TECHNISCHE UNIVERSITÄT MÜNCHEN
TUM School of Computation, Information and Technology

Advancing Clinical Optoacoustic Tomography

Jan Kukačka

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Klaus Diepold

Prüfer der Dissertation:

1. Prof. Dr. Vasilis Ntziachristos
2. Prof. Dr.-Ing. Christian Jirauschek
3. Prof. Dr. Björn Menze

Die Dissertation wurde am 29. 6. 2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 22. 2. 2024 angenommen.

To grandma Jana

—

Breast cancer took your life but never your good spirit.

Abstract

Medical imaging is an essential tool for clinical decision making in scenarios such as screening, diagnostics, or therapy monitoring. Suitability of an imaging modality for a given scenario depends on the information value it provides—based on the contrast mechanism, penetration depth, and resolution—and its price in terms of examination cost, invasiveness, acquisition speed, and other factors. Clinical scenarios outside the suitability limits of the existing imaging modalities are left without suitable decision support. For example, every year, thousands of patients with breast lesions are exposed to a biopsy because the status of their lesion cannot be established in a non-invasive manner, despite developed screening programs and established routines utilizing X-ray mammography, ultrasonography, and magnetic resonance imaging.

Optoacoustic imaging is a non-invasive technology providing optical contrast with high resolution at depths of several centimeters, available at moderate cost. With these unique properties, it provides otherwise unavailable functional and anatomic measurements and presents a natural complement to existing clinical modalities. Over the past three decades, optoacoustic imaging of small animals has matured into a valuable tool of pre-clinical research, delivering high-quality images and enabling novel biomedical research. On the other hand, its clinical translation has been hindered by insufficient image quality. Unlike small animal systems, clinical optoacoustic tomography operates with limited angular coverage of light delivery and signal detection, resulting in signal imperfections that are amplified by simple reconstruction algorithms. Furthermore, handheld imaging devices operate in the presence of motion which corrupts spectral information and prevents noise reduction through signal averaging.

To overcome these limitations and provide optoacoustic images of high-quality needed for clinical applications, we develop a series of image processing and visualization improvements that form a novel "second-generation" multispectral optoacoustic-ultrasound tomography pipeline. The proposed pipeline utilizes a motion-quantification algorithm for automated selection of stable multispectral frames, a model-based inversion procedure with system impulse response characterization, a frame averaging algorithm combined with a co-registration step, and adaptive non-linear contrast adjustment filters. The resulting images achieve a new level of quality for a handheld optoacoustic scanner and visualize vascular features deep in tissue with a high resolution of 200 μm.

The clinical value of the improved imaging technology stems from the ability to reliably detect features *in vivo* that are otherwise unavailable. We explore the added value in two directions: First, we consider the possibility of extracting quantitative morphological biomarkers using automated semantic image segmentation. Our experiments with self-supervised deep learning on fundus images as a proxy dataset validate the advantages of the methodology in terms of reducing the need for manually annotated data. Second, we perform two oncological clinical studies focused on breast cancer lesions and oral cancer metastases. With the aid of histopathology correlation, we characterize the features observed in presence of breast tumors. Furthermore, we study the possibility of detecting lymph node metastases using a tumor-specific contrast agent and propose an alternative marker of malignancy based on intranodal deoxyhemoglobin variability.

Altogether, this work presents the second generation of optoacoustic tomography methods and defines an improved standard for future clinical applications.

Zusammenfassung

Die medizinische Bildgebung spielt eine wichtige Rolle in klinischen Anwendungen, wie Screening, Diagnose oder Therapieüberwachung. Die Eignung einer Bildgebungsmodalität für einen Anwendungsfall erfolgt nach einer Kosten-Nutzen-Rechnung: dem Nutzen im Sinne des vermittelten Informationsgehalts in Bezug auf Kontrastmechanismus, Penetrationstiefe, und Auflösung einerseits, sowie den Kosten wie etwa dem finanziellen Gegenwert, der Dauer der Untersuchung und der Invasivität des Verfahrens andererseits. Für klinische Anwendungsfälle abseits der Eignungsgrenzen bestehender Bildgebungsmethoden können keine angemessenen Entscheidungshilfen gegeben werden. So müssen sich beispielsweise tausende Patientinnen mit Brustläsionen jedes Jahr einer Biopsie unterziehen, weil der Status ihrer Läsionen nicht durch ein nichtinvasives Verfahren festgestellt werden kann, trotz fortschrittlichen Vorsorgeprogrammen und etablierten Untersuchungsstandards mittels Mammographie, Sonographie, und Kernspintomographie.

Optoakustische Bildgebung ist eine einzigartige, nichtinvasive Technologie, die optischen Kontrast in Tiefen von einigen Zentimetern mit hoher Auflösung bietet und dabei vergleichsweise geringe Kosten aufweist. Dadurch liefert die Optoakustiktomographie ansonsten nicht verfügbare funktionale and anatomische Messungen und komplementiert somit bestehende Bildgebungsmodalitäten. In den letzten dreißig Jahren ist die Optoakustiktomographie kleiner Tiere zu einem wertvollen Instrument der präklinischen Forschung geworden und unterstützt mit hochqualitativen Bildern neuartige biomedizinische Erkenntnisse. Allerdings wurde die Übertragung ins klinische Umfeld durch eine bislang unzureichende Bildqualität beeinträchtigt. Im Gegensatz zu Kleintierscannern arbeiten die klinischen Systeme mit einer eingeschränkten Winkelabdeckung der Lichtabgabe und des Ultraschallempfangs, was zu Signalunvollkommenheiten führt, die durch einfache Rekonstruktionsalgorithmen verstärkt werden. Darüber hinaus verschlechtert die Bewegung der tragbaren Bildgebungsgeräten die Aufnahmequalität, wodurch die Spektralmessungen verfälscht werden und eine Rauschreduzierung durch Signalmittelung verhindert wird.

Um diese Einschränkungen zu überwinden und optoakustische Bilder in ausreichender Qualität für klinische Anwendungen zu bieten, haben wir mehrere Bildverarbeitungs- und Visualisierungsverbesserungen entwickelt, die zusammen eine multispektrale Optoakustisch-ultraschalltomographie Pipeline der „zweiten Generation“ bilden. Die Pipeline nutzt einen Bewegungsquantifizierungsalgorithmus für die automatisierte Auswahl eines stabilen Einzelbildes; ein modellbasiertes Inversionsverfahren inklusive Systemimpulsantwortkompensation; einen Rahmenmittelungsalgorithmus kombiniert mit einem Koregistrierungsschritt; und einen adaptiven, nichtlinearen Kontrastanpassungsfilters. Die resultierenden Bilder erreichen ein neues Qualitätsniveau für einen tragbaren optoakustischen Scanner und visualisieren Gefäßmerkmale tief im Gewebe mit einer hohen Auflösung von 200 μm.

Der klinische Mehrwert der verbesserten Bildgebungstechnologie besteht in ihrer Fähigkeit, Merkmale *in vivo* zuverlässig zu erkennen, die sonst nicht verfügbar wären. Der Mehrwert wird aus zwei Perspektiven untersucht: Erstens wird die Möglichkeit der Feststellung von quantitativen morphologischen Biomarkern mittels automatisierter semantische Bildsegmentierung geprüft. Unsere Experimente mit Fundus-Bildern als Proxy-Datensatz zeigen die Vorteile des selbstüberwachten Lernens, insbesondere die Reduzierung des Bedarfs an manuell annotierten Daten. Zweitens werden die Vorteile der neuen Technologie im Zuge von zwei onkologischen klinischen Studien zu Brustkrebs, beziehungsweise zu Mundkrebsmetastasen ausgewertet. Mit Hilfe der Histopathologie werden die Bildmerkmale charakterisiert, die bei Präsenz von Brusttumoren beobachtet werden können. Weiterhin wird die Erkennungsmöglichkeit der Lymph-

knotenmetastasen mit Hilfe eines tumorspezifischen Kontrastmittels untersucht und ein alternativer Bösartigkeitsmarker auf der Basis von intranodalen Desoxyhämoglobinvarianz wird vorgestellt.

Insgesamt stellt diese Dissertationsarbeit die zweite Generation der Methoden der optoakustischen Tomographie vor und definiert einen verbesserten Standard für zukünftige klinische Anwendungen.

Acknowledgements

I could not have finished my dissertation without the support from other people, and I am grateful to every one of them:

I am grateful to Prof. Vasilis Ntziachristos, who once said that great things happen when expertise in one field gets translated and applied in another one. As my supervisor, he challenged me to leave the ideal world of perfect, clean datasets we computer scientists like to live in, roll up my sleeves, and get my hands dirty with real clinical data, where often one half is missing and the other cannot be trusted. He pushed me from developing algorithms with hypothetical properties to finding solutions to very specific problems. I had to step out from every comfort zone I used to have to keep learning and keep growing. I am very thankful for this opportunity.

◆

I am grateful to Dominik Jüstel, who guided me on this journey as the best group leader I could have wished for. He gave me a great freedom to pursue my own scientific curiosity as well as the freedom to fail—and learn from my failures. He supported my ambitions in all possible ways and provided council whenever I was at my wits' end. He taught me how to focus on the things that matter and how to stay on track when pursuing my goals in the labyrinth of scientific research.

◆

I am grateful to my fiancé Ellena for accepting my PhD as a third party in our life and supporting me through and through. I am beyond grateful that she brought to this world our son Alan, who reminds me every day what matters the most in life. Also, he has shown me that I can survive (and write a PhD thesis) with much less sleep than I ever thought.

◆

I am grateful to my colleagues at IBMI, who have been a wonderful bunch to work with, celebrate with, and support each other at difficult times. Special shout-out belongs to Suhanyaa, my "PhD-sister", who was always there to share the best laughs. I cannot but thank Bob "The Writer" Wilson, who mentored and pushed me while writing my *opus magnum* paper on breast cancer. I am glad that I could have met Miguel (now a professor!), who has been a great coffee-drinking and road-biking buddy. Discussing science and the intricacies of academia with him has always been a joy, especially onboard a train heading to the mountains.

◆

I am grateful to Prof. Patrick van der Smagt for being a mentor to me, providing advice, inspiration, feedback, and encouragement at the various stages of my PhD studies.

◆

I am grateful to my parents, who have patiently kept asking all those years when I am going to be finally done studying and never lost interest in their son abroad.

◆

Finally, I am grateful to all my friends and Dave, who have kept making my life a wonderful experience.

Table of contents

List of figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of tables

Chapter 1

Chapter 3

Chapter 6

Chapter 7

List of acronyms

#	2D	Two-dimensional
	2G-OPUS	Second generation optoacoustic-ultrasound tomography
	3D	Three-dimensional
A	ANSI	American National Standards Institute
	ANTs	Advanced Neuroimaging Tools
	AUC	Area under receiver operating characteristic curve
	AUPRC	Area under precision-recall curve
B	BN	Batch normalization
	BU	Blind unmixing
C	CI	Confidence interval
	CLAHE	Contrast-limited adaptive histogram equalization
	CNN	Convolutional neural network
	CNR	Contrast-to-noise ratio
	CPU	Central processing unit
	CT	Computed tomography
D	DCIS	Ductal carcinoma in situ
E	EGFR	Epithelial growth factor receptor
	END	Elective neck dissection
	ER	Estrogen receptor
	EX	Hard exudates
F	FBP	Filtered back-projection
	FDA	U.S. Food and Drug Administration
	FN	False negative
	FO	Fovea
	FOV	Field of view
	FP	False positive
	FWHM	Full width at half-maximum
G	GPU	Graphics processing unit
	GUI	Graphic user interface

H	Hb	Deoxyhemoglobin
	HbO_2	Oxyhemoglobin
	HE	Hemorrhages
	HER2	Human epidermal growth factor receptor 2
I	IBC	Inflammatory breast cancer
	ICA	Independent component analysis
	ICG	Indocyanine green
L	LCN	Local contrast normalization
	LCO	Lower cut-off
	LN	Lymph node
M	MA	Microaneurysms
	mAP	Mean average precision
	MB	Model based
	MC	Monte Carlo
	MCC	Matthews correlation coefficient
	MoCo	Momentum contrast
	MRI	Magnetic resonance imaging
	MS	Multispectral
	MSOT	Multispectral optoacoustic tomography
N	NAT	Neoadjuvant chemotherapy
	Nd:YAG	Neodymium-doped yttrium aluminum garnet
	NIR	Near infra-red
	NNMF	Non-negative matrix factorization
	NST	No special type
O	OA	Optoacoustic
	OAM	Optoacoustic microscopy
	OAT	Optoacoustic tomography
	OCT	Optical coherence tomography
	OD	Optic disc
	OOP	Out-of-plane
	OPUS	Optoacoustic-ultrasound
P	PAM	(Twente) Photoacoustic mammoscope
	PCA	Principal component analysis

	pCR	Pathological complete response
	PDE	Partial differential equation
	PET	Positron emission tomography
	pp	Percentage points
R	RAM	Random access memory
	ReLU	Rectified linear unit
	RGB	Red-green-blue
	ROI	Region of interest
S	SE	Soft exudates
	SIS	Semantic image segmentation
	SLN	Sentinel lymph node
	SN	Sigmoid norm
	SNR	Signal-to-noise ratio
	sO_2	Oxygen saturation
	SPECT	Single photon emission computed tomography
	SSIM	Structural similarity
T	TIR	Total impulse response
	TP	True positive
U	UM	Unsharp masking
	US	Ultrasound
X	XC	Cross-correlation
	XRM	X-ray mammography
Z	ZNXC	Zeroed-normalized cross-correlation

CHAPTER 1

Introduction

1.1 Breast cancer—screening, diagnosis, and therapy monitoring

In Germany, on average, a woman dies to breast cancer every 28 minutes [1]. Despite a relatively high 5-year survival rate of 88% (cf. only 10% in pancreatic cancer), it is the leading cause of cancer-related death in women, accounting for nearly 18% of cancer-related deaths. Overall, cancer is responsible for nearly a quarter of all deaths in Germany [2]. With such epidemiologic statistics, it is no wonder that breast cancer patient care attracts considerable attention at all stages—from screening and diagnosis down to therapy monitoring.

German healthcare offers a mammography screening to women between 50 and 69 years old every two years. Three percent screening examinations are suspicious and 1.1% require an additional core needle biopsy, yet nearly half of those biopsies are benign [3]. In the total volume of 2.8 Mio screening examinations per year, 14 000 benign biopsies are performed. Despite being minimally invasive, the core needle biopsy procedure is still painful, and the uncertainty of the outcome induces mental stress for the patient [4]. Furthermore, about 10% of the screening-related biopsies in Germany detect in situ lesions [3], many of which are harmless but are treated nonetheless due to the risk of developing into an invasive cancer. Inability to properly select high-risk precancerous neoplasms results in overtreatment and complications stemming from excessive surgeries and therapies [5, 6]. Novel non-invasive, point-of-care examination methods could help with more precise diagnosis of suspicious findings, reduction of benign biopsies, and risk stratification of in situ lesions.

Six permille of screening examinations reveal an invasive breast cancer requiring a treatment. Depending on the tumor stage and its genetic and molecular properties, a systemic chemotherapy may be indicated. If the tumor nature allows it, neoadjuvant chemotherapy (NAT) is applied prior to the surgical removal of the tumor. The aim of NAT is reducing the tumor size to allow a breast-conserving surgery instead of a mastectomy [7]. Although the therapy regime is tailored to each patient individually, monitoring its effectiveness is not easy. One possibility is evaluating the pathological complete response (pCR) after NAT. pCR is indicated upon the surgical excision as a full disappearance of cancer cells from the primary tumor site. pCR is a sign of therapy effectiveness and a positive prognosis factor [8]. However, the NAT is applied over several months before the surgery occurs and pCR can be evaluated. Having access to instant monitoring of the therapy efficiency would enable improved personalized therapy management and reduce side effects of chemotherapy [9].

1.1.1 Limitations of present imaging modalities

Radiologists in breast healthcare have a range of imaging modalities available. Whereas screening is done nearly exclusively using X-ray mammography (XRM), diagnostic imaging is performed also using ultrasonography, magnetic resonance imaging (MRI), or computed X-ray tomography (CT). Less frequently, positron emission tomography (PET), single photon emission computed

tomography (SPECT), or scintimammography are also employed to provide functional parameters about intake of tracers. For intraoperative navigation, ultrasonography or gamma cameras are typically used. Table 1.1 summarizes the characteristics of these imaging modalities.

For the two clinical scenarios outlined above—diagnostic imaging and therapy monitoring—the current available imaging modalities have certain limitations which are discussed next.

Despite its wide acceptance, XRM may produce false positive findings. Due to the nature of mammography, the images are formed as projections of tissue volume onto a plane and overlaying structures may form patterns resembling malignant lesions [10]. Furthermore, fat necrosis following a trauma or a surgery, benign lesions, and in situ proliferations may also mimic malignancy [11]. In the screening regime, the false positive rates of mammography screening are 2–15%, with lower numbers for subsequent screens (images from previous visits are available for a reference) and Europeran Union countries (regulatory requirement for ≤5% false positive rate) [3, 10]. However, women with dense and more complex breast tissue, younger age, and higher estrogen level are much more susceptible to false positive findings [12]. In this group of women, ultrasonography is a preferred diagnostic imaging modality [13].

Ultrasonography (also ultrasound, US) is an imaging modality used for diagnosis of clinically detected lesions, screening recalls, and sometimes for screening of patients with dense breasts [14]. US does an excellent job at distinguishing (benign) liquid-filled cysts from solid masses [15]. Moreover, US lesion parameters such as shape, margin, boundaries, orientation, and presence of spiculations allow classification of some solid masses as benign or malignant [15, 16]. Generally, the quality of an US examination is dependent on the operator [17]. A widespread utilization of US for screening is prevented by its low specificity [18]. Combined US and mammography screening results in up to 20–30% benign biopsy findings [14, 19]. Similar false positive rates of US examinations are observed also in diagnosis of symptomatic patients [20, 21]. For diagnosis of screening recalls, i.e. lesions that could not be classified based on mammography examination alone, a study conducted in Sao Paulo reported a false positive rate as high as 71% [22]. Despite the broad consensus on the value of US imaging in breast lesion classification, these numbers leave a large space for improvement.

Together with XRM and US, MRI is frequently used for breast cancer imaging. It is recommended as an adjunct screening modality to XRM in high-risk patients and for additional evaluation of suspicious findings and diagnosed lesions [23]. The breast imaging MRI protocols rely on intravenously administered contrast agent (gadolinium chelate) triggering rapid local enhancement

Table 1.1: Summary of properties of clinical imaging modalities.

	Resolution	Image type	Scanner cost	Probe medium
XRM	50 µm	Anatomic	€ 10s-100s k	X-rays (ionizing)
US (Doppler)	<1 mm	Anatomic (Functional)	€ 1s-10s k	Pressure waves
CT	0.5 mm	Anatomic	€ 100s k-1s mil	X-rays (ionizing)
MRI	<1 mm	Anatomic Functional Molecular	€ 100s k-1s mil	Magnetic field alterations, Radio waves
PET	~1 mm	Functional	€ 100s k-1s mil	γ-radiation (ionizing)
SPECT	~5 mm	Functional	€ 100s k	γ-radiation (ionizing)
MSOT	200 µm	Functional Molecular	€ 10s k	Pulsed light, pressure waves

of T1-weighted images around abnormally permeable tumor-related blood vessels [24]. MRI has been recognized to have a lesion detection sensitivity superior to XRM and US (89%–100%) but only moderate specificity (65%–86%) due to similarities in patterns observed in malignant and benign lesions [25]. As a result, a routine use of MRI pre-treatment has been a subject of a controversy, often warranting a more radical surgical approach despite no improvement in the oncological outcome [25]. Finally, as shown in Table 1.1, MRI is associated with much higher equipment costs that XRM or US, limiting its widespread availability and cost-effectiveness.

