to many classical approaches. This fusion will spark many new developments in the future. In particular, the fusion on network-level using either the direct connection of networks or precision learning allows end-to-end training of algorithms. The only requirement for this deep fusion is that each operation in the hybrid net has a gradient or sub-gradient for the optimization. In fact, there are already efforts to design whole programming languages to be compatible with this kind of *differential programming* [121]. With such integrated networks, multi-task learning is enabled, for example, training of networks that deliver optimal reconstruction quality and the best volumetric overlap of the resulting segmentation at the same time, as already conjectured in [122]. This point may even be expanded to computer-aided diagnosis or patient benefit.

In general, we observe that the CNN architectures that emerge from deep learning are computationally very efficient. Networks find solutions that are on par or better than many state-of-the-art algorithms. However, their computational cost at inference time is often much lower than state-of-the-art algorithms in typical domains of medical imaging in detection, segmentation, registration, reconstruction, and physical simulation tasks. This benefit at run-time comes at high computational cost during training that can take days even on GPU clusters. Given an appropriate problem domain and training setup, we can thus exploit this effect to save run-time at the cost of additional training time.

Deep learning is extremely data hungry. This is one of the main limitations that the field is currently facing, and performance grows only logarithmically with the amount of data used [123]. Approaches like weakly supervised training [124] will only partially be able to close this gap. Hence, one hospital or one group of researchers will not be able to gather a competitive amount of data in the near future. As such, we welcome initiatives such as the grand challenges[3] or medical data donors,[4] and hope that they will be successful with their mission.

5 Conclusion

In this short introduction to deep learning in medical image processing we were aiming at two objectives at the same time. On the one hand, we wanted to introduce to the field of deep learning and the associated theory. On the other hand, we wanted to provide a general overview on the field and potential future applications. In particular, perceptual tasks have been

[3] https://grand-challenge.org.
[4] http://www.medicaldatadonors.org.

studied most so far. However, with the set of tools presented here, we believe many more problems can be tackled. So far, many problems could be solved better than the classical state-of-the-art does alone, which also sparked significant interest in the public media. Generally, safety and understanding of networks is still a large concern, but methods to deal with this are currently being developed. Hence, we believe that deep learning will probably remain an active research field for the coming years.

If you enjoyed this introduction, we recommend that you have a look at our video lecture that is available at https://www.video.uni-erlangen.de/course/id/662.

Acknowledgements

We express our thanks to Katharina Breininger, Tobias Würfl, and Vincent Christlein, who did a tremendous job when we created the deep learning course at the University of Erlangen-Nuremberg. Furthermore, we would like to thank Florin Ghesu, Bastian Bier, Yixing Huang, and again Katharina Breininger for the permission to highlight their work and images in this introduction. Last but not least, we also express our gratitude to the participants of the course "Computational Medical Imaging" (https://www5.cs.fau.de/lectures/sarntal-2018/), who were essentially the test audience of this article during the summer school "Ferienakademie 2018".