In neoadjuvant therapy response assessment, change in the tumor size is typically considered as a proxy measure to pCR available to imaging [8, 9]. However, XRM is not reliable at assessing the tumor size and although US may be a better choice, both methods tend to underestimate the true lesion spread [8]. Experiments with Doppler US have shown that observing vascularization and blood flow in the tumor region may be better predictive features for a pCR [26]. Diffuse optical tomography has also shown promise in predicting the therapy response, especially in patients with well-perfused tumors [27]. Alternatively, nuclear imaging methods (PET, scintigraphy) can be used to monitor radioactive tracer uptake, which may correlate with a therapeutic drug uptake. Furthermore, pre- and post-therapy changes of the uptake may also relate to the response [9]. Overall, MRI is usually the imaging modality preferred for NAT response assessment over XRM and US, although the accuracy varies between histopathological subgroups [9].

1.2 Optoacoustic tomography

Optoacoustic[1] tomography (OAT) is an emerging modality combining several unique properties to overcome the limitations of the existing imaging approaches. Optoacoustic imaging uses light absorption as the contrast mechanism and US waves for signal detection. Since the US waves are not affected by optical scattering, the limits of the OA imaging resolution are dictated by acoustic diffraction. Hence, OAT enables imaging with high resolution (hundreds of microns) at depths of several centimeters. Unlike XRM and CT, OAT does not involve ionizing radiation. The cost of OAT devices is an order of magnitude lower than in the case of MRI and CT scanners. Furthermore, OAT devices are far more portable than MRI and CT scanners and require only moderate adjustments to the premises to adhere to the laser safety standards. Commercially available OAT devices for clinical research come in the form of hybrid optoacoustic-ultrasound scanners. In that form, they expand the capabilities of a standard grayscale ultrasound, routinely used for breast imaging, with optoacoustic features. As such, they might be readily incorporated into existing breast care infrastructure.

The main endogenous absorber visualized by OAT is hemoglobin. OAT excels at visualizing blood vessels and provides images of vascular patterns around and inside breast lesions without the need for any contrast agents. Rapid growth of new blood vessels—angiogenesis—is a process accompanying cancer and a potential optoacoustic biomarker of tumor malignancy [29-31]. Besides showing blood vessels, OAT also has the promise of quantifying the local oxygen saturation. Reduced oxygen saturation—hypoxia—is also linked to malignancy and poor prognosis [32-35]. Furthermore, changes in the oxygen saturation levels in response to a systemic treatment may be early signs of the therapy response [27, 33, 36].

Besides visualizing endogenous molecules, OAT can visualize further biological processes with the aid of exogenous reporters. Those reporters are typically formed as conjugates of a signaling

[1] Also known as photoacoustic tomography, sometimes thermoacoustic tomography. A historical account of the emergence of multiple terms for the same technology used in parallel by different research groups around the world can be found in the review by Manohar and Razansky [28].

compound producing OA signal and a targeting moiety ensuring biological specificity of the probe. Small-molecule dyes, metallic or semiconducting nanostructures, and organic nanostructures are the main classes of signaling compounds. For targeting, small molecules, peptides, adhirons, affibodies, aptamers, and proteins can be used [37]. The probes may target cell surface receptors overexpressed in tumor cells. For example, targeting of the epidermal growth factor receptor was demonstrated using probes based on the CF-750 near-infra red (NIR) dye [38] or golden nanoparticles [39]. Alternatively, "smart" probes, which are activated by enzymes, can be used to detect tumors cells with increased enzyme levels [40] or target certain biological processes [41]. For example, probes activated by caspases—enzymes essential to apoptosis—were shown as potential markers of chemotherapy efficiency [41, 42]. As of now, however, there are no FDA-approved probes specific to OA imaging and existing approved optical agents, such as indocyanine green, have relatively low extinction coefficients compared to golden nanoparticles and other specialized agents. Clinical usage of nanoparticle-based agents is usually limited by their biocompatibility, such as prolonged accumulation in the reticuloendothelial system [37].

1.2.1 State-of-the-art

OAT has been first demonstrated in 1994 in works of Oraevsky [43] and Kruger [44]. Whereas these early prototypes used as little as 12 US transducers and produced images of poor quality, current systems with hundreds of transducers achieve sub-millimeter resolution, several centimeters of depth penetration, and obtain images at multiple wavelengths [45-48]. Over the past thirty years, OAT has been used in numerous clinical applications: breast cancer imaging [45-59], imaging of the cardiovascular system (vascular function in the extremities and carotid plaques assessment) [60-66], musculoskeletal system [67, 68] and adipose tissue function [69], gastrointestinal imaging [70], imaging of the lymphatic system [71-75], thyroid [76-83] and prostate [84, 85], skin lesions [86-89], and intraoperative guidance [90-94]. Recently, the first OAT scanner has received an FDA approval for diagnostic use in breast cancer patients [95]. A detailed overview of clinical applications of OAT is presented in Chapter 2, section 2.4.

OAT systems come in a whole range of implementations, besides optoacoustic imaging devices of other scales, such as optoacoustic microscopes [96, 97], mesoscopy systems [98, 99], and endoscopes [100]. Pre-clinical OAT systems are mature and offer good imaging quality for small animal scanning owing to low requirements for penetration depth and feasibility of ultrasound detection over the full angular coverage [101-103]. Clinical devices come either as a single-purpose stationary scanners or as multi-purpose handheld devices [104]. Several stationary scanners exist that were developed for breast imaging. These scanners typically take the form of an examination bed with a scanning chamber for the pendant breast [45, 47, 48] or the form of a mammograph [105]. Stationary scanners have the advantage of potentially larger detector aperture, complex light delivery system, and are less sensitive to motion as the breast is immobilized in the scanner. Furthermore, they can provide full 3D scans of the breast and suffer low inter-operator variability.

Handheld clinical OAT devices emerged as an extension of standard handheld ultrasound linear probes by attaching additional laser light delivery through optical fibers [106]. Probes with custom-built detector arrays in curvilinear arrangement appeared later, better fitting to the specific needs of optoacoustic imaging and producing images with less artifacts [66, 107]. In both cases, since the probes contain all the hardware required for US imaging, they can operate in a hybrid mode and provide co-registered OA and US images simultaneously [108]. This is advantageous since the US provides anatomical images which are enriched by functional parameters provided by OA. On the other hand, the narrow-band sensitivity of the available piezoelectric transducers makes them suboptimal for detection of optoacoustic signals that are broadband in their nature [109, 110] and better broadband detectors are needed, such as optical resonators [111, 112]. Overall, the handheld systems offer various clinical utilizations and operate in a fashion similar to the standard clinical

US. As such, handheld systems are more accessible to existing radiology personnel and may fit better into established clinical workflows.

1.3 Outstanding challenges for OAT

Despite its great promise, OAT is facing several challenges which hinder achieving its full clinical potential. Here, we outline the challenges which are in the center of focus of this thesis.

1.3.1 Motion

Clinical optoacoustic tomographic imaging with handheld probes is a subject to motion stemming from the handheld probe operation and the physiological motion of the patient, e.g., breathing and vessel pulsation. Motion occurring between individual laser pulses results in misalignment of consecutive measurements, preventing their averaging and corrupting spectral information when multiple illumination wavelengths are used. Possible ways to minimize the negative effects of motion include stationary imaging systems, shorter acquisition schemes using less wavelengths, or performing motion correction as an image processing step. Furthermore, quantification of motion in a sequence of images would enable automatically selecting ones with the least motion.

Chapter 3 covers the issue of motion in OAT. A novel method for quantification of motion in multispectral optoacoustic-ultrasound (OPUS) scans—Motion score—is presented and validated on scans of an agar phantom and healthy volunteers. It is demonstrated that using Motion score facilitates automated frame selection with optimal spectral quality needed for the later image analysis steps.

1.3.2 Light fluence attenuation

The core principle of OA imaging is detection of US pulses produced upon absorption of light by molecules in the tissue. The US pulses are produced only in the regions of the tissue that are sufficiently illuminated. Due to photon scattering and absorption in the biological tissue, the light intensity decreases approximately exponentially with the tissue penetration depth. The maximum light intensity that can be safely applied to the human skin within a period of time is regulated by the American National Standards Institute (ANSI) light exposure limits and cannot be increased over a safe threshold [113]. Thus, the maximum light penetration depth has a strict limit, which lies around five centimeters, depending on the imaged tissue [114].

Besides limiting the maximum achievable imaging depth, the attenuation of the light fluence has several other implications for OA imaging.

Image quality
As the light fluence decays, the signal is getting weaker, and noise becomes an issue in greater depths. Low signal-to-noise ratio (SNR) and reduced image quality are also aggravated by imperfections of the applied reconstruction algorithms if they fail to faithfully capture the real physical properties of the scanner and the imaged tissue. In such cases, the noise gets amplified in the form of imaging artifacts.

Chapter 4 introduces advanced image processing tools for improvement of the image quality. First, advanced image reconstruction algorithms are presented, that ensure improved image quality. Then, a method for improvement of the image contrast based on motion correction and frame averaging is presented. Together, the proposed image processing pipeline forms the second-generation multispectral optoacoustic-ultrasound tomography (2G-OPUS), leading to an unprecedented image quality of handheld OAT and paving the way for novel clinical applications.

Pixel value visualization and interpretation
The distribution of light fluence intensities in the field-of-view follows an exponential distribution since the fluence decays exponentially with the depth. The product of the light fluence and the absorption coefficient dictates the OA signal strength, which in turn defines the pixel values in the reconstructed image. Assuming the absorption coefficients are distributed uniformly in the tissue, the distribution of the pixel values also follows an exponential distribution. Visualization of images with such heavily skewed pixel value distribution is difficult and prone to over- and undersaturated regions.

Additionally, such images are difficult to use with machine learning algorithms since image patterns appear on different intensity scales. Absolute image intensities carry little information except about the depth, which is typically not of interest.

Chapter 5 presents optimized methods for image visualization in OAT.

Image analysis
The ultimate goal of multispectral optoacoustic tomography is distinguishing various absorbers and recovering their concentrations in the tissue. Standard spectroscopic methods as linear unmixing do not provide precise results due to *spectral coloring* problem. Spectral coloring is caused by wavelength-dependent, spatially uneven light fluence attenuation: The illumination spectrum is different for each point in the tissue—it has a different color. In presence of spectral coloring, the recorded optoacoustic spectra are not a linear combination of the absorption spectra of the present absorbers; instead, they are convolved with the (unknown) light fluence spectra.

The difficulty in deconvolving the two spectra is the existence of multiple configurations that produce the same signal—the problem is ill-posed and recovering the correct solution requires additional information to constrain the solution space. A powerful constraint could be formed by segmenting the image into regions sharing similar chromophore distribution, such as blood vessels and subcutaneous fat. **Chapter 6** focuses on image segmentation in OA images and presents a self-supervised learning approach to train segmentation models in absence of large, annotated datasets.

1.3.3 Clinical feature discovery

Clinical optoacoustic tomography is a relatively novel imaging modality and most of the clinical studies up to date are of explorative nature, finding out potential imaging targets. Suitable target features need to be established and validated on larger studies.

Chapter 7 focuses on two clinical applications of OAT: imaging of breast cancer and imaging of oral cancer lymph node metastases. This chapter presents the results of clinical application of the methods developed in the earlier chapters.

1.4 Structure of the thesis

This thesis is structured into the following chapters. Chapter 2 summarizes the background information relevant to the rest of the thesis. Historical perspective is provided to highlight the importance of medical imaging using light since the nascence of specialized imaging devices, and to summarize the milestones leading to the current state of OA imaging. Clinical applications of OAT are briefly reviewed. Physical principles behind OA imaging are summarized and standard methods for image formation and analysis are explained. Chapters 3–7 cover the proposed solutions to the challenges outlined above and present the experimental results. Finally, Chapter 8 concludes this thesis with a summary of the main findings and an outlook.

CHAPTER 2

Background

2.1 History of optical biomedical imaging

Light has been used as the probing medium for biological and medical imaging from the very nascence of specialized devices expanding the viewing capabilities of the human eyesight. Around the turn of the 17th century, microscope was developed (Fig. 2.1a) [115]. Using one or more glass lenses, the microscope alters the angles of propagation of visible light to magnify viewed objects, enabling the examination of details that are too small for the human eye to see. The pioneers of microscopy—Robert Hooke, Antonie van Leeuwenhoek, and others—used the early microscopes during the 17th century to observe biological samples up to a single-cell resolution. In fact, Hooke was the first to describe a biological cell in his work *Micrographia* and coined the term itself, alluding to the similarity of biological tissue organization to a honeycomb[2]. Leeuwenhoek and Hooke provided the first microscopic drawings of insects, microorganisms, plants, bacteria, and discovered important cell types such as erythrocytes and spermatozoa [115, 117]. Microscopes have played an essential role in science ever since and they are still an everyday tool in nowadays histopathological clinical practice.

Hand in hand with microscopy, dyes have been developed for staining microscopy specimens. Dyes bind to different molecules (proteins), highlighting specific cell types and organelles. van Leuwenhoek used saffron, von Gerlach introduced carmine, Wissowzky described hematoxylin-eosin stain, to name a few. Chemical properties of dyes complement the physical function of the microscope, greatly expanding the scope of possible applications.

Another early medical imaging device is otoscope—aural speculum. It is used to inspect the ear canal and the eardrum. Although its early implementations in the form of tongs appear parallel to the microscope, it only became a true imaging/viewing device with light delivery and magnification capabilities in the 19th century. The solution proposed by John Brunton in 1865 (Fig. 2.1b) resembles modern otoscopes (up to integration of the light source) [118], but a simpler design popularized by Arthur Hartmann was adopted at the time, consisting of a cylindrical cone-shaped speculum and a separate light-delivery in the form of a concave mirror with a hole to look through, originally devised by Hofmann in 1841 [119, 120]. Brunton's otoscope instead featured an integrated hollow mirror for light delivery. The advantage of leaving the funnel unobstructed by a lens and a mirror is the option to guide instruments through, which is why also nowadays' otoscopes typically have a removable magnification lens.

The advance in medical viewing instrumentation in the 19th century is manifested in another invention from the same time. In 1851, Herman Helmholtz presented *Augenspiegel*—the first ophthalmoscope [121]. The device (Fig. 2.1c), consisting of a transparent mirror and a lens,

[2] A common historical misconception is that the term *cell* originates from similarity of biological cells with monk's living quarters, which is, however, not mentioned anywhere in Hooke's writings [116].

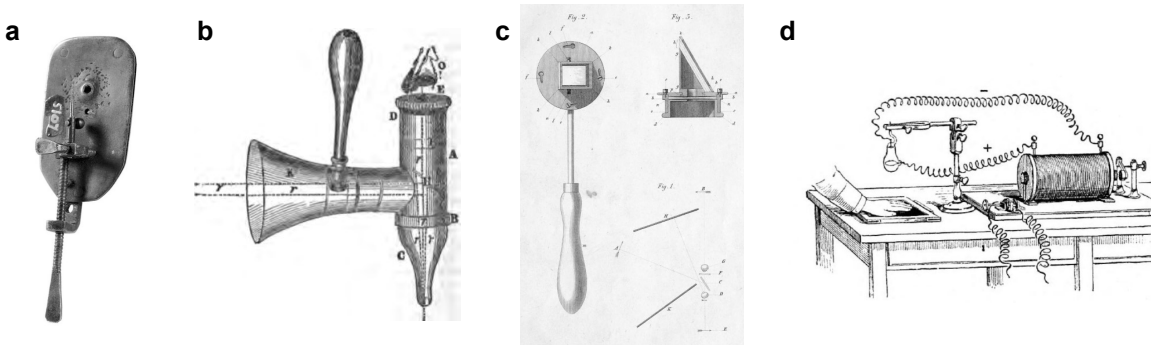

Figure 2.1: Historical medical imaging (viewing) instruments. a, Leeuwenhoek's microscope. **b,** Brunton's otoscope. **c,** Helmholtz' *Augenspiegel*. **d,** X-ray device for obtaining radiographs of hand. Image sources listed in the section Permissions.

enabled Helmholtz for the first time ever to view the human retina *in vivo*. He reported seeing branching networks of arteries and veins and the optic disc, and correctly predicted that the ophthalmoscope would play a major role in understanding the changes in the eye accompanying diseases leading to blindness.

Nowadays, to inspect the morphology and function of the vascular network of the retina, fluorescein angiography is used. Fluorescein, a fluorescent dye, is first injected into the patient's blood stream. The retina is observed with a specialized camera delivering ultraviolet light illumination that excites the fluorophore, and a filter that allows only the fluorescent light through. This way, the clinician can observe how the blood carrying the fluorescein propagates in the retina and may spot anomalous regions.

A true revolution in medical imaging came in 1895 with the discovery of x-rays by Wilhelm Conrad Röntgen [122]. While studying cathode rays (electron beams), Röntgen noticed fluorescence on a barium platinocyanide-coated sheet triggered by operation of a nearby Crookes vacuum tube. The fluorescence was not caused by luminescence, since the tube was fully covered by opaque cardboard, nor by cathode rays that are quickly attenuated in the air outside the tube. Röntgen examined this phenomenon further and concluded that the fluorescence is caused by a new, unknown type of radiation, which he named *x-rays*. He quickly noticed that x-rays penetrate a whole variety of materials, including glass and metals, whereas other materials, such as lead, cast shadows on fluorescence screen or photo plates, implying the radiation in fact behaves as rays. Importantly, he noticed that soft tissue is transparent to x-rays whereas bones are not, allowing seeing the bones inside a living patient. The clinical potential of the discovery was huge and the first diagnostic x-ray images—radiographs—were obtained already within few months [123]. A setup of an early x-ray imaging device is shown in Fig. 2.1d.

To image other structures than bones using x-rays, special contrast agents have been used. Barium sulfate solution can be used to view the bowel, and iodine-based agents can be injected into the blood stream to make the blood vessels visible on the x-rays.

2.2 Principles of optical imaging

Microscope, otoscope, ophthalmoscope, and radiography are all examples of the first medical imaging (or viewing) devices. They are all routinely used in clinics ever since their invention. A property they share is that their function is based on the interaction of electromagnetic radiation (such as visible light, Fig. 2.3) with the examined tissue and detection of the altered radiation via human eye or a specialized detector. Additional chemicals, such as dyes and contrast agents, greatly

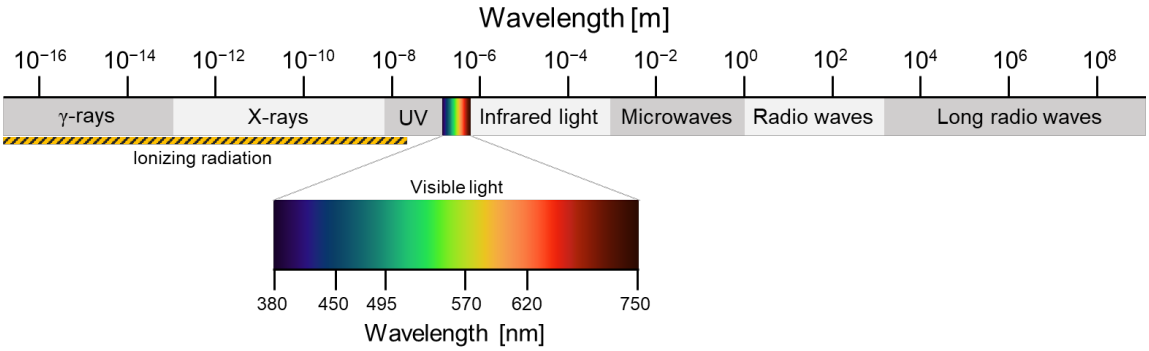

Figure 2.3: Electromagnetic spectrum. Only small part of the electromagnetic spectrum (380–750 nm) takes the form of the visible light. However, other types of electromagnetic waves can be also used to probe biological tissues. Abbreviations: UV, ultraviolet.

enhance the capabilities of those techniques since they link specific biological functions to specific physical properties accessible to electromagnetic radiation interrogation. For example, in fluorescein-based fundus angiography, the functional properties of blood flow are linked to presence of fluorescence. In histology, hematoxylin dye binds to chromatin in cell nuclei, linking their presence to purple color.

Modern optical biomedical imaging methods rely on the same principles as the early techniques. They typically utilize electromagnetic radiation in the visible and infrared light range. In this wavelength range, biological tissue exhibits low absorption which means the light can penetrate relatively deep into the tissue. However, the tissue is a strongly scattering (turbid) medium. Elastic photon scattering is an event in which a photon is absorbed and re-emitted into some other direction without the loss of energy. The photon propagation in a turbid medium can be characterized by the mean free path, denoting the average photon travel distance between two scattering events, and the transport mean free path, denoting the distance after which the direction of the photon becomes effectively random (Fig. 2.2). The transport mean free path of light in biological tissues is in the order of 1 mm.

High scattering has two important implications for optical imaging methods. First, the amount of energy arriving at a point in tissue is decreasing exponentially with the depth because photons are more likely to get either absorbed or backscattered before reaching the point. Second, since most of the photons lose their direction already after passing through ~1 mm of the tissue, it is impossible

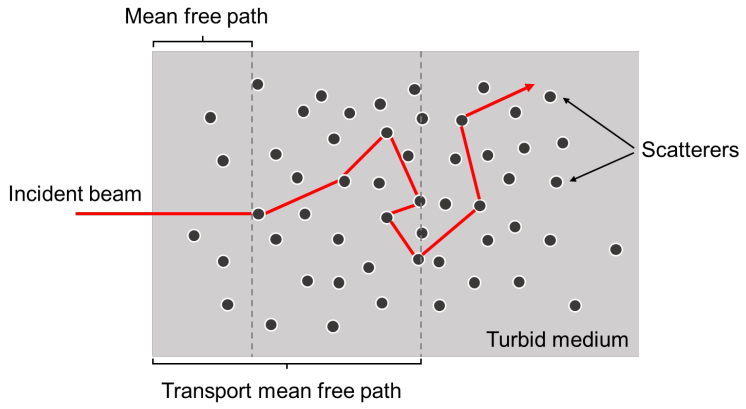

Figure 2.2: Light transport in scattering medium. Scatterers are depicted as black circles. Two parameters characterize the photon transport: Mean free path refers to the mean distance a photon travels in the medium before a scattering event occurs. Transport mean path refers to the mean distance a photon travels in the medium before its direction becomes effectively random.

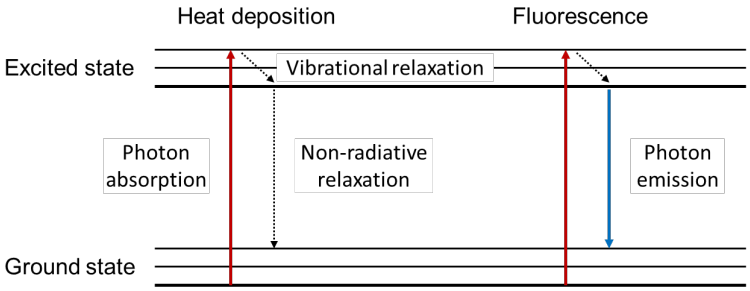

Figure 2.4: Jablonski diagram of photon absorption and relaxation in form of heat deposition and fluorescence.

to focus the illumination into a certain part of the tissue with high precision. Correspondingly, when detecting photons that have interacted with the tissue, it is impossible to tell with high precision where they came from. I.e., one can detect how the tissue altered the photons, but one cannot localize where the interactions occurred in the tissue.

Light may interact with the tissue via the following physical mechanisms—absorption, scattering, polarization, and frequency shift—and each can provide unique information about the probed tissue.

Light absorption, in particular, is related to molecular conformation and thus provides biochemical information. Biomolecules in the tissue are characterized by their unique absorption spectra and spectroscopic approaches can be used to distinguish them. Human tissue contains several strongly absorbing molecules—*chromophores*: hemoglobin, melanin, water, and lipids. Hemoglobin is a protein present in red blood cells that facilitates oxygen transport. Depending on the presence of a bound oxygen molecule, hemoglobin is present either in the oxyhemoglobin or deoxyhemoglobin form, each having a distinct absorption spectrum. Quantification of the oxy- and deoxyhemoglobin ratio in the blood, the oxygen saturation (sO_2), is an important metabolic biomarker accessible to optical imaging.

Upon absorption of a photon by one of a molecule's electrons, the electron enters an excited state and can only return to the ground state by releasing the absorbed energy in a process called relaxation (Fig. 2.4). Relaxation can be non-radiative, when the energy is released in the form of heat, or radiative, such as fluorescence, when the electron emits a photon with a lower energy than it had absorbed earlier. The absorption interaction can be detected by any of the three mechanisms: measuring the residual, unabsorbed light leaving the sample, measuring the photons emitted by fluorescence, or measuring the deposited heat.

Optical imaging methods relying on the detection of the photons leaving the sample (either residual or fluorescent) suffer from the above-described limitations of reduced depth penetration and limited directivity resulting in low resolution. The last type of interaction, heat deposition, is utilized by optoacoustic imaging.

2.3 History of optoacoustic imaging

Heat deposition upon the absorption of a light pulse triggers a transient volume expansion producing a mechanical pressure impulse—sound. This is known as the optoacoustic (photoacoustic) effect and was first described by Alexander G. Bell and his assistant Charles S. Tainter in 1880 in their work on photophone [124]. The size of the absorber determines the wavelength of the produced sound wave [109]. Sensing the (ultra-)sound waves produced upon the illumination of a tissue by a light pulse is the core mechanism of optoacoustic imaging. Unlike

photons, the ultrasound pressure waves are not affected by optical scattering in the tissue. The acoustic attenuation is two to three orders of magnitude lower than the optical attenuation and depth is not the main limiting factor [125]; however, acoustic coupling between the probed sample and the detectors is required. Finally, tomographic reconstruction of the recorded ultrasound waves allows recovering the absorber locations in high resolution.

Bell's photophone discovery in 1880 sparked immediate scientific interest and multiple works have emerged, investigating the exact nature of the sound production in response to periodic (chopped) illumination. Works by Mercardier, Preece [126], Röntgen, and Tyndall explained the process through expansion and contraction of thin air layer in contact with the absorbing object which acts as a periodic heat source. This model was rigorously described by Rosencwaig and Gersho almost a century later [127]. It was also soon discovered that the optoacoustic effect was not limited to solid materials and applied also to liquids and gases. However, the only demonstrated application of the effect was wireless sound transmission—which was nonetheless inferior to radio waves—and it was not before 1938 until optoacoustics found a new application domain in gas analysis.

Veingerov [128], Pfund [129], and Luft [130] presented gas analyzers based on the optoacoustic effect and revived the interest in the method. These devices detected even small quantities of radiation-absorbing gases such as CO and CO_2. Furthermore, absorption spectra of gases could be measured using monochromatic illumination with a device called *spectrophone* [131]. The idea of spectrophone dates back to Bell [132], however, the device only became practical through an improved readout sensitivity achieved with a microphone and a galvanometer. The advantage of spectrophone over a standard spectroscope was the ability to probe the absorption in the invisible (infrared) parts of the electromagnetic spectrum.

Important technical advances towards harnessing the potential of the photoacoustic effect include the detection of the produced sound through a microphone instead the human ear (and later by a sufficiently broadband piezoelectric transducer[3], allowing much higher sound frequencies to be recorded); development of better computational hardware for data recording and processing; and the discovery of the laser in 1960, which facilitated stable illumination of specific narrow spectrum.

Restored interest in photoacoustics in solid materials arrived shortly after the laser discovery, with applications to non-destructive semiconductor examination and biological tissues. Among the first biological applications were experiments with mammalian eyes [134-136]. Spectroscopic characterization of human blood and hemoglobin followed soon thereafter [137], and the method was also investigated for detection of drug incorporation in the skin [138]. Around that time, experiments with microwave-range "illumination" of biological tissues were also performed under the name of *thermoacoustics* [139].

The first reflection-mode medical photoacoustic implementation appeared in 1993 and was demonstrated on human finger and fingernail A-scan [140]. The prototype was demonstrating possibilities of using laser-induced ultrasound for intraarterial diagnosis and ablative plaque-therapy.

2.4 Clinical optoacoustic tomography

Modern optoacoustic imaging of spatial distribution of absorbers in biological tissues was demonstrated in 1994 concurrently by Oraevsky [43] and Kruger [44]. With the use of tomographic reconstruction methods, 2D and 3D optoacoustic images of biological tissues were obtained.

[3] Coincidentally, the piezoelectric effect was described by Currie brothers in 1880—the same year the optoacoustic effect was discovered [133].

Further advances in electronics, signal processing, and image reconstruction methods opened the door for a plethora of clinical applications of optoacoustic tomography, which we briefly review next. Most of the listed applications have been preceded by extensive pre-clinical research, which we take the liberty to omit in the following review.

2.4.1 Breast cancer

Breast cancer has been among the main applications of OAT due to its high prevalence and favorably shallow location of the tumors. Initial results with breast OAT scanners developed in the late 90's and early 2000's provided observations of increased OA signal in breast lesions [50, 141]. Later advances were predominantly revolving around bed-based scanners which image the breast in the pendant position using detectors arranged circularly, hemispherically, or in a plane. With sub-millimeter resolution, modern OAT scanners can resolve large blood vessels supplying the tumors with blood, as well as signals from the microvascular peri- and intratumoral environment [47, 48]. Furthermore, using multiple-wavelength illumination, OAT has the potential to spectroscopically resolve the contributions of individual chromophores, providing localized information about the oxygen saturation in the tumors. Although truly quantitative measurements have not been obtained yet—due to the difficulty of estimating the light fluence—various semi-quantitative proxy measures have been proposed, such as the *S-factor* introduced by the group around the Kyoto University–Canon consortium [59]. Overall, only one clinical OAT study has reported consistently lower oxygen saturation levels inside malignant tumors [58].

Besides bed-based scanners, clinical OAT breast cancer imaging has also been performed with handheld devices, often implemented in tandem with US imaging. Such hybrid scanner implementation is advantageous since it provides an extension of plain ultrasonography which is routinely used in the breast healthcare workflow. As result, OPUS imaging can be easily incorporated into the existing radiological practice. Clinical applications of handheld OAT to breast cancer imaging have been relying primarily on two systems—*Imagio*® (Seno Medical Instruments Inc.; Texas, USA) and *Acuity*® (iThera Medical GmbH; Germany). The *Imagio* system has been utilized in two large, multi-center studies in USA and Netherlands, which aimed at evaluating its use for adjustment of grading of suspicious breast lesions [55, 56]. Based on these trial results, *Imagio* has received an FDA approval for use in breast cancer diagnostics [95], being the first optoacoustic scanner approved for a clinical application. The *Acuity* system has been employed in a small, prospective study at University Hospital Münster, revealing increased hemoglobin perfusion of malignant tumors relative to healthy tissue [49].

2.4.2 Cardiovascular system

Hemoglobin is by far the most important contrast source in current clinical optoacoustic imaging, making blood vessels a natural imaging target for OAT. The capability of OAT to image blood vessels in extremities of human volunteers has been demonstrated in 2005 by Niederhauser *et al.* [106]. In 2012, Dima and Ntziachristos have presented multispectral images of human carotid artery and jugular veins [66]. Later, Masthoff *et al.* have used multispectral optoacoustic tomography (MSOT) to distinguish arteriovenous and venous vascular malformations based on their oxygen saturation [64]. In 2020, Karlas *et al.* have proposed MSOT for monitoring of muscle perfusion and oxygen saturation during venous and arterial occlusion as a method for detection of vascular disorders altering normal blood flow [65]. Furthermore, they have suggested that the lipid contrast provided by MSOT could be used for improved diagnostics of carotid plaques [63]. These studies were all conducted using handheld OAT providing 2D cross-sectional images. A different approach to imaging human peripheral vasculature has been proposed by the Kyoto University–Canon consortium and their bed-based 3D scanners *PAI-03* [61], *PAI-04* [60], and *PAI-05* [62].

Their images provide access to complete subcutaneous vascular network morphology and allow extraction of morphological biomarkers such as vessel tortuosity.

2.4.3 Musculoskeletal system

The clinical applications of OAT for imaging of the musculoskeletal system focused primarily on imaging of joints and arthritis. In 2007, Wang *et al.* have demonstrated the capability of OAT to present the morphology of human finger joints on *ex vivo* samples [142]. Later demonstrations have shown that OAT can localize the bone, tendons, and blood vessels in human finger joints also *in vivo* using custom-built circular detector systems [143-145] as well as linear arrays [146]. First experiments with osteoarthritic patients were published by Xiao *et al.* in 2010, showing increased OA signals in osteoarthritic joint cavities [147]. In 2017, concurrent works by van den Berg *et al.* and Jo *et al.* imaged finger joints of inflammatory arthritis patients, showing increased OA signals in presence of synovitis [67, 148].

2.4.4 Thyroid

The thyroid gland is particularly well accessible to OAT imaging owing to its shallow location on the anterior side of the neck. Two early studies on excised thyroid samples in 2014 explored the possibility of finding thyroid malignancies with OAT. Dogra *et al.* imaged thyroidectomy samples with malignant and benign nodules and found increased deoxyhemoglobin levels in malignant tissue compared to benign and healthy tissue regions [77]. On the other hand, Kang *et al.* attempted to extend their earlier work on detecting microcalcifications in breast tissue to thyroid but did not find significant contrast [76]. In 2016, Dima and Ntziachristos demonstrated handheld imaging of human thyroid *in vivo*, observing optical contrast in agreement with the anatomy of the thyroid region [80]. Besides resolving the thyroid and the surrounding muscles, they demonstrated the ability to visualize the thyroid vascularization to a higher level of detail than possible with Doppler US. A 2017 *in vivo* study on cancer patients by Yang *et al.* provided similar observations of thyroid vasculature, but has not evaluated the diagnostic capability of OAT [78]. A larger 2021 study on 52 subjects by Kim *et al.* has shown that a combined OPUS system has better sensitivity and specificity (83% and 93%, resp.) at papillary thyroid carcinoma detection than US alone (≤50%, ≥90%, resp.) [83]. Studies utilizing the *Acuity* MSOT have also shown alterations of the hemoglobin parameters in thyroid cancer nodules as well as in patients with Graves diseases [81, 82].

2.4.5 Lymphatic system

OAT has been utilized in two applications related to the lymphatic system: detection of the sentinel lymph node (LN) and detection of LN metastases. Following successful pre-clinical trials of detecting the sentinel LN using OAT with the aid of methylene blue optical contrast agent [149], Garcia-Urbe *et al.* have in 2015 successfully demonstrated the same procedure in breast cancer patients [73]. Clinical applications on LN metastasis detection focused primarily on melanomas since melanin—produced by the cancer cells—is a very strong optical absorber, providing excellent OA contrast. Jose *et al.* have in 2011 demonstrated the ability of OAT to detect melanoma metastasis in an excised human LN [71]. In 2012, the same group has presented further encouraging results on multiple excised LNs [72]. Yet the translation of the technique to *in vivo* imaging has not yet been entirely achieved. A 2015 MSOT study by Stoffels *et al.* examined to use the intrinsic melanin contrast for metastasis detection in sentinel lymph nodes (SLN) of melanoma patients [74]. In both the *ex vivo* and *in vivo* scenarios, they observed 0% false negative rate, suggesting that MSOT can be used to safely rule out a possibility of a SLN metastasis. However, a very high false positive rate, stemming from a low specificity of melanoma cell MSOT contrast, hinders the use of OAT as a method for reduction of negative LN biopsies. In 2017, Kang *et al.*

have presented a pre-clinical proof-of-concept of using a combination of fluorescence imaging with OAT for SLN detection [92]. In a 2019 study, Stoffels *et al.* examined the ability of a combined MSOT and infrared camera approach to identify indocyanine green (ICG) labeled SLNs in melanoma patients, and found that the proposed procedure achieves a comparable performance to the standard technique based on the detection of a radioactive tracer [75].

2.4.6 Skin lesions

OA imaging of the skin is usually performed with systems operating on the mesoscopic scale, such as the raster-scanning optoacoustic mesoscopy system [99] or a hybrid systems combining OA with optical coherence tomography [86]. These systems utilize US detectors with higher central frequencies than OAT to achieve much higher spatial resolution at the cost of reduced depth penetration. As such, they are optimized for imaging the skin and the subcutaneous vasculature, providing unique insights into the morphology of skin lesions in a non-invasive manner. However, imaging of skin lesions has been demonstrated also with the use of OAT systems. Attia *et al.* and Breathnach *et al.* have utilized OAT to measure the thickness of non-melanoma and melanoma skin lesions, respectively [87, 88]. Using 3D handheld probes, they found a good correspondence to the lesion thickness estimated by histology, demonstrating the possibility of using OAT for excision planning. Besides imaging the skin lesions themselves, Hai *et al.* have devised an OAT setup for a label-free, non-invasive detection of circulating melanoma cells in patient veins *in vivo* [89]. While their work requires further validation, it is a promising direction towards novel optoacoustic prognostic biomarkers for melanoma patients.

2.4.7 Intraoperative guidance

Surgical instruments such as needles are typically tracked during surgery using US. However, the visibility of the instrument fades unless it is perpendicular to the US probe, limiting the maneuvering space. Since metals are strong optical absorbers, the tools can be visualized using OAT without the limitations of conventional US [150]. Navigation of a biopsy needle during sentinel LN biopsy was demonstrated by Kim *et al.* in a mouse model in 2010 [91] and in 2015 by Garcia-Urbe *et al.* also in a human patient [73]. A further improved tool tip localization precision is afforded by the use of 3D probes [151].

Besides detection of the tool using an external illumination integrated into the OAT probe, a separate line of research is concerned with the detection of the tool tip directly illuminated by an integrated optical fiber. Proposed by Piras *et al.* in 2013, the technique has been improved by Lediju Bell *et al.* [152] and advocated for endonasal [153] or spinal surgeries [154].

2.4.8 Further clinical applications

Besides the broader research directions summarized above, OAT has also been applied to other isolated clinical applications. Reber *et al.* were able to visualize the brown fat metabolism using MSOT [69]. Knieling *et al.* used MSOT to assess the intestinal inflammation in a patient with Crohn's disease [70]. Tummers *et al.* have studied pancreatic tumor detection with the aid of fluorescent tumor-targeted dyes and observed increased OA signals in the excised tumor specimen compared to the healthy pancreatic tissue [93]. Finally, OAT has also been implemented in the form of a transrectal handheld probe for prostate imaging. Horiguchi *et al.* have used the transrectal OAT to visualize the neurovascular bundle *in vivo*, which is an important landmark for nerve-preserving radical prostatectomy [94]. Later, they have also used their system to observe angiogenesis around prostate tumors *in vivo* [155]. A different transrectal probe utilizing capacitive micromachined ultrasonic transducer array was presented by Kothapalli *et al.* in 2019 [84]. This probe has shown improved image quality in terms of resolution and OA contrast. Besides

visualizing the neurovascular bundle, the authors have also obtained a specific contrast from a prostate tumor region and observed altered ICG dynamics in a malignant region of a prostate.

2.5 Physical model of OAT

2.5.1 Thermoacoustic model

Optoacoustic tomography relies on detecting pressure waves induced upon heating the tissue with pulsed laser light. The spatiotemporal thermoacoustic model of variations in temperature (T; unit K) and pressure (p; unit Pa) from their ambient values in a non-attenuating inviscid medium can be derived from linearized equations of fluid dynamics in the following form [156-158]:

$$\frac{\partial}{\partial t}\big(\rho_0(\boldsymbol{r})c_p T(\boldsymbol{r},t) - \gamma\beta T_0(\boldsymbol{r})p(\boldsymbol{r},t)\big) = \nabla\cdot\big(\kappa\nabla T(\boldsymbol{r},t)\big) + H(\boldsymbol{r},t)\ , \tag{2.5.1}$$

$$\nabla\cdot\frac{\nabla p(\boldsymbol{r},t)}{\rho_0(\boldsymbol{r})} - \gamma K_T\frac{\partial^2 p(\boldsymbol{r},t)}{\partial t^2} = -\beta\frac{\partial^2 T}{\partial t^2}\ . \tag{2.5.2}$$

Here, $\boldsymbol{r}\in\mathbb{R}^3$ is the spatial variable, t is the time variable, ρ is the mass density (kg·m^{-3}), c_p is the specific heat capacity at constant pressure (J·K^{-1}), γ is the heat capacity ratio (dimensionless), β is the thermal coefficient of volume expansions (K^{-1}), κ is the thermal conductivity (W·K^{-1}·m^{-1}), H is the heating function (J·m^{-3}·s^{-1}), and K_T is the isothermal compressibility (Pa^{-1}); subscript 0 denotes the ambient quantity.