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436.
[2] Dahl GE, Yu D, Deng L, Acero A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Actions Audio Speech Lang Process 2012;20:30–42.
[3] Krizhevsky A, Sutskever I, Hinton GE. ImageNET classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–105.
[4] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature 2016;529:484.
[5] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015;518:529.
[6] Mordvintsev A, Olah C, Tyka M. Inceptionism: going deeper into neural networks. Google Research Blog; 2015. p. 5. Retrieved June 20.
[7] Tan WR, Chan CS, Aguirre HE, Tanaka K. ArtGAN: artwork synthesis with conditional categorical GANs. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE; 2017. p. 3760–4.
[8] Briot J, Hadjeres G, Pachet F. Deep learning techniques for music generation – a survey; 2017. CoRR abs/1709.01620.
[9] Seebock P. Deep learning in medical image analysis (Master's thesis). Vienna University of Technology, Faculty of Informatics; 2015.
[10] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221–48.
[11] Pawlowski N, Ktena SI, Lee MC, Kainz B, Rueckert D, Glocker B, et al. DLTK: state of the art reference implementations for deep learning on medical images; 2017 arXiv:1711.06853.
[12] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
[13] Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017;37:505–15.
[14] Suzuki K. Survey of deep learning applications to medical image analysis. Med Imaging Technol 2017;35:212–26.
[15] Hagerty J, Stanley RJ, Stoecker WV. Medical image processing in the age of deep learning. In: Proceedings of the 12th international joint conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP). 2017. p. 306–11.
[16] Lakhani P, Gray DL, Pett CR, Nagy P, Shih G. Hello world deep learning in medical imaging. J Digit Imaging 2018;31:283–9.
[17] Kim J, Hong J, Park H, Kim J, Hong J, Park H. Prospects of deep learning for medical imaging. Precis Future Med 2018;2:37–52.
[18] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access 2018;6:9375–89.
[19] Rajchl M, Ktena SI, Pawlowski N. An introduction to biomedical image analysis with TensorFlow and DLTK; 2018 https://medium.com/tensorflow/an-introduction-to-biomedical-image-analysis-with-tensorflow-and-dltk-2c25304e7c13.
[20] Breininger K, Würfl T. Tutorial: how to build a deep learning framework; 2018 https://github.com/kbreininger/tutorial-dlframework.
[21] Cornelisse D. An intuitive guide to Convolutional Neural Networks; 2018 https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050.
[22] Zhou SK, Greenspan H, Shen D. Deep learning for medical image analysis. Academic Press; 2017.
[23] Lu L, Zheng Y, Carneiro G, Yang L. Deep learning and convolutional neural networks for medical image computing. Springer; 2017.
[24] Chollet F. Deep learning with python. Manning Publications Co.; 2017.
[25] Géron A. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.; 2017.
[26] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging. Med Phys 2018;46(1):e1–36.
[27] Niemann H. Pattern analysis and understanding, vol. 4. Springer Science & Business Media; 2013.
[28] Rosenblatt F. The perceptron, a perceiving and recognizing automaton (Project Para). Cornell Aeronautical Laboratory; 1957.
[29] Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 1989;2:303–14.
[30] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw 1991;4:251–7.
[31] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35:1798–828.
[32] Ivanova I, Kubat M. Initialization of neural networks by means of decision trees. Knowl Based Syst 1995;8:333–44.
[33] Sonoda S, Murata N. Neural network with unbounded activation functions is universal approximator. Appl Comput Harmon Anal 2017;43:233–68.
[34] Rockafellar R. Convex analysis, Princeton landmarks in mathematics and physics. Princeton University Press; 1970.
[35] Bertsekas DP, Scientific A. Convex optimization algorithms. Athena Scientific Belmont; 2015.
[36] Schirrmacher F, Köhler T, Husvogt L, Fujimoto JG, Hornegger J, Maier AK. QuaSI: quantile sparse image prior for spatio-temporal denoising of retinal OCT data. In: Medical Image Computing and Computer-Assisted Intervention, MICCAI 2017: 20th international conference, proceedings, vol. 10434. Springer; 2017. p. 83.
[37] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT Press; 2016.
[38] Maier A, Schuster M, Eysholdt U, Haderlein T, Cincarek T, Steidl S, et al. QMOS – a robust visualization method for speaker dependencies with different microphones. J Pattern Recognit Res 2009;4:32–51.