The coupled partial differential equations (PDEs) (2.5.1) and (2.5.2) can be simplified into a single PDE under the assumption of thermal confinement. This assumption is satisfied when the pulse duration τ_p is much shorter than thermal relaxation time τ_{th}:

Assumption 1: *Thermal confinement.*

$$\tau_p \ll \tau_{th}. \tag{2.5.3}$$

The thermal relaxation time is defined as

$$\tau_{th} = \frac{d_c^2}{\alpha_{th}}\ , \tag{2.5.4}$$

where d_c is the characteristic dimension[4] of the excited structures inside the region of interest (m), and α_{th} is the thermal diffusivity (m^2·s^{-1}). The thermal relaxation time specifies the period needed for the deposited heat to affect the surrounding structures by thermal diffusion. Assuming reasonable values for OAT in biological tissue of $d_c = 0.1$ mm and $\alpha_{th} = 0.1$ mm^2/s, thermal confinement is satisfied when $\tau_p \ll 1$ µs. Under the thermal confinement assumption, the thermal conductivity in equation (2.5.1) is set to zero, $\kappa = 0$, and the two equations can be reduced to the following PDE:

$$\frac{1}{c(\boldsymbol{r})^2}\frac{\partial^2}{\partial t^2}p(\boldsymbol{r},t) - \rho_0(\boldsymbol{r})\nabla\cdot\left(\frac{\nabla p(\boldsymbol{r},t)}{\rho_0(\boldsymbol{r})}\right) = \frac{\beta}{c_p}\frac{\partial}{\partial t}H(\boldsymbol{r},t)\ . \tag{2.5.5}$$

[4] "The dimension of the structure of interest or the decay constant of the optical energy deposition, whichever is smaller." [125]

Here, c denotes the speed-of-sound ($\mathrm{m \cdot s^{-1}}$), which is related to the previously introduced quantities as:

$$c = \sqrt{\frac{1}{\rho K_T}} \,. \tag{2.5.6}$$

Since the spatial distribution of the mass density ρ (and of the speed-of-sound c) in the imaged tissue is typically unknown, acoustically homogenous medium is often assumed and the equation $(2.5.5)$ is further simplified to:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta\right) p(\boldsymbol{r}, t) = \frac{\beta}{c_p}\frac{\partial H(\boldsymbol{r}, t)}{\partial t} \,. \tag{2.5.7}$$

Here, Δ denotes the Laplace operator in the spatial dimensions. This equation is often also written in an equivalent form containing the Grüneisen parameter Γ (dimensionless), which summarizes the efficiency of heat conversion to acoustic pressure:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\Delta\right) p(\boldsymbol{r}, t) = \Gamma\frac{\partial H(\boldsymbol{r}, t)}{\partial t} \,, \tag{2.5.8}$$

$$\Gamma \equiv \frac{\beta c^2}{c_p} \,. \tag{2.5.9}$$

The left part of the equation (2.5.8) describes the pressure wave propagation whereas the right part represents the pressure source. The source term contains a time derivative of the heating function, implying that only time-variant heating produces pressure waves [125].

2.5.2 Optoacoustic wave equation

The time-variant heating in optoacoustic tomography is generated by illuminating the tissue with pulsed laser light. The heating function has the following form:

$$H(\boldsymbol{r}, t) = \mu_A(\boldsymbol{r})\Phi(\boldsymbol{r}, t)\eta_{th} \,. \tag{2.5.10}$$

Here, μ_A is the optical absorption coefficient ($\mathrm{m^{-1}}$), Φ is the optical fluence rate ($\mathrm{J \cdot m^{-2} \cdot s^{-1}}$), and η_{th} is the heat conversion efficiency (dimensionless). In practice, the fluence rate can be decomposed into independent spatial and temporal components:

$$\Phi(\boldsymbol{r}, t) = \Phi_r(\boldsymbol{r})\Phi_t(t) \,. \tag{2.5.11}$$

If the length of the illumination pulse τ_p is much shorter than the stress relaxation time τ_{st}, the heating is occurring under the stress confinement. The stress relaxation time characterizes the time needed for the induced pressure to affect the surrounding structures and is defined as:

$$\tau_{st} = \frac{d_c}{c} \,. \tag{2.5.12}$$

This assumption means the entire thermal energy must be deposited before the density or volume of the illuminated medium can change [159]. Formally, the stress confinement assumption is defined as:

Assumption 2: *Stress confinement.*

$$\tau_p \ll \tau_{st} \,. \tag{2.5.13}$$

Assuming values of $d_c = 0.1$ mm and $c = 1500$ m/s, stress confinement is satisfied when $t_p \ll 67$ ns. The length of laser pulses used in OAT is on the order of 10 ns and generally satisfies the stress confinement assumption (as well as the thermal confinement assumption). When the stress confinement assumption is satisfied, the temporal fluence function Φ_t can be approximated by a Dirac delta function, $\Phi_t(t) \cong \delta(t)$.

The standard **optoacoustic wave equation** then takes the following form:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\Delta\right) p(\boldsymbol{r}, t) = p_0(\boldsymbol{r}) \frac{\mathrm{d}\delta(t)}{\mathrm{d}t} \ . \tag{2.5.14}$$

The initial pressure p_0 can further be expressed as

$$p_0(\boldsymbol{r}) \equiv p(\boldsymbol{r}, t = 0) = \Gamma \eta_{th} \mu_A(\boldsymbol{r}) \Phi(\boldsymbol{r}) \ , \tag{2.5.15}$$

and the optoacoustic wave generation and propagation can be modeled as the following system of a PDE and initial conditions, known as the Cauchy problem for the wave equation:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\Delta\right) p(\boldsymbol{r}, t) = 0 \tag{2.5.16a}$$

$$p(\boldsymbol{r}, t = 0) = \Gamma \eta_{th} \mu_A(\boldsymbol{r}) \Phi(\boldsymbol{r}) \ , \tag{2.5.16b}$$

$$\frac{\partial}{\partial t} p(\boldsymbol{r}, t)\Big|_{t=0} = 0 \tag{2.5.16c}$$

2.5.3 Acoustic forward problem

Solving the equation system (2.5.16) yields a unique solution for any $p(\boldsymbol{r}, t)$. In optoacoustic tomography, one assumes acoustic point-detectors placed on a surface Ω enclosing the imaged object. The forward acoustic problem means finding $p(\boldsymbol{r}_d, t)$ for $\boldsymbol{r}_d \in \Omega$, which corresponds to the acoustic signals arriving at the detectors. The solution to this problem is available through the Green's function approach. Green's function (G) represents the response of a differential system to a spatiotemporal delta impulse source located at $(\boldsymbol{r}', t')$:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta\right) G(\boldsymbol{r}, t, \boldsymbol{r}', t') = -\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t') \ . \tag{2.5.17}$$

In the free-space (an infinite space with no boundary), the Green's function of this differential system has the form of a diverging spherical wave [125]:

$$G(\boldsymbol{r}, t, \boldsymbol{r}', t') = \frac{1}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} \delta\left(t - t' - \frac{|\boldsymbol{r} - \boldsymbol{r}'|}{c_s}\right) \ . \tag{2.5.18}$$

The solution of (2.5.16) then becomes a superposition of such spherical waves in space and time:

$$p(\boldsymbol{r}, t) = \int_{-\infty}^{t^+} \int G(\boldsymbol{r}, t, \boldsymbol{r}', t') \frac{\Gamma}{c^2} \frac{\partial}{\partial t'} H(\boldsymbol{r}', t') \ \mathrm{d}\boldsymbol{r}' \, \mathrm{d}t' \ . \tag{2.5.19}$$

Substituting (2.5.18) into (2.5.19) and using the sifting property of the Dirac delta yields

$$p(\boldsymbol{r}, t) = \frac{\Gamma}{4\pi c_s^2} \int \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \frac{\partial}{\partial t} H\left(\boldsymbol{r}', t - \frac{|\boldsymbol{r} - \boldsymbol{r}'|}{c_s}\right) \ \mathrm{d}\boldsymbol{r}' \ . \tag{2.5.20}$$

Replacing the heating function with its optoacoustic form yields the following solution:

$$p(\boldsymbol{r},t) = \frac{\Gamma\eta_{th}}{4\pi c_s^2}\int \frac{\mu_A(\boldsymbol{r}')\Phi(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\frac{\mathrm{d}}{\mathrm{d}t}\delta\left(t-\frac{|\boldsymbol{r}-\boldsymbol{r}'|}{c_s}\right)\,\mathrm{d}\boldsymbol{r}'\ , \tag{2.5.21}$$

which can be equivalently written in any of the two forms:

$$p(\boldsymbol{r},t) = \frac{\Gamma\eta_{th}}{4\pi c_s^2}\,\frac{\partial}{\partial t}\int_{|\boldsymbol{r}-\boldsymbol{r}'|=c_s t} \frac{\mu_A(\boldsymbol{r}')\Phi(\boldsymbol{r}')}{c_s t}\,\mathrm{d}\boldsymbol{r}'\ , \tag{2.5.22}$$

$$p(\boldsymbol{r},t) = \frac{1}{4\pi c_s^2}\,\frac{\partial}{\partial t}\int_{|\boldsymbol{r}-\boldsymbol{r}'|=c_s t} \frac{p_0(\boldsymbol{r}')}{c_s t}\,\mathrm{d}\boldsymbol{r}'\ . \tag{2.5.23}$$

Intuitively, the equations (2.5.21)–(2.5.23) can be understood as follows: The pressure at a point $\boldsymbol{r}$ at time t is equal to the mean pressure emitted by point pulse-sources located on the sphere of radius $c_s t$ centered at $\boldsymbol{r}$. This represents the spherical mean operator [160]. By integrating both sides of the solution with respect to time, one obtains the spherical Radon transform—related to the standard Radon transform with the difference of applying the spatial integral to spheres rather than planes [159, 161]. In practice, the pressure field needs to be evaluated at multiple locations and computationally more efficient *k-wave* methods based on Fourier transforms are used [162, 163].

2.5.4 Limitations of the model

The above-described physical model of optoacoustic tomography is reasonably accurate for practical purposes while remaining computationally tractable thanks to several simplifying assumptions. However, it is important to understand the nature and implications of those assumptions as they may cause less or more severe errors.

First, as introduced in equation (2.5.7), acoustically homogenous medium is assumed: $\rho(\boldsymbol{r}) = \rho$, implying $c(\boldsymbol{r}) = c$. A justification for this assumption is that the optical contrast in soft biological tissue is much higher than the density variations, allowing to neglect them [164]. However, variations in the acoustic impedance have practical implications for optoacoustic tomography. First, errors in the speed of acoustic waves propagation cause spatial distortion of the imaged objects. This issue is further pronounced in presence of acoustic coupling medium with a distinct speed of sound. Second, acoustic interfaces alter the direction of the wave propagation through refraction and reflection. Whereas the refraction is manifested through further spatial distortion of the obtained images, reflections on acoustic interfaces cause imaging artifacts—fictious structures appearing in the images. Whereas the image distortion due to acoustic inhomogeneities is usually non-detrimental, reflection artifacts are problematic as they can lead to erroneous evaluation of the images.

The second simplifying assumption in the model is that of an acoustically non-attenuating medium. In practice, the acoustic attenuation in soft tissue is frequency dependent according to the power law [165]. Although at frequencies typically used in OAT, the acoustic attenuation is much weaker than the optical attenuation and is not the main factor limiting the penetration depth [125], at frequencies around 43 MHz the optical and acoustic attenuation become comparable [166]. Depending on the signal frequency, the acoustic attenuation may cause a reduction of resolution and hinder quantification attempts [166]. Modelling the acoustic attenuation requires use of the Stoke's equation in place of eq. (2.5.14) [167]:

$$\left(\frac{\partial^2}{\partial t^2} - c^2\Delta - \tau c^2\Delta\frac{\partial}{\partial t}\right)p(\boldsymbol{r},t) = p_0(\boldsymbol{r})\frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\ , \tag{2.5.24}$$

Here, the coefficient τ captures the relation between the medium viscosity and density. Various methods have been proposed for solving this variant of the wave equation and correction of the sound attenuation effects [166-169]. It is nevertheless important to recognize that the frequency response of the ultrasound detector and signal filters needs to be also considered while modeling the acoustic attenuation.

2.6 Image formation

Optoacoustic tomography forms images by reconstructing the physical properties of the tissue upon the laser light illumination from acoustic signals recorded by ultrasound detectors placed on a detection surface Ω_D which is fully or partially enclosing the imaged object. The reconstruction is performed by inverting the forward model described in the previous section. The physical quantity of interest in current optoacoustic tomography is the initial pressure distribution, p_0, which can be obtained by solving the acoustic inverse problem. However, the ultimate (and much more ambitious) goal of OAT is recovering the absorption distribution μ_A from the initial pressure and, if images at multiple wavelengths are available, the closely related tissue chromophore concentrations. This is often called *quantitative* or *functional* OAT and requires solving the optical inverse problem. Whereas a multitude of solutions to the acoustic inverse problem exists, the optical inverse problem is ill-posed and remains unsolved in the general case [170].

2.6.1 Acoustic inverse problem

Many solutions to the acoustic problem exist, depending on the imaging system properties. When considering the inversion algorithm, the most important system property is the geometrical configuration of the detector elements—the detection surface. Detector elements are either placed on the detection surface and record in parallel, or a single detection element may be scanned along the surface. The most common detection surfaces are sphere, cylinder, and plane. In the case of handheld OAT devices, the detectors are typically placed on a line or an arc and the obtained images are 2D, although hemispherical handheld detectors for 3D acquisition have also been demonstrated.

The algorithms can be roughly divided into four categories: closed-form time-domain and frequency-domain solutions, and numerical time-reversal and model-based methods [162]. Here, the time-domain back-projection and a model-based approach are viewed in more detail.

Time-domain methods
The time-domain (filtered) back-projection algorithm works in a similar manner as the inverse Radon transfer, projecting the recorded data to spheres instead of planes (alluding to the similarity of the forward problem to the Radon transform, mentioned above). Generally, the algorithm can be summarized in three steps: 1) filtering of the recorded data, 2) projection of the filtered data on spheres according to the time-of-flight, 3) (weighted) summation of the projections. Formally:

$$p_0(\boldsymbol{r}) = \int_{\Omega} b\left(\boldsymbol{r}_S, \frac{|\boldsymbol{r}-\boldsymbol{r}_S|}{c}\right) \frac{\mathrm{d}\Omega(\boldsymbol{r},\boldsymbol{r}_S)}{\Omega_S(\boldsymbol{r})} \tag{2.6.1}$$

Here, Ω denotes the detection surface, $\Omega_S(\boldsymbol{r})$ is the solid angle of the whole detection surface w.r.t. reconstruction point $\boldsymbol{r}$ (2π for planar geometry, 4π for spherical), and $\mathrm{d}\Omega(\boldsymbol{r},\boldsymbol{r}_S)$ denotes the solid angle subtended by an infinitesimal surface element located at $\boldsymbol{r}_S$ when viewed from $\boldsymbol{r}$. Analytically, the solid angle can be computed as

$$\mathrm{d}\Omega(\boldsymbol{r}_S) = \frac{\mathrm{d}S(\boldsymbol{r}_S)}{|\boldsymbol{r} - \boldsymbol{r}_S|^2} \left(\boldsymbol{n}_\Omega(\boldsymbol{r}_S) \cdot \frac{\boldsymbol{r} - \boldsymbol{r}_S}{|\boldsymbol{r} - \boldsymbol{r}_S|} \right) , \tag{2.6.2}$$

where, $\boldsymbol{n}_\Omega(\boldsymbol{r}_s)$ denotes the inward-pointing normal vector to the surface Ω at point $\boldsymbol{r}_S$, and $\mathrm{d}S(\boldsymbol{r}_S)$ represents the infinitesimal surface element of surface Ω at point $\boldsymbol{r}_S$.

Furthermore, $b(\boldsymbol{r}, t)$ denotes the filtered back-projection function. Depending on its form, different time-domain algorithms can be derived. The simplest one, the *delay-and-sum* method, uses the raw signals without any filtering [171]:

$$b_{DS}(\boldsymbol{r}, t) \equiv p(\boldsymbol{r}, t) \ . \tag{2.6.3}$$

The *universal back-projection* algorithm uses the following form [172]:

$$b_{UBP}(\boldsymbol{r}, t) \equiv 2p(\boldsymbol{r}, t) - t \tfrac{\partial}{\partial t} p(\boldsymbol{r}, t) \ . \tag{2.6.4}$$

The second term with a time derivative represents a ramp filter suppressing low frequencies [125]. This form of back-projection is an exact solution to the inverse acoustic problem outlined in the section 2.5.3 for spherical, cylindrical, and planar detection surfaces (hence *universal*). Computational complexity of these algorithms for N_D detectors and N_V voxels is $\mathcal{O}(N_D N_V) \approx \mathcal{O}(N^5)$, assuming a voxel grid $N \times N \times N$ and N^2 detection elements [173]. Part of the computational load can be eliminated under the *far-field approximation* [173], applicable when the distance of the object from the detectors is significantly larger than the size of the features of interest, or $|\boldsymbol{r} - \boldsymbol{r}_s| \cong |\boldsymbol{r}_s|$. In such case, the solid angle $\mathrm{d}\Omega(\boldsymbol{r}, \boldsymbol{r}_S)$ no longer depends on $\boldsymbol{r}$ and does not need to be computed for each voxel separately. Furthermore, the second term in equation (2.6.14) outweighs the first one, $t \frac{\partial}{\partial t} p(\boldsymbol{r}, t) \gg p(\boldsymbol{r}, t)$ and the back-projection function becomes

$$b_{FFBP}(\boldsymbol{r}, t) \equiv t \tfrac{\partial}{\partial t} p(\boldsymbol{r}, t) \ . \tag{2.6.5}$$

Whereas the overall computation complexity of back-projection in 3D remains $\mathcal{O}(N^5)$ even under the far-field approximation, the algorithm can be parallelized using a graphics processing unit (GPU) and run in real-time [174]. Furthermore, in 2D the complexity is only $\mathcal{O}(N^3)$. Commercially available handheld OAT scanners utilize back-projection for real-time image reconstruction.

Model-based methods

Model-based methods treat the forward acoustic problem as a discrete linear system [164], where the initial pressure vector $\boldsymbol{p}_0$ and the vector of the detected pressure $\boldsymbol{p}_d$ are related through a system (model) matrix $\boldsymbol{M}$:

$$\boldsymbol{p}_d = \boldsymbol{M}\boldsymbol{p}_0 \ . \tag{2.6.6}$$

Discretization is performed as a grid containing N_V pixels and T time-samples. The number of detectors, N_D, corresponds to the actual number of transducers. Then, $\boldsymbol{p}_d$ is a vector of length $N_D T$, $\boldsymbol{p}_0$ is a vector of length N_V, and $\boldsymbol{M}$ is a matrix of shape $N_D T \times N_V$.

For the simplified case of perfect point-sources and point-detectors, the values of the matrix $\boldsymbol{M}$ are solutions to the equations (2.5.21)–(2.5.23). In practice, however, the sources are modelled as uniform spherical absorbers of diameter d matching the discretization grid step size, generating a typical N-shaped pressure wave [109]:

$$p(\boldsymbol{r}, t) = \begin{cases} \dfrac{p_0}{2|\boldsymbol{r} - \boldsymbol{r}'|} \left(\dfrac{|\boldsymbol{r} - \boldsymbol{r}'|}{c_s} - t \right) & \text{if } |\boldsymbol{r} - \boldsymbol{r}'| - \dfrac{d}{2} \leq c_s t \leq |\boldsymbol{r} - \boldsymbol{r}'| + \dfrac{d}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6.7}$$

Equation (2.6.7) describes the propagation of the N-shape pressure wave with the initial amplitude p_0 emanating from a spherical absorber at point $\boldsymbol{r}'$ in a medium with a speed of sound c_s.

Furthermore, the detectors have a finite size and their response p_d needs to be modelled accordingly. A simple model is averaging over the detector surface:

$$p_d(t) = \int\limits_{\boldsymbol{r}\in\Omega_d} p(\boldsymbol{r}, t) \ \mathrm{d}\boldsymbol{r} \ , \tag{2.6.8}$$

where Ω_d denotes the surface of the detector. A computationally affordable approximation to the averaging model is a temporal convolution of the pressure pulse with a spatial impulse response function of the detector [175, 176].

Solving the linear system for a given $\boldsymbol{p}_d$ (2.6.6) directly is impractical due to a potentially very large model matrix and iterative methods are preferred. The solution $\boldsymbol{p}_0^*$ can be found by minimization of the following least-squares objective:

$$\boldsymbol{p}_0^* = \underset{\boldsymbol{p}_0}{\operatorname{argmin}} \|\boldsymbol{M}\boldsymbol{p}_0 - \boldsymbol{p}_d\|_2^2 \ . \tag{2.6.9}$$

In practice, the solution may not be uniquely and accurately determined by the recorded data and the model. In that case, the problem is ill-posed and additional regularization is required [175]. Tikhonov regularization is a typical choice [177, 178]:

$$\boldsymbol{p}_0^* = \underset{\boldsymbol{p}_0}{\operatorname{argmin}} \|\boldsymbol{M}\boldsymbol{p}_0 - \boldsymbol{p}_d\|_2^2 + \lambda\|\boldsymbol{R}\boldsymbol{p}_0\|_2^2 \ . \tag{2.6.10}$$

Here, λ is a parameter controlling the regularization strength and $\boldsymbol{R}$ is a regularization matrix which can encode additional prior information about the solution. When set to identity, $\boldsymbol{R} = \boldsymbol{I}$, equation (2.6.10) becomes simple L_2 regularization. A proper value of the λ parameter can be identified using the L-curve approach [179].

Other regularization terms have been also utilized, such as L_1 norm [180], total variation [180, 181], or shearlet regularization [182, 183]. Additionally, it is reasonable to constrain the solution space to positive values, $\boldsymbol{p}_0 \in \mathbb{R}_+^{N_V}$, as negative initial pressure violates the physical model [184]. Depending on the exact form of the optimization problem, a suitable least-squares solver needs to be applied [185].

Model-based image reconstruction methods are computationally much more demanding than analytical solutions and typically are too slow to be used for real-time image formation. However, the increase in computing power and availability of parallel processors such as GPUs in the recent years fuel the interest in model-based methods. In many aspects, they have major advantages over back-projection schemes. They afford integration of physical properties of the imaging system into the forward model, such as its detector geometry, detector shape [175, 186], and electrical response [181, 186]. Furthermore, the model can also account for the light [187] and sound [168, 188] attenuation in the medium. Overall, model-based reconstruction leads to images with less artifacts and considerably better image quality than back-projection [177, 181].

2.6.2 Optical inverse problem

The optical inverse problem refers to recovering the distribution of the absorption coefficient inside the imaged tissue, $\mu_A(\boldsymbol{r})$, given the initial pressure distribution p_0. Solving the optical inverse problem is required to ultimately obtain the distribution of concentration coefficients of the

individual chromophores. Unlike the acoustic inverse problem, the optical inverse problem is ill-posed and remains unsolved in the general case [170].

The relationship between the acoustic pressure and the optical absorption is modelled by the equation (2.5.15) as $p_0(\boldsymbol{r}) = \Gamma\eta_{th}\mu_A(\boldsymbol{r})\Phi(\boldsymbol{r})$. The light fluence Φ at a point $\boldsymbol{r}$ depends on the distribution of the absorption and scattering (μ_S, unit m^{-1}) coefficients within the tissue, which can be explicitly written as (omitting the energy conversion efficiency parameters):

$$p_0(\boldsymbol{r}) = \mu_A(\boldsymbol{r})\Phi(\boldsymbol{r}; \mu_A, \mu_S). \tag{2.6.11}$$

The difficulty of recovering μ_A from p_0 has multiple roots. First and foremost, the light fluence Φ depends on both μ_A and μ_S, which are unknown. Estimating these coefficients together is an ill-posed problem, since the same fluence profile can be generated by many different configurations of μ_A and μ_S. Hence, given p_0, μ_A does not have a unique solution [189]. Second, the scattering coefficient affects the solution only through the fluence term, which smoothens out any high-frequency noise component of μ_S. This leads to numerical instabilities in the inversion methods, resulting in amplifying the high-frequency component of μ_S [189]. Finally, the Grüneisen parameter $\Gamma(\boldsymbol{r})$ is also unknown and may vary throughout the tissue, although we refrain from discussing the issue here.

Light fluence model

The light fluence distribution is required for solving the optical inverse problem. Since the fluence inside the tissue is unknown, it needs to be modelled. A typical particle-based model of light transport is the radiative transfer equation. Because the illumination time-scale in OA imaging is much shorter than the acoustic propagation (stress confinement), the time-independent radiative transfer equation can be used [189]:

$$(\hat{\boldsymbol{s}} \cdot \nabla + \mu_A + \mu_S\ )\phi(\boldsymbol{r}, \boldsymbol{s}) - \mu_S \int \Theta(\hat{\boldsymbol{s}}, \hat{\boldsymbol{s}}')\phi(\boldsymbol{r}, \hat{\boldsymbol{s}}')d\hat{\boldsymbol{s}}' = q(\boldsymbol{r}, \hat{\boldsymbol{s}})\ , \tag{2.6.12}$$

where $\Theta(\hat{\boldsymbol{s}}, \hat{\boldsymbol{s}}')$ is the scattering phase function (dimensionless), giving the probability that a photon travelling along the angle $\hat{\boldsymbol{s}}$ will be deflected to the direction of $\hat{\boldsymbol{s}}'$; $q(\boldsymbol{r}, \hat{\boldsymbol{s}})$ is the light source, and $\phi(\boldsymbol{x}, \hat{\boldsymbol{s}})$ represent the time-integrated light radiance (unit $\mathrm{J{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}s^{-1}}$). Its integral over all angles $\hat{\boldsymbol{s}}$ corresponds to the light fluence:

$$\Phi(\boldsymbol{r}) = \int \phi(\boldsymbol{r}, \hat{\boldsymbol{s}})\ \mathrm{d}\hat{\boldsymbol{s}}. \tag{2.6.13}$$

As solving the radiative transfer equation is usually infeasible, its approximations are used. In particular, the diffusion approximation is considered for biological tissue, where the scattering dominates the photon transport: $\mu_S \gg \mu_A$. The diffusion approximation has the following form:

$$\mu_A\Phi - \nabla \cdot (D\nabla)\Phi = q_0. \tag{2.6.14}$$

Here, $D = [3(\mu_A + \mu_S')]^{-1}$ is the optical diffusion coefficient (unit m), $\mu_S' = (1-g)\mu_S$ is the reduced scattering coefficient with anisotropy factor g, and q_0 is an isotropic light source term. This approximation is accurate at depths where all the photons have lost their original directionality (> 1 mm), which is given by the mean transport path. The light fluence typically reaches its maximum just below the tissue surface where both incident and backscattered photons contribute.

Inversion

Solving the optical inverse problem requires constraining the solution space by incorporating suitable inductive biases based on prior knowledge. Numerous methods to this end have been proposed and the reader is referred to the section 5.2 for a brief overview; a reader with a substantial

interest in the topic can then further benefit from the review by Cox *et al.* [189]. Here, one particular approach is discussed, which makes use of illumination with multiple wavelengths [170].

The main assumption is that the absorption spectrum at any imaged point should be an additive combination of absorption spectra of a limited amount (K) of known optical absorbers present in the imaged tissue:

$$\mu_A(\boldsymbol{r}, \lambda) = \sum_{k=1}^{K} c_k(\boldsymbol{r}) \alpha_k(\lambda). \tag{2.6.15}$$

Here, λ represents the illumination wavelength, c_k represents the (non-negative) concentration of a chromophore k and α_k is its absorption spectrum. This assumption is valid in the near-infrared wavelength range where the absorption is dominated by hemoglobin, lipids, and water [190].

The second assumption is that the reduced scattering coefficient in tissue depends on the wavelength via the following relationship:

$$\mu_S'(\boldsymbol{r}, \lambda) = a(\boldsymbol{r}) \lambda^{-b}. \tag{2.6.16}$$

This assumption has been derived from experimental results and the values of b are known; a represents an unknown scattering amplitude which is wavelength-independent.

Using these two assumptions, the initial pressure is modelled as:

$$p_m(\boldsymbol{r}, \lambda; \boldsymbol{c}, a) = \mu_A(\boldsymbol{r}, \lambda; \boldsymbol{c}) \Phi(\boldsymbol{r}, \lambda; \boldsymbol{c}, a). \tag{2.6.17}$$

Here, $\boldsymbol{c}$ represents the vector of coefficients c_1, …, c_K. The model utilizes equations (2.6.16) and (2.6.17) to compute μ_A and μ_S' from $\boldsymbol{c}$ and a, and the diffusion approximation from eq. (2.6.14) is used to compute the fluence Φ.

Least-squares minimization is used to solve the inverse problem:

$$\boldsymbol{c}^*, a^* = \underset{\boldsymbol{c}, a}{\operatorname{argmin}} \frac{1}{2} \| \boldsymbol{p}_m(\boldsymbol{c}, a) - \boldsymbol{p}_0 \|_2^2 + f^{reg}(\boldsymbol{c}, a) \ , \tag{2.6.18}$$

where $\boldsymbol{p}_0$ is a vector of initial pressure measurements for all pixels $\boldsymbol{r}$ and wavelengths λ, and $\boldsymbol{p}_m$ is a vector-valued function corresponding to the model evaluation at the matching values of $\boldsymbol{r}$ and λ. Regularization function f^{reg} is used to include additional assumptions on the distribution of coefficients. The original paper utilizes Tikhonov regularization to avoid noise amplification and improve the stability of the inversion.

Overall, this method uses only very mild assumptions to solve the problem with robust mathematical tools. Despite the undisputable elegance of the presented framework, its demonstration has been limited to numerical simulations. The practical issues of imperfect acoustic inversion and unknown Grüneisen parameter distribution prevent its straightforward translation to real data.

2.7 Image analysis

The aim of medical image analysis is the extraction of measurements to provide a clinician with data needed to make decisions about the patient diagnosis and treatment. In multispectral OAT, the measurements of interest are related to concentrations of chromophores in the examined regions. The main chromophores of interest are oxygenated and deoxygenated hemoglobin, but in certain

cases lipids, water, and exogenous contrast agents are also targeted. To obtain the chromophore concentrations, the measured spectra need to be disentangled into the individual components according to some unmixing model. In this section, the typically used unmixing models are discussed.

Linear model
The linear model is a standard spectroscopic approach for disentanglement of spectrum constituents. It assumes that the absorption spectrum in each pixel is a linear combination of absorption spectra of a-priori known constituents [191]. The mathematical form of this mixing model was already introduced in the previous section in the equation (2.6.15). Inverting this model is known as **linear unmixing**:

$$\underset{\boldsymbol{c}}{\operatorname{argmin}} \|\boldsymbol{A}\boldsymbol{c} - \boldsymbol{s}\|_2. \tag{2.7.1}$$

Here, $\boldsymbol{A}$ is a matrix with the absorption spectra α_k of the constituent chromophores stacked as columns:

$$\boldsymbol{A} = \begin{bmatrix} \alpha_1(\lambda_1) & \cdots & \alpha_K(\lambda_1) \\ \vdots & \ddots & \\ \alpha_1(\lambda_L) & & \alpha_K(\lambda_L) \end{bmatrix} \tag{2.7.2}$$

The vector $\boldsymbol{c}$ contains the concentrations of the K chromophores and the vector $\boldsymbol{s}$ is the measured spectrum at L wavelengths λ_1, λ_2, …, λ_L.

Linear unmixing requires the constituent spectra to be linearly independent (i.e., the matrix $\boldsymbol{A}$ has a full rank) and the number of wavelengths must be equal or higher than the number of unmixing components ($L \geq K$), otherwise the problem is underdetermined and an infinite number of solutions exists. Additionally, the non-negativity constraint on the elements of $\boldsymbol{c}$ is often used to ensure physically meaningful solutions. In that case, a non-negative least squares solver is employed to find the optimal solution [185, 191].

The main limitation of linear unmixing is that the measured spectrum $\boldsymbol{s}$ is assumed to be a vector of optical absorption coefficients $\mu_A(\lambda)$, but in practice it is a vector of reconstructed pressure signals $p_0(\lambda)$. As explained in the previous section, the absorption coefficients are related to the initial pressure through the unknown light fluence and the photoacoustic conversion efficiency coefficient (see eq. (2.5.15)). Furthermore, image reconstruction artifacts may create additional spectral variations violating the linear model. For all these reasons, linear unmixing produces satisfactory results only in simplified computational simulations and in idealized phantoms but fails at quantitative analysis of *in vivo* images.

Eigenspectral model
To overcome the limitations of the linear model, the eigenspectra optoacoustic tomography method (eMSOT) has been proposed by Tzoumas *et al.* [192], which accounts for the unknown light fluence and hence provides more precise estimates of oxygen saturation. The basic assumption of eMSOT is that the light fluence distribution is smooth and thus low-dimensional. Using computer simulations of tissue at various levels of oxygen saturation, the authors generated a dataset of light fluence maps. This dataset was analyzed via principal component analysis and the authors have found that the fluence spectrum at any point can be reliably modelled using only the first three principal components (which they call eigenspectra):

$$\Phi(\boldsymbol{r}, \lambda) = \Phi_{\mathrm{M}}(\lambda) + m_1(\boldsymbol{r})\Phi_1(\lambda) + m_2(\boldsymbol{r})\Phi_2(\lambda) + m_3(\boldsymbol{r})\Phi_3(\lambda). \tag{2.7.3}$$

Here, Φ_M is the mean spectrum of the simulated dataset, Φ_1–Φ_3 are the three eigenspectra, and m_1–m_3 are their coefficients. Estimation of the eigenspectra coefficients is performed in a two-

step process for a regular grid of points. First, initial estimates are found for each grid point individually using oxygen saturation values obtained by the linear model. Then, the estimates for the grid points are jointly optimized in a constrained model-based inversion scheme. The model itself assumes that oxy- and deoxyhemoglobin are the only two absorbers in the tissue and the fluence can be modelled according to the equation (2.7.3):

$$p_0(\boldsymbol{r}, \lambda; \boldsymbol{c}, \boldsymbol{m}) = \boldsymbol{\alpha}^T(\lambda)\boldsymbol{c}(\boldsymbol{r})\Big(\Phi_M(\lambda) + \boldsymbol{m}^T(\boldsymbol{r})\boldsymbol{\Phi}_e(\lambda)\Big). \tag{2.7.4}$$

Here, $\boldsymbol{\alpha}$ is a vector of absorption coefficients of oxy- and deoxyhemoglobin, $\boldsymbol{c}$ is the vector of their concentrations, $\boldsymbol{m}$ is the vector of the eigenspectra coefficients and $\boldsymbol{\Phi}_e$ is the vector of the three eigenspectra. The model contains a normalization scaling factor on the concentration coefficients:

$$\hat{c}_{\mathrm{Hb}}(\boldsymbol{r}) = c_{\mathrm{Hb}}(\boldsymbol{r})\frac{\Gamma(\boldsymbol{r})\|p_0(\boldsymbol{r})\|_2}{\|\Phi(\boldsymbol{r})\|_2}. \tag{2.7.5}$$

This scaling does not affect the oxygen saturation computation. The optimization constraints are of three types. First, spatial variations of coefficients m_1 and m_3 are penalized according to the distance of the points on the grid. Second, the coefficient m_2 is constrained to become smaller for grid points deeper in the tissue. Finally, the coefficients m_1 and m_3 are required to be within a value range depending on the initial estimates. Besides these constraints on $\boldsymbol{m}$, the concentrations of hemoglobin must be non-negative.

After estimation of the fluence parameters on the grid, fluence can be computed for all points by interpolation of the fitted eigenspectra parameters, and oxygen saturation can be obtained with much higher precision than what is afforded by the simple linear model. eMSOT has shown good results in mouse imaging [192], but has not been translated into the clinical imaging practice. Further improvements to the optimization scheme have been published, including Bayesian optimization [193] and using neural networks to estimate the parameters [194].

Blind unmixing

Blind unmixing[5] (BU) is a category of algorithms that do not rely on prior knowledge of the constituent chromophores. Instead of fitting the observations to a predefined absorption matrix $\boldsymbol{A}$, the matrix entries are estimated during the unmixing procedure. The unmixing is applied concurrently on a whole set of measurements—all pixels in one image, or a whole imaging study with multiple subjects—to ensure that the estimate of the matrix $\boldsymbol{A}$ represents the whole dataset.

The advantage of BU is that it does not attempt to overcome the notoriously hard problem of light fluence correction (like eMSOT) nor pretends the problem is not there (like the linear model). Instead, BU gives an answer to a different question:

Which dominant spectral patterns are observed in the measured data?

It is quite reasonable to assume that these dominant patterns are linked to the actual imaged chromophores. Note that this question drops altogether the impractical requirement of measuring the optical absorption coefficients—and the associated modelling of complicated physical, electronical, and computational processes between the illuminated chromophores and the reconstructed images—we can just as well analyze the spectral patterns in the initial pressure data. In this way, the question addressed by BU is a much easier one. Still, a correct answer to this easier question can give a better insight into the chromophore distribution than a wrong result produced by linear unmixing.

[5] Also called *blind source separation* in other contexts.

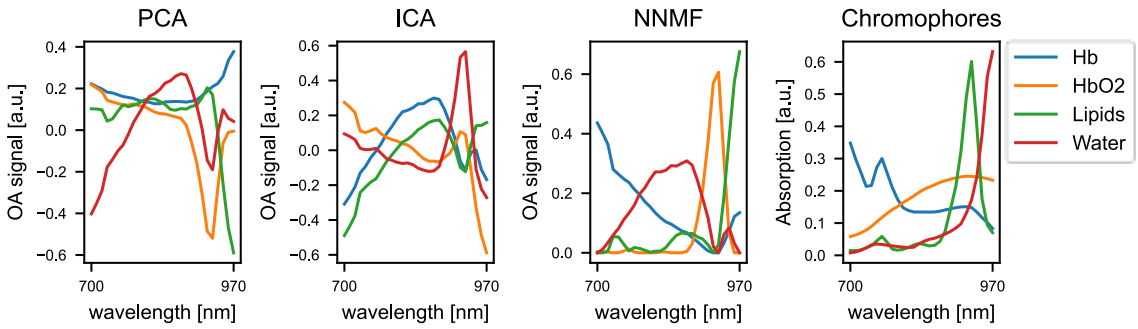

Figure 2.5: Four spectral components detected by different blind unmixing algorithms (PCA, ICA, NNMF) in a scan of a breast tumor *in vivo*. Absorption spectra of Hb, HbO_2, lipids, and water are shown for reference in the rightmost plot. Abbreviations: OA, optoacoustic; a.u., arbitrary unit; PCA, principal component analysis; ICA, independent component analysis; NNMF, non-negative matrix factorization.

On the other hand, since the spectral components identified in a BU procedure do not have a specific biological meaning assigned *per se*, the analysis of such unmixed data is inherently more challenging. An extensive experience is required to interpret the meaning of the unmixed components, as the observed patterns are outcomes of biochemical processes (which chromophores are present together), physical processes (light fluence, acoustic reflections), imaging system (artifacts), and the unmixing algorithm (which components it picks up). Accordingly, there is not a single correct solution to the BU problem and often a different algorithm or parameter setup needs to be used to arrive at a useful result. Furthermore, as the results of BU methods are usually "married" to a specific dataset, they cannot be easily transferred to new data or generalized to different studies.

Although BU algorithms are free of assumptions about the optical properties of the constituent chromophores, they rely on other assumptions that allow them to disentangle the measured spectra into individual components. Here, we discuss the assumptions behind three BU algorithms that have been applied to multispectral OAT: principal component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NNMF). To demonstrate the differences of the components they identify in the data, Fig. 2.5 shows the results of BU of an OAT image of a breast tumor using these three methods. The absorption spectra of the four main endogenous absorbers are shown for reference.