[39] Schlemper J, Castro DC, Bai W, Qin C, Oktay O, Duan J, et al. Bayesian deep learning for accelerated MR image reconstruction. In: Knoll F, Maier A, Rueckert D, editors. Machine learning for medical image reconstruction. Cham: Springer International Publishing; 2018. p. 64–71.
[40] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning; 2016 [ArXiv e-prints].
[41] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on machine learning. ACM; 2008. p. 1096–103.
[42] Holden D, Saito J, Komura T, Joyce T. Learning motion manifolds with convolutional autoencoders. In: SIGGRAPH Asia 2015 Technical Briefs. ACM; 2015. p. 18.
[43] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010;11:3371–408.
[44] Huang FJ, Boureau Y-L, LeCun Y, Huang Fu Jie, Boureau Y-Lan, LeCun Yann, et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: IEEE conference on Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE; 2007. p. 1–8.
[45] Goodfellow I. NIPS 2016 tutorial: generative adversarial networks; 2016 arXiv:1701.00160.
[46] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: International conference on machine learning. 2017. p. 214–23.
[47] Gauthier J. Conditional generative adversarial nets for convolutional face generation. In: Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014; 2014. p. 2.
[48] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks; 2017.
[49] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1–9.
[50] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
[51] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-NET: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 424–32.
[52] Milletari F, Navab N, Ahmadi S-A. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D Vision (3DV). IEEE; 2016. p. 565–71.
[53] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 770–8.
[54] Veit A, Wilber MJ, Belongie S. Residual networks behave like ensembles of relatively shallow networks. In: Advances in neural information processing systems; 2016. p. 550–8.
[55] Kobler E, Klatzer T, Hammernik K, Pock T. Variational networks: connecting variational methods and deep learning. In: German conference on pattern recognition. Springer; 2017. p. 281–93.
[56] Mandic DP, Chambers J. Recurrent neural networks for prediction: learning algorithms, architectures and stability. John Wiley & Sons, Inc.; 2001.
[57] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.
[58] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling; 2014 arXiv:1412.3555.
[59] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification; 2018. CoRR abs/1803.01229.
[60] Maier A, Schebesch F, Syben C, Würfl T, Steidl S, Choi J-H, et al. Precision learning: towards use of known operators in neural networks. In: Tan JKT, editor. 24th International Conference on Pattern Recognition (ICPR). 2018. p. 183–8.
[61] Yuan X, He P, Zhu Q, Bhat RR, Li X. Adversarial examples: attacks and defenses for deep learning; 2017 arXiv:1712.07107.
[62] Brown TB, Mané D, Roy A, Abadi M, Gilmer J. Adversarial patch; 2017 arXiv:1712.09665.
[63] Sutton RS, Barto AG, Bach F, Sutton, Richard S, Barto Andrew G, et al. Reinforcement learning: an introduction. MIT Press; 1998.
[64] Zheng Y, Comaniciu D. Marginal space learning. In: Marginal space learning for medical image analysis. Springer; 2014. p. 25–65.
[65] Ghesu FC, Krubasik E, Georgescu B, Singh V, Zheng Y, Hornegger J, et al. Marginal space deep learning: efficient architecture for volumetric image parsing. IEEE Trans Med Imaging 2016;35:1217–28.
[66] Bier B, Unberath M, Zaech J-N, Fotouhi J, Armand M, Osgood G, et al. X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 55–63.
[67] Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E. A region based convolutional network for tumor detection and classification in breast mammography. In: Deep learning and data labeling for medical applications. Springer; 2016. p. 197–205.
[68] Aubreville M, Krappmann M, Bertram C, Klopfleisch R, Maier A. A guided spatial transformer network for histology cell differentiation. In: Association TE, editor. Eurographics workshop on visual computing for biology and medicine. 2017. p. 21–5.
[69] Aubreville M, Stöve M, Oetter N, de Jesus Goncalves M, Knipfer C, Neumann H, et al. Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images. Int J Comput Assist Radiol Surg 2018, http://dx.doi.org/10.1007/s11548-018-1836-1.
[70] Ghesu FC, Georgescu B, Zheng Y, Grbic S, Maier A, Hornegger J, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. IEEE Trans Pattern Anal Mach Intell 2017;41(1):176–89.
[71] Breininger K, Albarqouni S, Kurzendorfer T, Pfister M, Kowarschik M, Maier A. Intraoperative stent segmentation in X-ray fluoroscopy for endovascular aortic repair. Int J Comput Assist Radiol Surg 2018;13.
[72] Roth HR, Lu L, Farag A, Shin H-C, Liu J, Turkbey EB, et al. DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: International conference on medical image computing, computer-assisted intervention. Springer; 2015. p. 556–64.
[73] Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging 2016;35: 1252–61.
[74] Chen S, Zhong X, Hu S, Dorn S, Kachelriess M, Lell M, et al. Automatic multi-organ segmentation in dual energy CT using 3D fully convolutional network. In: van Ginneken B, Welling M, editors. MIDL. 2018.
[75] Nirschl JJ, Janowczyk A, Peyster EG, Frank R, Margulies KB, Feldman MD, et al. Deep learning tissue segmentation in cardiac histopathology images. In: Deep learning for medical image analysis. Elsevier; 2017. p. 179–95.
[76] Middleton I, Damper RI. Segmentation of magnetic resonance images using a combination of neural networks and active contour models. Med Eng Phys 2004;26:71–86.
[77] Fu W, Breininger K, Schaffert R, Ravikumar N, Würfl T, Fujimoto J, et al. Frangi-Net: a neural network approach to vessel segmentation. In:

Maier A, Deserno Th, Handels H, Maier-Hein KH, Palm C, Tolxdorff Th, editors. Bildverarbeitung für die Medizin. 2018. p. 341–6.
[78] Poudel RP, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: Reconstruction, segmentation, and analysis of medical images. Springer; 2016. p. 83–94.
[79] Andermatt S, Pezold S, Cattin P. Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data. In: Deep learning and data labeling for medical applications. Springer; 2016. p. 142–51.
[80] Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans Biomed Eng 2016;63:1505–16.
[81] Schaffert R, Wang J, Fischer P, Borsdorf A, Maier A. Metric-driven learning of correspondence weighting for 2-D/3-D image registration. In: German Conference on Pattern Recognition (GCPR). 2018.
[82] Miao S, Wang JZ, Liao R. Convolutional neural networks for robust and real-time 2-D/3-D registration. In: Deep learning for medical image analysis. Elsevier; 2017. p. 271–96.
[83] Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: fast predictive image registration – a deep learning approach. NeuroImage 2017;158:378–96.
[84] Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T, et al. An artificial agent for robust image registration. In: AAAI. 2017. p. 4168–75.
[85] Krebs J, Mansi T, Delingette H, Zhang L, Ghesu FC, Miao S, et al. Robust non-rigid registration through agent-based action learning. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI. Springer; 2017. p. 344–52.
[86] Zhong X, Bayer S, Ravikumar N, Strobel N, Birkhold A, Kowarschik M, et al. Resolve intraoperative brain shift as imitation game. In: MICCAI Challenge 2018 for Correction of Brainshift with Intra-Operative Ultrasound (CuRIOUS 2018). 2018.
[87] Diamant I, Bar Y, Geva O, Wolf L, Zimmerman G, Lieberman S, et al. Chest radiograph pathology categorization via transfer learning. In: Deep learning for medical image analysis. Elsevier; 2017. p. 299–320.
[88] De Fauw JR, Ledsam B, Romera-Paredes S, Nikolov N, Tomasev S, Blackwell H, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342.
[89] Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. Sci Rep 2017;7:41598-017.
[90] Carneiro G, Nascimento J, Bradley AP. Deep learning models for classifying mammogram exams containing unregistered multi-view images and segmentation maps of lesions. In: Deep learning for medical image analysis. Elsevier; 2017. p. 321–39.
[91] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115.
[92] Wu J, Yildirim I, Lim JJ, Freeman B, Tenenbaum J. Galileo: perceiving physical object properties by integrating a physics engine with deep learning. In: Advances in neural information processing systems; 2015. p. 127–35.
[93] Chu M, Thuerey N. Data-driven synthesis of smoke flows with CNN-based feature descriptors. ACM Trans Graph 2017;36:69.
[94] Meister F, Passerini T, Mihalef V, Tuysuzoglu A, Maier A, Mansi T. Towards fast biomechanical modeling of soft tissue using neural networks. In: Medical Imaging meets NeurIPS workshop at 32nd conference on Neural Information Processing Systems (NeurIPS). 2018.
[95] Maier J, Berker Y, Sawall S, Kachelrieß M. Deep scatter estimation (DSE): feasibility of using a deep convolutional neural network for real-time X-ray scatter prediction in cone-beam CT. Medical imaging 2018: physics of medical imaging, vol. 10573. International Society for Optics and Photonics; 2018. p. 105731L.
[96] Unberath M, Zaech J-N, Lee SC, Bier B, Fotouhi J, Armand M, et al. DeepDRR – a catalyst for machine learning in fluoroscopy-guided procedures. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 98–106.
[97] Horger F, Würfl T, Christlein V, Maier A. Towards arbitrary noise augmentation – deep learning for sampling from arbitrary probability distributions. In: International workshop on machine learning for medical image reconstruction. Springer; 2018. p. 129–37.
[98] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19.
[99] Stimpel B, Syben C, Würfl T, Mentl K, Dörfler A, Maier A. MR to X-ray projection image synthesis. In: Noo F, editor. Proceedings of the 5th international conference on image formation in X-ray computed tomography (CT-meeting). 2018. p. 435–8.
[100] Schiffers F, Yu Z, Arguin S, Maier A, Ren Q. Synthetic fundus fluorescein angiography using deep neural networks. In: Maier A, Deserno TM, Handels H, Maier-Hein KH, Palm C, Tolxdorff T, editors. Bildverarbeitung für die Medizin 2018. Berlin, Heidelberg: Springer Berlin Heidelberg; 2018. p. 234–8.
[101] Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 529–36.
[102] Wang G, Ye JC, Mueller K, Fessler JA. Image reconstruction is a new frontier of machine learning. IEEE Trans Med Imaging 2018;37:1289–96.
[103] McCann MT, Jin KH, Unser M. A review of convolutional neural networks for inverse problems in imaging; 2017 arXiv:1710.04011.
[104] Zhang Z, Liang X, Dong X, Xie Y, Cao G. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans Med Imaging 2018;37:1407–17.
[105] Kofler A, Haltmeier M, Kolbitsch C, Kachelrieß M, Dewey M. A U-Nets cascade for sparse view computed tomography. In: International workshop on machine learning for medical image reconstruction. Springer; 2018. p. 91–9.
[106] Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature 2018;555:487.
[107] Huang Y, Würfl T, Breininger K, Liu L, Lauritsch G, Maier A. Some investigations on robustness of deep learning in limited angle tomography. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 145–53.
[108] Ye JC, Han Y, Cha E. Deep convolutional framelets: a general deep learning framework for inverse problems. SIAM J Imaging Sci 2018;11:991–1048.
[109] Kang E, Chang W, Yoo J, Ye JC. Deep convolutional framelet denosing for low-dose CT via wavelet residual network. IEEE Trans Med Imaging 2018;37:1358–69.
[110] Han Y, Ye JC. Framing U-Net via deep convolutional framelets: application to sparse-view CT. IEEE Trans Med Imaging 2018;37:1418–29.
[111] Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a variational network for reconstruction of accelerated mri data. Magn Reson Med 2018;79:3055–71.
[112] Vishnevskiy V, Sanabria SJ, Goksel O. Image reconstruction via variational network for real-time hand-held sound-speed imaging. In: International workshop on machine learning for medical image reconstruction. Springer; 2018. p. 120–8.
[113] Adler J, Öktem O. Learned primal-dual reconstruction. IEEE Trans Med Imaging 2018;37:1322–32.