PCA [195] is a method often used for explorative data analysis and as a pre-processing dimensionality reduction step. PCA identifies a linear transformation of the data into a new orthogonal representation where the first component matches the direction of the largest data variance, the next component is an orthogonal direction with the largest remaining variance, etc. Statistically, PCA is based on the assumption that the components (i.e., the constituting spectra) are uncorrelated. PCA can be computed through the singular value decomposition of the data matrix $\boldsymbol{S}$ (after subtracting its column-wise mean):

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1^T \\ \vdots \\ \boldsymbol{s}_{N_V}^T \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T. \tag{2.7.6}$$

Here, $\boldsymbol{U}$ is a matrix of shape $N_V \times N_V$ of the left singular vectors, $\boldsymbol{V}$ is a matrix of shape $L \times L$ of the right singular vectors, and $\boldsymbol{\Sigma}$ is a diagonal matrix of shape $N_V \times L$ containing the singular values in descending order. The (orthonormal) columns of the matrix $\boldsymbol{V}$ are the principal components and the transformation is applied as:

$$\hat{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{V}. \tag{2.7.7}$$

In practice, when only K main components are of the interest, a truncated matrix $\boldsymbol{V}_K$ with the first K columns of $\boldsymbol{V}$ is used instead. Singular values on the diagonal of the matrix $\boldsymbol{\Sigma}$ are equal to the amount of data variance captured by each component, which can be used to select K such that the transformed matrix $\hat{\boldsymbol{S}}$ contains a certain amount of the original data variance (e.g., 95%).

ICA is another method to identify a linear transformation of the data matrix into a new representation. Unlike PCA, ICA relies on a stronger assumption that the individual components are statistically independent, but they do not necessarily have to be orthogonal. There are various algorithms to find the components maximizing statistical independence, a popular one being FastICA [196]. FastICA utilizes the central limit theorem stating that a sum of statistically independent variables tends towards normal distribution, i.e., the individual variables are expected to be less gaussian than their sum. A fixed-point iteration scheme is used, measuring the non-normality of the components using kurtosis. A pre-requisite for ICA is preprocessing the data matrix to have a zero mean and a unit variance along the columns (whitening), which can be achieved via singular value decomposition. Since ICA produces L components without any importance measure like PCA, the two methods are often used in tandem: PCA performs dimensionality reduction and whitening of the data which are then transformed with ICA.

ICA has been used for BU of optoacoustic data in several works: Glatz *et al.* have used it to detect exogenous optical agents in a mouse model [197]; Arabul *et al.* have attempted to separate morphological parts of human carotid plaques *ex vivo* using a non-negative variant of ICA [198]. However, the ICA assumption of statistical independence conflicts with the expected distribution of endogenous chromophores in the biological tissue. Notably, oxy- and deoxyhemoglobin are always present together in the blood, so their distributions are not independent. Furthermore, hemoglobin is not expected to be present in the subcutaneous fat layer, breaking the independence of lipid and hemoglobin distributions. Thus, ICA is better suited for detection of exogenous contrast agents and targeted reporters whose distribution can be modelled as independent of the tissue chromophores [199].

NNMF is an approach based on the assumption that the data were generated as an additive combination of a limited number of non-negative components, i.e., the data matrix $\boldsymbol{S}$ can be factorized into two non-negative matrices:

$$\boldsymbol{S} \approx \hat{\boldsymbol{C}}\hat{\boldsymbol{A}}. \tag{2.7.8}$$

Here, $\hat{\boldsymbol{A}}$ is a matrix of the spectral components ($K \times L$) and $\hat{\boldsymbol{C}}$ is a matrix of their coefficients at each pixel ($N_V \times K$). The value of K is a parameter of the algorithm and when $K \leq L$, it plays the role of a compression bottleneck—the algorithm attempts to find the best representation of the data while using only K components. In other words, the factor matrices have less elements than their product, enforcing data compression. The factorization is performed as a constrained optimization problem:

$$\underset{\hat{\boldsymbol{C}}\geq 0, \hat{\boldsymbol{A}}\geq 0}{\operatorname{argmin}} \left\| \hat{\boldsymbol{C}}\hat{\boldsymbol{A}} - \boldsymbol{S} \right\|_F. \tag{2.7.9}$$

The above optimization may be extended with additional regularizer terms to promote certain properties of the unmixing solution. For example, Jüstel and Irl *et al.* have used L_1 and L_2 penalties on the factor matrices to promote sparsity of the unmixed representation [200]. Grasso *et al.* have combined NNMF with a light fluence correction model and super-pixel analysis to arrive at a more robust framework for BU of OA data [201]. Fig. 2.5 shows that the results of NNMF unmixing are

the most similar to the absorption spectra of the main chromophores and hence offer better interpretability than the results of PCA and ICA.

Overall, BU has not seen widespread adoption in the OAT image analysis. However, as the OAT image quality increases, so are the requirements for the precision of the chromophore analysis. Since the linear unmixing model cannot achieve the required level of accuracy in the absence of light fluence correction, BU algorithms might be used more frequently in the future.

CHAPTER 3

Motion quantification and frame selection

Multispectral optoacoustic tomography (MSOT) facilitates resolving unique functional features based on optical contrast, including detailed visualization of vasculature and of tissue oxygenation. Many of the pathophysiological features recorded require the recording of absorption spectra of light absorbing tissue moieties, such as oxygenated or deoxygenated hemoglobin or lipids. To achieve this, MSOT records OA images sequentially at different illumination wavelengths to form a multispectral (MS) frame. Nevertheless, motion can impact the accurate unmixing of the relative contributions of tissue constituents on the collected spectrum. Even if modern MSOT systems can acquire a 25-wavelength MS frame within 1 second, any motion occurring within that period corrupts the measured spectra. A typical MSOT recording may contain a combination of stationary frames and frames corrupted by probe motion, especially if the operator moves the probe over the imaged volume, localizing suitable imaging positions and acquiring the data continuously. The stationary frames, affording the best spectral quality, need to be selected for further analysis. Currently, this selection is done manually, which is neither quantitative nor reproducible and is time consuming. An automatic method for selection of stationary frames in MSOT scans would simplify and standardize the image analysis pipeline and ensure optimal spectral quality of analyzed images.

It has been recognized that motion presents an issue for MSOT and researchers have proposed various methods to suppress the negative effects of motion. In pre-clinical MSOT studies, artifacts caused by periodic motion following a regular pattern, such as heartbeat and breathing, have been suppressed with the help of frame rejection and motion clustering [202, 203]. These methods cannot be simply translated to the clinical handheld setting with highly irregular and non-periodic motion. Another option to mitigate the influence of motion, as demonstrated in a pre-clinical setting, is to use very fast illumination frequency in the kilohertz range [204]. In the clinical MS setting however, the illumination frequency is limited by the wavelength switching speed of the employed laser (up to 50 Hz in the current *Acuity Echo*® implementations) and by the ANSI maximum energy exposure limits for human skin, which regulates the permissible amount of energy per unit of time and presents a trade-off between energy-per-pulse and pulse frequency [113].

In clinical studies, motion correction has been employed to align the images within one MS frame to reduce motion-induced spectral artifacts [47, 205-207]. While effective at suppressing some of the motion artifacts, this approach struggles to align images at different wavelengths that vary too much in their appearance since they highlight different absorbers (which is desired in MS imaging).

This chapter is based on the following manuscript intended for a later publication: Kukačka, Jan, Dominik Jüstel, and Vasilis Ntziachristos. "Motion score: Ensuring optimal spectral quality in handheld MS-OPUS scans through automated selection of stationary frames" in preparation (2022).

Furthermore, motion correction algorithms tend to be sensitive to their parameter settings and may require manual tuning to work well on a specific dataset. However, in principle, motion correction is orthogonal to stationary frame selection and can be applied to reduce the remaining motion artifacts in stationary frames.

On the other hand, the problem of automatic selection of the optimal MS frames and the related task of motion quantification in MSOT scans have been neglected so far. A simple approach to address the quantification of motion in MSOT scans could be the evaluation of cross-correlation (or some other similarity measure) between consecutive images in the scan sequence. However, such an approach would suffer from the inability of a simple metric to correctly estimate various types of motion affecting the image appearance in complex ways. Furthermore, small jitter cannot be distinguished from more serious steady drift when considering only neighboring images.

Therefore, a robust method is needed for quantification of motion in MSOT scans to enable a standardized, automated selection of the most stationary frames, ensuring optimal spectral quality. To address this need, we present *Motion score*—the first algorithm for robust quantification of motion in hybrid MS-OPUS scans and automatic selection of optimal stationary MS frames for downstream analysis. Relying solely on robust evaluation of similarity of the images in the US stream, *Motion score* does not require any external tracking or hardware modifications and can be directly applied to any existing MS-OPUS scans. We validate the ability of our algorithm to select optimal frames on a dataset of eight scans of an agar tube phantom and three *in vivo* scans of arteries. We demonstrate that *Motion score* overcomes the limitations of other methods for stationary position detection, outperforming them in terms of precision, recall, and mean average precision. Furthermore, we show on examples of phantom and *in vivo* scans that the MS frames selected by our algorithm do not suffer from motion-induced artifacts and thus yield the optimal spectral quality required for precise clinical analysis. Finally, we make a ready-to-use, open-source implementation of *Motion score* available in a public code repository.[6]

3.1 Methods

3.1.1 Multispectral optoacoustic-ultrasound tomography

MS-OPUS is an imaging technology that acquires two streams of data in parallel using a single handheld probe (Fig. 3.1a). An MS-OPUS scan contains a sequence of US images, $I_{US}^{(j)}$, $j = 1, \dots, N_{US}$, and a sequence of single-wavelength OA images, $I_{OA}^{(k)}$, $k = 1, \dots, N_{OA}$. As shown in Fig. 3.1b, the OA images are recorded cyclically at wavelengths from a predefined set Λ, depending on the system preset. One MS frame[7] consists of $N_\Lambda = |\Lambda|$ consecutive single-wavelength images spanning the whole set of acquired wavelengths: $I_{MS}^{(k)} = \left\{I_{OA}^{(k)}, I_{OA}^{(k+1)}, \dots I_{OA}^{(k+N_\Lambda-1)}\right\}$, $k = 1, \dots, N_{MS} = N_{OA} - N_\Lambda + 1$. The OA and US images are acquired in an interleaved fashion, and each image is assigned a timestamp $t_{OA}^{(k)}$, $t_{US}^{(j)}$. Thus, we can define a sequence $\boldsymbol{s}^{(k)}$ of indices of US images corresponding to the k-th MS frame to be the ordered sequence of indices in the set $\left\{i \in \mathbb{N} \middle| t_{OA}^{(k)} \leq t_{US}^{(i)} \leq t_{OA}^{(k+N_\Lambda-1)}\right\}$. In the following text, we omit the frame index (k) where it is obvious from the context.

[6] www.github.com/jankukacka/optimal_frames

[7] Contrary to the usual protocol, we do not require the start of a MS frame to be aligned to the lowest wavelength image (i.e., $k \equiv 1 \pmod{N_\Lambda}$), but we still sort the images within the MS frame by wavelength before further processing. This has two implications: a MS frame can start at any single-wavelength image in the sequence and two consecutive MS frames share $N_\Lambda - 1$ single-wavelength frames.

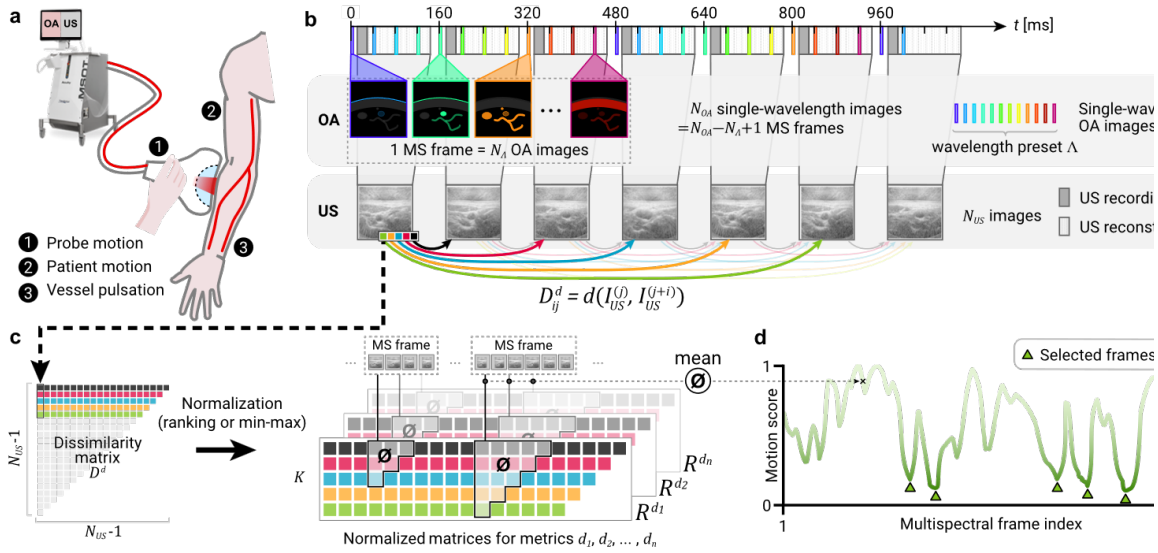

Figure 3.1: MS-OPUS imaging and *Motion score* computation. a, Schematic of the handheld MS-OPUS operation with sources of motion highlighted. **b,** MS-OPUS scanning scheme. OA and US images are acquired at different frequencies in an interleaved fashion. Several US images are acquired during recording of a single MS frame. **c,** Dissimilarities between US images and their subsequent neighbors form the matrix D^d. Values of D^d are row-wise ranked to obtain a rank-normalized matrix R^d. Multiple dissimilarity measures $d_1, \dots, d_n$ can be combined since the ranking normalization equalizes their scales. The *Motion score* of a MS frame is then computed as the mean of all dissimilarities between its corresponding US images (i.e., the US images that were acquired while recording the MS frame). The number of corresponding US images may vary between MS frames. **d,** A peak-finding algorithm is applied on the *Motion score* vector of the whole MS-OPUS scan (where a lower *Motion score* means less motion) to identify up to N_{peaks} stationary MS frames. **Abbreviations:** MS, multispectral; OA, optoacoustic; US, ultrasound.

3.1.2 Motion quantification

Let d be a dissimilarity function on the space of images (e.g., L_2-norm). Assuming that two images will be similar in terms of d if there is no motion occurring between their acquisitions, we can utilize d for motion quantification. However, this assumption does not necessarily hold true for OA images acquired at different wavelengths which may display different absorbers, giving them a distinct appearance even in the absence of motion. The US images, on the other hand, have a consistent appearance and can be used to estimate the motion occurring during the MS-OPUS scanning. A naïve approach to compute motion m_0 in a MS frame would be to take the mean of dissimilarities between consecutive US frames in its corresponding sequence $\boldsymbol{s}$:

$$m_0 = \frac{1}{|\boldsymbol{s}|} \sum_{j=\boldsymbol{s}_1}^{\boldsymbol{s}_{|\boldsymbol{s}|-1}} d\left(I_{US}^{(j)}, I_{US}^{(j+1)}\right). \tag{3.1.1}$$

This approach has two limitations. First, it cannot differentiate between (acceptable) jitter and (much more problematic) steady drift. Second, it suffers from biases of the chosen dissimilarity function d towards penalizing certain types of motion more than others. To overcome these limitations, we propose *Motion score*, a robust motion quantification framework for MS-OPUS.

3.1.3 Motion score

Fig. 3.1c outlines the *Motion score* computation. Let $\boldsymbol{D}^d \in \mathbb{R}^{(N_{US}-1)^2}$ be a matrix of dissimilarities between US images within one MS-OPUS scan measured by d:

$$\boldsymbol{D}_{ij}^{d} = \begin{cases} d\left(I_{US}^{(j)}, I_{US}^{(j+i)}\right) & \text{if } i+j \leq N_{US} \\ \text{undefined} & \text{else.} \end{cases}, \tag{3.1.2}$$

Element j in row i corresponds to the dissimilarity between the j-th image and its i-th subsequent image in the US image sequence. Elements under the anti-diagonal are undefined since they correspond to subsequent images which would be beyond the end of the sequence. In practice, only the first $K = \max_k \left|\boldsymbol{s}^{(k)}\right|$ rows of the matrix $\boldsymbol{D}^d$ must be computed, while the rest is unused. Furthermore, let $\boldsymbol{R}^d$ and $\boldsymbol{N}^d$ be matrices with the same shape as $\boldsymbol{D}^d$ and elements in the range [0,1], where $\boldsymbol{R}^d$ is a matrix of normalized row-wise ranks of the elements of $\boldsymbol{D}^d$, $\boldsymbol{R}_{ij}^d = (\text{rank of } \boldsymbol{D}_{ij}^d \text{ in } \boldsymbol{D}_i^d)/(N_{US} - i - 1)$, and $\boldsymbol{N}^d$ is a matrix of row-wise min-max normalized elements of $\boldsymbol{D}^d$, $\boldsymbol{N}_{ij}^d = \left(\boldsymbol{D}_{ij}^d - \min_k \boldsymbol{D}_{ik}^d\right) \Big/ \left(\max_k \boldsymbol{D}_{ik}^d - \min_k \boldsymbol{D}_{ik}^d\right)$. Finally, let $\Delta = \{d_1, \ldots, d_n\}$ be a set of dissimilarity measures. Then we define the ranked *Motion score*, m_R^Δ, and the normalized *Motion score*, m_N^Δ, for a MS frame with a corresponding sequence of US image indices $\boldsymbol{s}$ as:

$$m_R^\Delta = \frac{2}{|\Delta||\boldsymbol{s}|^2} \sum_{d \in \Delta} \sum_{i=1}^{|\boldsymbol{s}|} \sum_{j=\boldsymbol{s}_1}^{\boldsymbol{s}_{|\boldsymbol{s}|-i}} \boldsymbol{R}_{ij}^d \ , \tag{3.1.3}$$

$$m_N^\Delta = \frac{2}{|\Delta||\boldsymbol{s}|^2} \sum_{d \in \Delta} \sum_{i=1}^{|\boldsymbol{s}|} \sum_{j=\boldsymbol{s}_1}^{\boldsymbol{s}_{|\boldsymbol{s}|-i}} \boldsymbol{N}_{ij}^d \ . \tag{3.1.4}$$

3.1.4 Frame selection

Fig. 3.1d shows a motion vector $\boldsymbol{m}$ of length N_{MS} obtained by computing the motion of all MS frames in a scan. To automatically select the best stationary frames, a peak-finding algorithm `find_peaks` of the SciPy library (v1.8.0) [208] was used to identify local minima of the vector $\boldsymbol{m}$. The minimal peak distance was set to 1/20 of the scan length ($\lfloor N_{MS}/20 \rfloor$), which was empirically observed to be a robust value over a large variety of scans used during the development of our method, and the other parameters were left at their default settings. The detected peaks were sorted by ascending values and the first N_{peaks} were taken, where the value of N_{peaks} was adjusted to the number of expected stationary positions depending on the scanning procedure. We found $N_{peaks} = 5$ to work well in practice; higher values decrease the risk of missing a desired location but increase the chance of false positives.

3.2 Experiments

3.2.1 Motion score validation

Our first goal was to validate the ability of *Motion score* to automatically select stationary frames in MS-OPUS scans and to compare its performance to other metrics.

MS-OPUS setup. We used two datasets acquired with *MSOT Acuity Echo*® scanners (iThera Medical GmbH, Munich, Germany) with a wavelength preset $\Lambda = \{700 \text{ nm}, 710 \text{ nm}, \ldots, 970 \text{ nm}\}$, $N_\Lambda = 28$. The acquisition frequency was 25 Hz and 6.25 Hz for the OA single-wavelength images and the US images, respectively.

Table 3.1: Composition of layers of the tissue-mimicking agar phantom.

Layer	Deionized water (l)	Agar (g)	Glycerol (ml)	Psyllium (g)
1	1.0	36.0	50	5.8
2	0.5	7.5	25	5.0
3	0.5	10.0	25	7.0
4	0.5	12.0	25	5.0
5	0.5	10.0	25	5.0

Dataset 1 (10 scans) was obtained using a probe with 256 transducers (3.4 MHz central frequency) arranged in a 4 cm wide 125° arc filled with a gel pad for acoustic coupling. Various types of probe motion (linear, rotational, slow and fast, steady and jerking, etc.) are captured in this dataset. The first eight scans were conducted on a tissue-mimicking agar phantom with various optical absorbers (see below). Another two scans targeted the carotid artery of a healthy volunteer. The duration of the scans is between 35 s and 120 s; in total, the dataset contains 11.6 minutes of recorded data.

Dataset 2 (1 scan) was obtained with an older prototype using a probe with 256 transducers (4 MHz central frequency) arranged in a 6 cm wide 145° arc filled with heavy water. The scan, showing the radial artery from a healthy volunteer, resembles a typical MS-OPUS clinical acquisition. We selected this dataset to demonstrate the efficiency of our method on a generic acquisition and to showcase the negative effects of motion on the spectral quality in an in vivo example.

Phantom manufacturing. A tissue-mimicking agar phantom (Fig. 3.2) containing various acoustic and optical imaging targets was manufactured using the following procedure. Five batches of phantom material were mixed according to the quantities listed in Table 3.1 and warmed up to 90°C while continuously stirring to avoid creation of air bubbles. The heated mixture was poured into a box of size 20x30x20 cm and allowed to cool down and solidify for 30 min before adding the next layer. Material of varying density was used for different layers to mimic acoustic interfaces inside the tissue. Psyllium husks were added to create acoustic inhomogeneities providing texture on US images. No other optical scatterers or absorbers were admixed to the phantom material, resulting in a semi-opaque appearance. Three types of objects made of denser agar (same as layer 1, see Table 3.1) were embedded into the phantom body: (1) agar blocks of various shape and size, (2) cylinders of denser agar containing indocyanine green (15 μg/ml), (3) blocks of denser agar with cylindrical cavities filled with optical absorbers (vegetable oil, oil red O stain). Whereas all three types are visible on the ultrasound, only targets (2) and (3) are visible in the OA images. Positions of all the insertions were noted during the phantom preparation.

Stable position annotation. To derive the ground truth annotations, three separate human experts selected all stationary frame ranges in the US videos in Dataset 1. Their annotations were summed to assign a score 0–3 to each US image; MS frames were assigned the mean score of all their corresponding US images, and scores ≥1.5 were considered stationary. For quantitative evaluation, the stationary position selection task was framed as an event detection problem. Any uninterrupted sequence of stationary MS frames was considered to be a target event and selecting any MS frame within such sequence was considered a true positive prediction (TP). On the other hand, selecting a frame outside a target event (stationary sequence) was a false positive prediction (FP) and any event not covered by any prediction was counted as a false negative (FN).

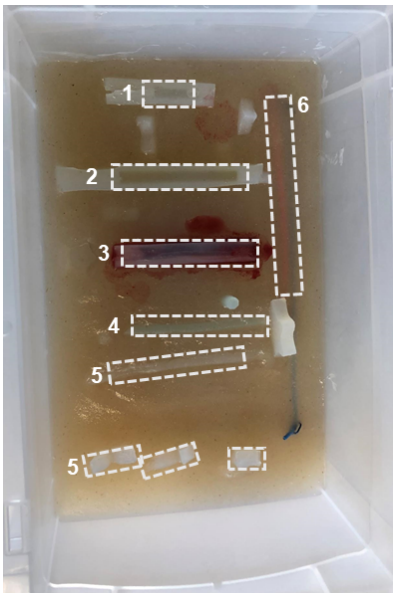

Figure 3.2: Tissue mimicking phantom used in Dataset 1, scans 1–8. Dashed lines mark following insertions: (1) cavity with sunflower oil, (2) cavity with olive oil, (3) cavity with oil red O, (4) cylinder of agar with indocyanine green, (5) stiff agar, (6) rubber tubes filled with water. The photo captures the phantom before adding the last agar layer.

Motion quantification. Motion in all scans of Dataset 1 was quantified using five different methods for comparison and the same frame-selection procedure was used on the motion vectors computed for all five methods (see section 3.1.4). To establish a baseline, the simple method (eq. (3.1.1)) was used with cross-correlation (XC), zeroed-normalized cross-correlation (ZNXC), and optical flow as dissimilarity measures:

$$XC\left(I_{US}^{(i)}, I_{US}^{(j)}\right) = -I_{US}^{(i)} \cdot I_{US}^{(j)}, \tag{3.2.1}$$

$$ZNXC\left(I_{US}^{(i)}, I_{US}^{(j)}\right) = -\left(\frac{I_{US}^{(i)} - \mu^{(i)}}{\sigma^{(i)}}\right) \cdot \left(\frac{I_{US}^{(j)} - \mu^{(j)}}{\sigma^{(j)}}\right), \tag{3.2.2}$$

where I_{US} is a vector of pixels contained in an ultrasound image and $\mu^{(i)}$, $\sigma^{(i)}$ are the mean and the sample standard deviation of the image pixel intensities, respectively. Optical flow was computed by applying a pre-trained FlowNet2 [209] (weights obtained from www.github.com/NVIDIA/flownet2-pytorch) on pairs of ultrasound images resized to 192x192 pixels. The resulting vector field was converted to a scalar by taking the mean of L_2 norms of the displacement vectors at every pixel.

Furthermore, the normalized *Motion score* (eq. (3.1.3)) and the ranked *Motion score* (eq. (3.1.4)) were computed using a set of two dissimilarity measures: structural similarity (SSIM) [210] and ZNXC. To compute SSIM, implementation of scikit-image (v0.19.2) [211] was used with default parameters. Since *Motion score* requires distances instead of similarity metrics, negative value of the SSIM was used.

Evaluation metrics. To evaluate the accuracy of the five motion quantification methods, precision, recall, and mean average precision (mAP) were used. Precision is defined as the ratio of true positives in all predicted positions (TP/TP+FP). Recall is defined as the fraction of target events that were correctly detected, computed as the ratio of unique true positives (TP_1) in all positives (TP_1/TP_1+FN). TP_1 is a subset of TP allowing at most one predicted position per target event. For computing precision and recall, N_{peaks} was set to match the number of stationary positions

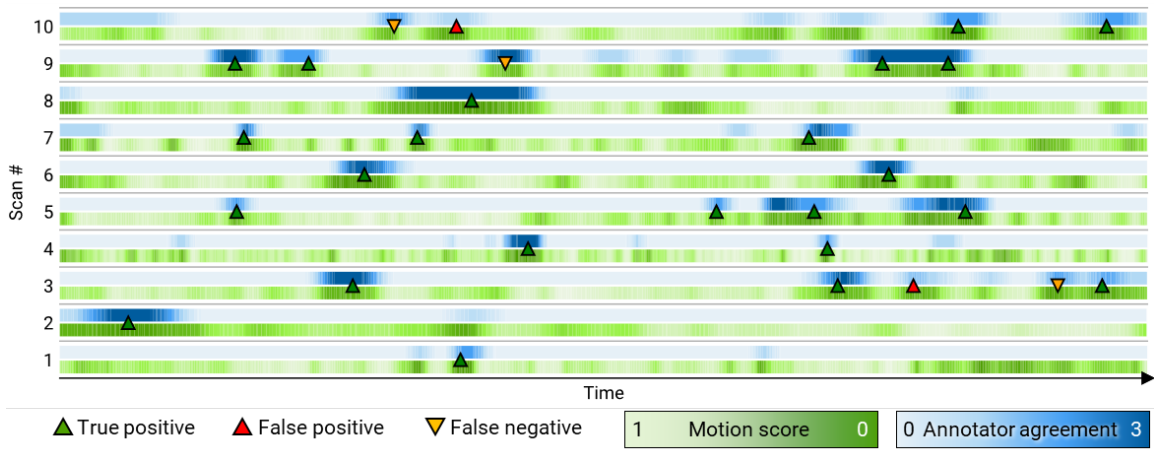

Figure 3.3: Stationary position detection evaluation on MS-OPUS scans. Three human annotators reviewed US videos of ten MS-OPUS scans (Dataset 1) and identified 25 stationary positions (blue). *Motion score* of those scans was also computed (green) and the most optimal frames were selected (n=N_{peaks}), such that N_{peaks} matches the number of true stationary positions. The selected frames are marked by green (true positive) and red triangles (false positive). Stationary positions missed by our algorithm (false negatives) are marked by yellow upside-down triangles.

identified by the human annotators in each scan. Since this number is generally unknown, the area under the precision-recall curve for all values of N_{peaks} was also computed for each scan (average precision), and the mean over all scans was computed to obtain the mAP.

3.2.2 Spectral quality evaluation

Our second goal was to demonstrate that the frames selected by the *Motion score* algorithm indeed suffer from fewer motion-related artifacts and thus have better spectral quality. Here, we assume that a measured spectrum in a pixel should be a linear combination of the absorption spectra of the chromophores present in that pixel. This linear combination can be computed by linear spectral unmixing of the form:

$$\boldsymbol{C}^* = \underset{\boldsymbol{C}\geq 0}{\operatorname{argmin}} \|\boldsymbol{C}\boldsymbol{A} - \boldsymbol{I}_{MS}\|_F. \tag{3.2.3}$$

Here, $\boldsymbol{I}_{MS} \in \mathbb{R}^{p\times|\Lambda|}$ is a MS frame with spectra of its p pixels arranged in rows, $\boldsymbol{C} \in \mathbb{R}^{p\times k}$ is a matrix of unmixing coefficients, and $\boldsymbol{A} \in \mathbb{R}^{k\times|\Lambda|}$ is a matrix of k spectral components which are either fixed known spectra of pre-defined absorbers or optimized together with $\boldsymbol{C}$ by a data-driven BU procedure; $\boldsymbol{C} \geq 0$ denotes a non-negativity constraint on the elements of $\boldsymbol{C}$; and $\|\cdot\|_F$ denotes the Frobenius norm.

Processes violating the linear decomposition assumption cause spectral corruption. Aside from spectral coloring resulting from uneven light fluence in the tissue, motion is the main source of spectral corruption. Spectral corruption can be quantified as the magnitude of residuals after linear spectral unmixing, i.e., the amount of signal that cannot be explained by a linear mixing model. The residuals are expected to be low if the measured spectra can be decomposed (unmixed) into a linear combination of several spectral components corresponding to the absorbers in the image. However, if the spectrum is corrupted by motion, it will contain irregularities that cannot be unmixed into the expected components and the unmixing residuals will be high. The unmixing residual (error) is computed as the $L_{2,1}$-norm (sum of L_2-norms over the matrix columns) of the residuals relative to the norm of the image:

$$E(\boldsymbol{I}_{MS}, \boldsymbol{C}, \boldsymbol{A}) = \frac{\|(\boldsymbol{C}\boldsymbol{A} - \boldsymbol{I}_{MS})^T\|_{2,1}}{\|\boldsymbol{I}_{MS}^T\|_{2,1}} \,. \tag{3.2.4}$$

In our experiments, two types of spectral unmixing were used. For phantom scans from Dataset 1, blind unmixing by NNMF was used with $k = 4$ to match the number of expected absorbers in the images (contrast agent, agar, psyllium husks, and probe membrane). For the clinical scan from Dataset 2, standard linear unmixing according to eq. (3.2.3) with spectra of oxy- and deoxyhemoglobin, lipids, and water were used.

3.3 Results

3.3.1 *Motion score* accurately identifies stationary frames in MS-OPUS scans

Motion score has shown nearly perfect agreement with human annotators at finding stationary positions in MS-OPUS scans. Three annotators identified in total n=25 stationary positions in the ten scans of Dataset 1 (1–4 positions per scan). On average, the annotators needed 8.5 min for annotation of 1 min of US video. *Motion score* (ranked; $\Delta = \{ZNXC, SSIM\}$) values were used to select a number of frames in each scan to match the ground truth number. Fig. 3.3 shows that of these 25 selected frames, 23 were located within 22 of the true stationary positions (92% precision, 88% recall). Furthermore, the two false positive frames were in positions that were labeled stationary by at least one annotator (red triangles). Similarly, the three false negative locations (yellow triangles) were also aligned with *Motion score* local minima, but their *Motion score* values were higher than in the 25 selected frames. Since the number of stable positions per scan is generally unknown, we also evaluated the mAP of stationary position detection, which measures the precision-recall over all values of N_{peaks}, and obtained a high mAP of 91.67%. Table 3.2 shows a performance comparison of *Motion score* to other metrics. *Motion score* outperformed naïve approaches (rows 1–3), where only directly consecutive US images were compared, by a wide margin. In particular, unnormalized cross-correlation and optical flow performed poorly. Finally, the ranked *Motion score* achieved a better mAP by 3.3 percentage points more than the normalized version.

3.3.2 Low *Motion score* correlates with optimal spectral quality

We observed that stationary MS frames—having low *Motion score*—did not suffer from motion-related artifacts. Fig. 3.4 shows the examples of motion-related artifacts in two phantom scans from Dataset 1 and the in vivo scan from Dataset 2. The phantom scans depict agar cavities filled with absorbers having distinct spectral appearances: oil red O organic dye (Fig. 3.4a) with a spectrum

Table 3.2: Stationary position detection performance. Performance of five methods for identifying stationary positions evaluated on Dataset 1. **Abbreviations:** ZNXC, zeroed normalized cross-correlation; SSIM, structural similarity; mAP, mean average precision.

Method	Precision	Recall	mAP
Cross-correlation	4/25	5/25	22.50%
Zeroed normalized cross-correlation	20/25	19/25	74.17%
Optical flow (FlowNet2)	17/25	15/26	58.33%
Motion score normalized (ZNXC + SSIM)	22/25	21/25	88.33%
Motion score ranked (ZNXC + SSIM)	**23/25**	**22/25**	**91.67%**

gradually decreasing with increasing wavelength, and olive oil (Fig. 3.4b) with a narrow absorption peak at 930 nm. Motion corrupts the spectra in both cases, regardless of the absorber. The images on the left, having a low *Motion score*, represent stationary frames, whereas the images on the right show frames corrupted by motion along the dashed arrows. In both cases, BU detected either spectral components with a different shape or different spatial distribution in the image (middle row). In Fig. 3.4a, the absorption spectrum in the stationary frame was decomposed into two components representing variations in light fluence, colocalized around the cavity edges. In the moving frame, three components were identified with disjoint absorption peaks at 700 nm, 760 nm, and 820 nm, appearing in the image as three crescents horizontally shifted along the motion direction. In Fig. 3.4b, two similar spectral components were identified in both images—component 1 (blue) fitting the true absorption peak of olive oil at 930 nm, and component 2 (orange) with a peak shifted towards 950 nm. In the stationary frame, only component 1 is present in the cavity, whereas the moving frame contains two, partly disjoint discs of the two components. In both cases, unmixing results in the presence of motion defy the true phantom composition, demonstrating the unsuitability of such frames for spectral analysis.

Furthermore, we observed that low *Motion score* indicates good spectral quality. To measure the spectral quality, we evaluated the unmixing residuals. The scatter plots in the bottom rows of Figs. 3.4a,b show the relationship between *Motion score* and the unmixing residuals. Although there were frames with low unmixing residuals despite high *Motion score* (Fig. 3.4b, gray circle—these frames contain motion along the tubular cavity and hence the position of the absorber cross-section remains stable), there were no frames with low *Motion score* and high unmixing residual, demonstrating that a low *Motion score* is a good indicator of optimal spectral quality.

Finally, we verified that *Motion score* is effective at selecting stationary frames with good spectral quality in clinical scans. Fig. 3.4c demonstrates motion-induced spectral artifacts in a clinical scan of a radial artery (dashed circle) from a healthy volunteer. The upper row shows unmixing residuals (same color scale) for a stationary (left) and moving (right) frame. The bottom row shows visualization of the unmixed chromophores (oxy- and deoxyhemoglobin, lipid, and water) over

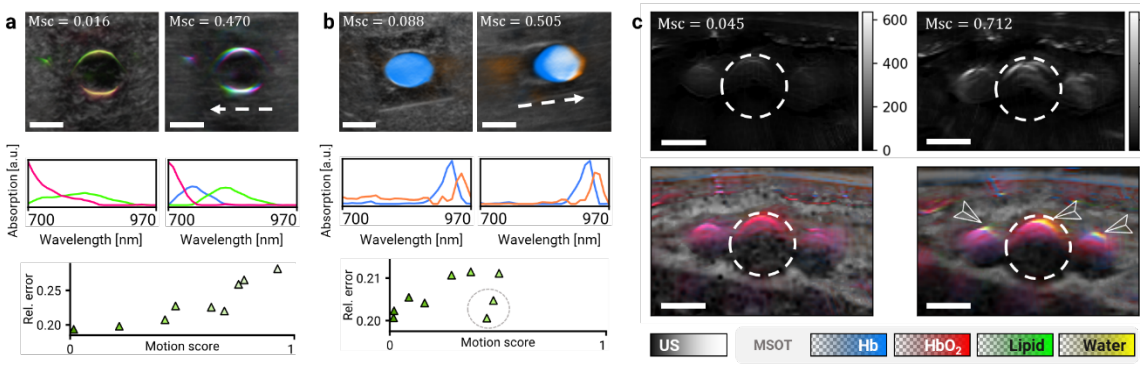

Figure 3.4: Lower *Motion score* leads to fewer motion artifacts and better spectral quality. a, Scans of a cavity filled with oil red O embedded in an agar phantom without motion (left) and with motion (right). Grayscale US images are overlaid by color-coded concentrations of components identified via non-negative matrix factorization (NMF). Below, their spectra are shown in the matching colors. Bottom: relationship between *Motion score* and relative NMF unmixing residuals (4 components). **b,** The same analysis as in (a) for a scan of a cavity filled with olive oil. The gray circle in the bottom scatterplot denotes frames in which motion was parallel to the tubular cavity and did not cause considerable spectral corruption. **c,** Scans of a radial artery (dashed circles) and concomitant veins without motion (left) and with motion (right). Top row: Residuals after linear unmixing (both using the same color scale) show that motion causes spectral artifacts around the blood vessel edges that cannot be properly unmixed. Bottom row: visualization of unmixing coefficients of Hb (blue), HbO_2 (red), water (yellow), and lipid (green). Motion-related spectral artifacts appear as additional stripes of water and fat signal around the artery and veins (arrows). Scalebars 5mm. **Abbreviations:** Msc, Motion score; US, ultrasound; Hb, deoxyhemoglobin; HbO_2, oxyhemoglobin.

grayscale US. Whereas only hemoglobin is detected inside the blood vessels in the stationary frame (as expected), unmixing of the motion-corrupted frame contains erroneous signals of lipids and water around the blood vessel edges (arrows).

3.4 Discussion

In this chapter, we present *Motion score*—a new algorithm for motion quantification and automated stationary frame selection in MS-OPUS scans. We have validated the accuracy of our method and demonstrated its superiority to simpler approaches. Furthermore, we have shown that frames selected by *Motion score* have fewer motion-related artifacts, yielding optimal spectral quality.

The selection of stationary frames is crucial, as moving frames suffer from various spectral artifacts that hinder molecular analysis. In the radial artery scan affected by motion, we observed increased water and lipid signals around the blood vessel edges. This type of unmixing error would have a negative effect on the accuracy of clinical MS-OPUS applications such as the evaluation of post-prandial lipemia [212] or carotid plaques [63]. On the other hand, frames selected using *Motion score* showed consistently good spectral quality that is needed for precise molecular analysis.

The automatic selection of the stationary frames eliminates the requirement for manual selection, which is non-reproducible and time consuming. In our experiment, a human annotator needed eight times more time to select good frames than the length of the scan itself. The economical improvement that *Motion score* provides is a clear, quantitative justification for its use in standardized MS-OPUS analysis.

On our motion quantification benchmark, *Motion score* outperformed all other methods that we evaluated. The algorithm tackles both issues of the naïve approach presented in eq. (3.1.1). First, *Motion score* considers the distances between all pairs of US images within one MS frame, which enables penalization of undesirable continuous drifting motion. Second, *Motion score* alleviates the bias resulting from the selection of only one particular metric by combining any number of dissimilarity measures using the rank-score or min-max normalization. The normalization equalizes different value scales and makes the measures comparable. The rank-score normalization is less sensitive to outliers than min-max normalization, as an image pair with exceptionally high dissimilarity cannot outweigh the relative differences between other pairs since only the order of the values is considered. Ranking in *Motion score* has an additional advantage when comparing MS frames with different numbers of corresponding US frames (which happens unless $\exists k \in \mathbb{N}: k/f_{US} = N_\Lambda/f_{OA}$): since the US images acquired further apart tend to have a higher dissimilarity, their perceived motion would also tend to be higher—using normalized ranks eliminates this bias. These advantages, together with better performance in our experiment, make the ranked *Motion score* the preferable variant.