[114] Würfl T, Ghesu FC, Christlein V, Maier A. Deep learning computed tomography. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 432–40.
[115] Würfl T, Hoffmann M, Christlein V, Breininger K, Huang Y, Unberath M, et al. Deep learning computed tomography: learning projection-domain weights from image domain in limited angle problems. IEEE Trans Med Imaging 2018;37:1454–63.
[116] Syben C, Stimpel B, Breininger K, Würfl T, Fahrig R, Dörfler A, Maier A. Precision learning: Reconstruction filter kernel discretization. In: Proceedings of the Fifth International Conference on Image Formation in X-Ray Computed Tomography. 2018. p. 386–90.
[117] Hammernik K, Würfl T, Pock T, Maier A. A deep learning architecture for limited-angle computed tomography reconstruction. In: Maier-Hein KH, geb. Fritzsche, Deserno TM, geb. Lehmann, Handels H, Tolxdorff T, editors. Bildverarbeitung für die Medizin 2017. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017. p. 92–7.
[118] Syben C, Stimpel B, Lommen J, Würfl T, Dörfler A, Maier A. Deriving neural network architectures using precision learning: parallel-to-fan beam conversion. In: German Conference on Pattern Recognition (GCPR). 2018.
[119] Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in X-ray computed tomography. IEEE Trans Med Imaging 2018;37:1370–81.
[120] Bier B, Aschoff K, Syben C, Unberath M, Levenston M, Gold G, et al. Detecting anatomical landmarks for motion estimation in weight-bearing imaging of knees. In: International workshop on machine learning for medical image reconstruction. Springer; 2018. p. 83–90.
[121] Li T-M, Gharbi M, Adams A, Durand F, Ragan-Kelley J. Differentiable programming for image processing and deep learning in halide. ACM Trans Graph 2018;37:139.
[122] Wang G. A perspective on deep imaging. IEEE Access 2016;4: 8914–24.
[123] Sun C, Shrivastava A, Singh S, Gupta A. Revisiting unreasonable effectiveness of data in deep learning era; 2017. p. 1 arXiv:1707.02968.
[124] Oquab M, Bottou L, Laptev I, Sivic J. Is object localization for free? Weakly-supervised learning with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 685–94.