Motion score is a simple, yet robust algorithm. Herein, we validated *Motion score* using scans acquired with a 28-wavelength preset, where most MS frames had 9 or 10 corresponding US images. Possible performance deterioration could occur in illumination presets with fewer wavelengths (and fewer US images per MS frame). This could be remedied by simply combining several MS frames together for the purpose of motion quantification, though this might not be so crucial since the acquisition of fewer wavelengths is faster and less prone to motion artifacts. Due to its generic nature, *Motion score* can be extended by including additional similarity measures aside from ZNXC and SSIM shown herein. On the other hand, adding poorly performing metrics, such as the unnormalized cross-correlation or optical flow computed by a general purpose pre-trained FlowNet2, would reduce the performance and should be avoided. Despite its motion

quantification capabilities, *Motion score* cannot be used as a general MS-OPUS scan quality measure since it produces relative values that are not comparable between scans. Furthermore, it also cannot be used as live feedback for the MS-OPUS operator since the whole scan must be finished before *Motion score* can be computed. Other motion quantification methods are needed for those scenarios to further advance the quality assurance in MS-OPUS.

Overall, using *Motion score* for automated stationary frame selection has clear benefits for spectral quality achieved by MS-OPUS. Furthermore, it is a step towards standardization of MS-OPUS processing, mitigating the inter-operator variability during scanning and image analysis. Our open-source implementation can be immediately used off-the-shelf in any MS-OPUS analysis pipeline. Altogether, this work is paving the way for better utilization of MS-OPUS for clinical applications.

CHAPTER 4

Image processing

Image processing controls the quality and the fidelity of OA images, which are essential factors for clinical translation of OAT. In this chapter, we present improvements to the image reconstruction and post-processing of the reconstructed images. Image reconstruction method utilizing total impulse response (TIR) correction and a heterogenous-speed-of-sound model is presented first in section 4.1. A post-processing method for averaging several multispectral frames utilizing elastic motion correction is then presented in section 4.2. We term the proposed pipeline *second-generation optoacoustic-ultrasound tomography* (2G-OPUS).

We hypothesize that using 2G-OPUS pipeline could lead to the most accurate image quality ever achieved in handheld OA imaging. We demonstrate the achieved performance on breast cancer scans from our study (Chapter 7). We characterize the size and contrast of the smallest structures that can be reliably resolved using 2G-OPUS. Our experimental results are presented in section 4.3.

4.1 Image reconstruction

Optoacoustic images are created from the recorded signals in a processing step called *image reconstruction*. Image reconstruction algorithms "reconstruct" the scanned object from the recorded signals using a mathematical model of the imaging system. The system model often relies on simplifying assumptions—either because they afford computationally simpler solutions, or because the exact properties of the system are unknown. However, the use of such simplifying assumptions leads to a mismatch between the recorded signals and the signals that would have been recorded had the model been exact. Equivalently, a hypothetical object that would have produced the recorded signals under the simplified model differs from the true imaged object. Hence, the use of a simplified model implies an incorrect solution. Typically, simplified models cause image distortions and fictious image features—*artifacts*.

A system model accounts for the detector properties such as the spatial arrangement, element shape, impulse response, or bandwidth. Furthermore, the speed of sound propagation in the tissue and the coupling medium are important parameters. Whereas some of the parameters are specified by the device manufacturer, others must be obtained experimentally via *system characterization* [162].

A simple mathematical model, assuming homogeneous speed-of-sound and point-like detectors, is used by the universal back-projection algorithm. This widely used method affords a fast solution, yet the analytical formula applies only to spherical, cylindrical, or planar detector geometries [172].

This chapter is partially based on the following publication whose parts appear verbatim without further notice: Kukačka, Jan, Stephan Metz, Christoph Dehner, Alexander Muckenhuber, Korbinian Paul-Yuan, Angelos Karlas, Eva Maria Fallenberg, Ernst Rummeny, Dominik Jüstel, and Vasilis Ntziachristos. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.

Time-reversal algorithms support arbitrary detector geometries, speed-of-sound heterogeneity, and frequency-dependent sound attenuation [173]. However, they only offer an analytical solution in their simple form, which is equivalent to back-projection. On the other hand, model-based approaches are computationally more demanding, but in many aspects, they have advantages over other schemes. They afford explicit integration of the physical properties of the imaging system into the forward model, such as detector geometry, shape [175, 186], and electrical response [181, 186]. Furthermore, the model can also account for the light [187] and sound [168, 188] attenuation in the medium. Overall, model-based reconstruction leads to images with less artifacts and considerably better image quality than back-projection [177, 181].

The integration of both the detector electrical impulse response and spatial impulse response is called TIR correction. Recently, a synthetic method for TIR derivation has been proposed [176], enabling characterization of the imaging system based on sparse experimental measurements of point sources. Avoiding the need for measurements on a dense grid vastly simplifies the model characterization. Furthermore, it has been shown that this approach allows appropriate modeling of the refraction caused by the acoustic impedance mismatch of the coupling medium and the tissue, which otherwise has a severe effect on the resulting image quality. We make use of this system characterization method and demonstrate its benefits on clinical data.

4.1.1 Method

We reconstruct the images by computing the initial pressure $\boldsymbol{p}_0$ from the filtered signals $\boldsymbol{s}$ using an iterative model-based approach. We use an acoustic model $\boldsymbol{M}$ of the scanner which accounts for the different SoS in the tissue and in the probe cavity filling (heavy water), the wave refraction on their interface, and for the physical properties of the transducers summarized in the TIR of the system [176]. Additionally, we use Tikhonov regularization to address the ill-posedness of the inverse problem and to mitigate the limited view artifacts and measurement noise [177, 178]. The regularization parameter α is chosen using an L-curve approach [179]. A non-negative LSQR algorithm is used to solve the optimization problem [184]:

$$\boldsymbol{p}_0 := \underset{\mathbf{p}\geq 0}{\operatorname{argmin}} \|\boldsymbol{M}\boldsymbol{p} - \boldsymbol{s}\|_2^2 + \alpha\|\boldsymbol{p}\|_2^2 \,. \tag{4.1.1}$$

The reconstructed OA images are of the size 401 x 401 pixels and correspond to a field of view (FOV) of 4 cm × 4 cm. Reconstruction of one multispectral image requires 15–30 min on a computer with an Intel® Xeon® E5-2630 CPU and 128 GB RAM. The implementation of our reconstruction code has been made available in a public code repository.[8]

4.2 Motion correction and frame averaging

Contrast in optoacoustic images decreases sharply with the depth. The main cause is elastic photon scattering, resulting in an exponential decay of the illumination fluence with the depth. Furthermore, the optoacoustic pressure waves are also attenuated while travelling through the tissue, albeit to a much lesser extent. The detection of the pressure waves is performed by an electronic system operating at a certain noise level. If the intensity of the incoming signals decreases below the noise level, image features can no longer be reconstructed. The reconstruction algorithms cannot distinguish between signal and noise and all recorded signals are treated equally. As a result, the noise-portion of the recorded signal is reconstructed along with the true signal,

[8] https://github.com/juestellab/mb-rec-msot

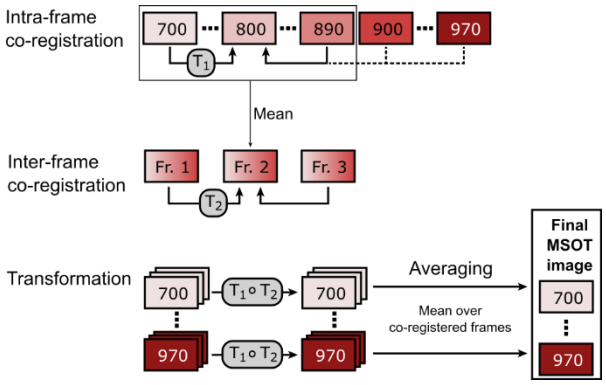

Figure 4.1: Overview of the motion correction and the frame averaging post-processing. First, images at wavelengths 700–890 nm are registered to the image at 800 nm (intra-frame elastic displacements T_1) and their mean is taken to represent the multispectral frame. Three consecutive multispectral frames are then registered to the middle frame (inter-frame elastic displacements T_2). Images at wavelengths above 890 nm cannot be reliably registered to 800 nm because they are visually very different; T_1 of 890 nm is used for them instead. All images are then transformed by composite transformations $T_1 \circ T_2$, and a mean multispectral frame is computed by averaging corresponding single-wavelength images from the three frames. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

giving rise to imaging artifacts. Signal pre-processing and image post-processing are two parts of the processing pipeline to alleviate the detrimental effect of noise on the image contrast.

The most straightforward way to suppress random noise is averaging of signals from multiple laser pulses [213]. Depending on the number of the averaged pulses, this can be time consuming. Furthermore, the effectiveness of the averaging is hindered by motion occurring between the pulses [213] as well as by temporal jitter of the laser optical build-up [214]. Averaging under such conditions results in blurring.

An alternative to averaging is signal filtering. Simple frequency filtering has limited applicability to optoacoustic signals which are naturally broadband since the signal frequency spectrum is directly related to the target size distribution [213]. To avoid suppression of true signals, an adaptive filter may be applied [215], or other methods based on wavelet transforms or decomposition techniques [216].

Instead of working directly with acoustic signals, noise reduction may also be performed on reconstructed images. Averaging of multiple successive images has been shown to improve the signal-to-noise ratio considerably [47, 217]. In pre-clinical setting, motion clustering and frame rejection have been used to reduce the detrimental effect of periodic motion on image averaging [202, 203]. However, motion patterns in handheld clinical operation are highly irregular and non-periodic, advocating the need for a more flexible approach.

4.2.1 Elastic motion correction and frame averaging

To improve the image contrast and reduce the noise, we propose aggregating several reconstructed, co-registered images together. The advantage of this approach over simple averaging of the recorded signals is that the image frames are first aligned to compensate for the motion occurring between them which avoids motion-induced blurring.

Figure 4.1 presents a schematic overview of the proposed method. Formally, we denote $I^n[\lambda]$ a single-wavelength image in a multispectral frame n (out of N) at wavelength λ (nm). We assume that multispectral frames are acquired at 28 wavelengths between 700 and 970 nm at 10 nm

intervals with illumination frequency of 25 Hz. At this acquisition rate, images $I^n[\lambda]$ and $I^{n+1}[\lambda]$ are acquired 1.12 s apart.

First, intra-frame elastic transformations $T_1^n[\lambda]$ are estimated by minimizing the following objective:

$$f\big(I^n[800], T_1^n(I^n[\lambda])\big). \tag{4.2.1}$$

Here, f denotes some image similarity measure. The images in the wavelength range $\lambda \in [700, \dots, 890]$ are aligned to the image at 800 nm. At these wavelengths, the skin and the blood vessels constitute good landmarks to guide the image registration. At wavelengths over 900 nm, fatty tissue becomes the main source of contrast, complicating alignment with the rest of the stack. Instead of estimating the transformation T_1 for images at wavelengths over 900 nm, we set $T_1^n[\lambda] \triangleq T_1^n[890]$ for $\lambda \in [900, \dots, 970]$.

Second, mean images $\bar{I}^n$ are generated to represent the individual multispectral frames:

$$\bar{I}^n = \frac{1}{20} \sum_{\lambda \in [700,\dots,890]} T_1^n[\lambda](I^n[\lambda]). \tag{4.2.2}$$

Next, inter-frame elastic transformations T_2^n are estimated by aligning mean images $\bar{I}_n$ to $\bar{I}_{\lceil N/2 \rceil}$, where N is the total number of the averaged images. Final images $I^*[\lambda]$ are created by computing the mean of the co-registered images:

$$I^*[\lambda] = \frac{1}{N} \sum_{n \in \mathbb{N}}^{n \leq N} T_2^n \circ T_1^n[\lambda](I^n[\lambda]). \tag{4.2.3}$$

In practice, we use $N = 3$ images. The elastic registration is performed by iteratively optimizing a smooth displacement field to maximize ANTs neighborhood cross-correlation (Advanced Neuroimaging Tools; [218]) between the pair of images being aligned. To handle larger displacements, the algorithm is applied sequentially on ¼, ½, and full-resolution images, using the results of each step as initialization for the next. The algorithm implementation from the SimpleITK library is used [219].

4.2.2 Spectral median adjustment

Additional contrast enhancement and artifact suppression can be achieved by aggregating images obtained at different wavelengths within a narrow band. This is possible thanks to a relatively smooth variation of the absorption spectra of the main endogenous chromophores in the optoacoustic illumination window (see Fig. 2.5).

The main issue addressed by this approach is the wavelength-independent electrical noise exhibited as strong ring-shaped artifacts in the images. Since these artifacts have high intensities, it is beneficial to use the non-linear median function (denoted med) for aggregation instead of the mean, as median is not affected by outlier values.

To generate a spectral-median-adjusted single-wavelength image $\tilde{I}[\lambda]$, we replace the intensity values in each pixel by the median of the intensities of the corresponding pixels in images within a narrow wavelength range. In other words, we apply a median filter on the spectrum in each pixel. In practice, we use a range of 40 nm, corresponding to 5 single-wavelength images in the above-described setup:

$$\tilde{I}[\lambda] = \mathrm{med}(I^*[\lambda - 20], I^*[\lambda - 10], I^*[\lambda], I^*[\lambda + 10], I^*[\lambda + 20]). \tag{4.2.4}$$

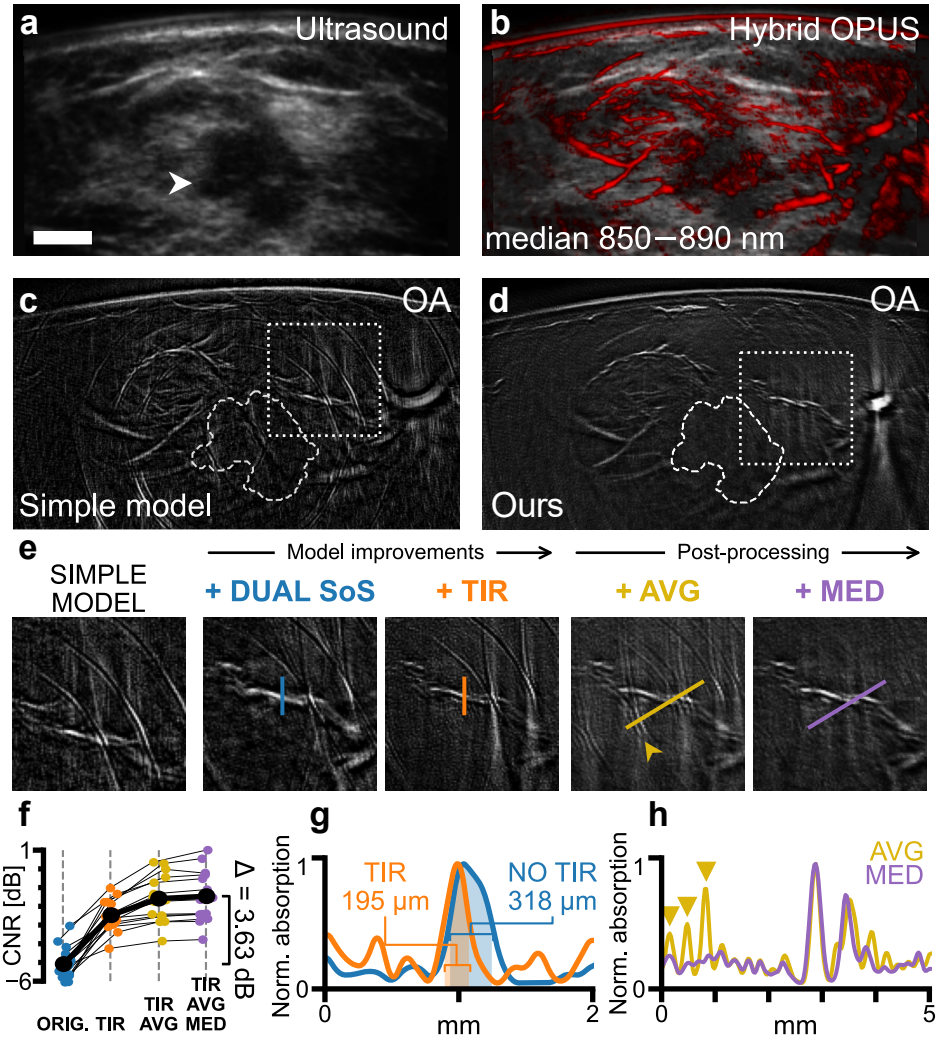

Figure 4.2: Improvements in image quality in 2G-OPUS images demonstrated on a case of invasive lobular carcinoma. a, Grayscale ultrasound image with hypoechoic tumor mass (white arrow). The tumor core is delineated by dashed contours in panels c and d. **b,** Hybrid OPUS visualization providing complementary morphologic and functional information about the lesions by overlaying optoacoustic signal (red, spectral median 850–890 nm) over grayscale ultrasound. **c,** Optoacoustic (OA) image ($\lambda = 880$ nm) reconstructed using a simple, uniform speed-of-sound (SoS) model without total impulse response (TIR) correction. **d,** OA image reconstructed using our proposed pipeline (spectral median 850–890 nm) **e,** Cut-out marked in panels c, d by dotted square showing the cumulative effect of using dual SoS, TIR correction, frame averaging (AVG), and spectral median correction (MED). **f,** Contrast-to-noise ratio (CNR) evaluated in all scans from our breast cancer study (see Chapter 7). Mean CNR of images reconstructed without proposed improvements is –5.09 dB, with TIR –2.47 dB, with TIR and AVG –1.59 dB, and with TIR, AVG, and MED –1.46 dB. **g**, Line profiles (normalized to maximum) of a blood vessel cross-section, marked in the panel e, images 2 and 3. Decrease of full width at half maximum from 318 µm (No TIR) to 195 µm (TIR) indicates improved resolution. **h,** Line profiles (normalized to maximum) along the lines marked in the panel e, images 4 and 5. Peaks marked by yellow arrows correspond to ring-shaped artifacts caused by electrical noise (also marked by yellow arrow in e). Suppression of these peaks on the purple curve shows that the spectral median filter removes this type of noise. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

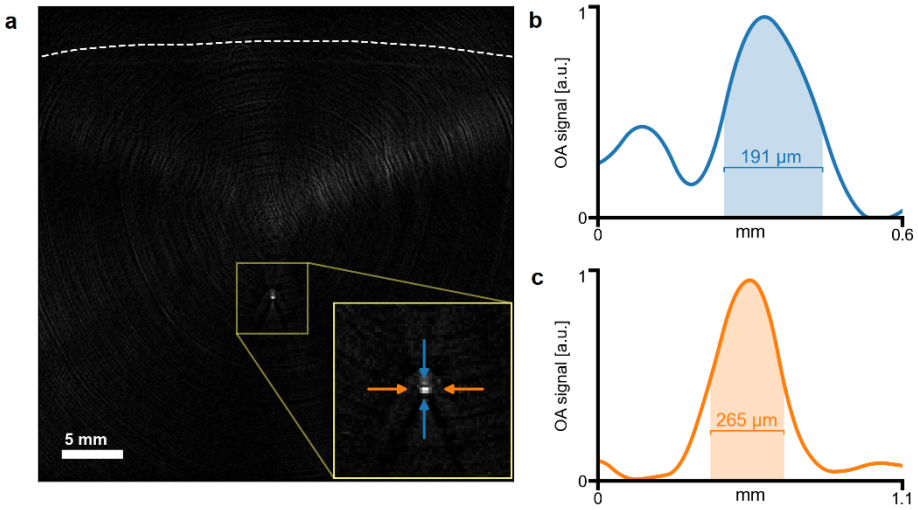

Figure 4.3: Phantom measurement of the attained resolution. a, A 100 µm microsphere was embedded into an agar cylinder and placed in water bath ca. 2 cm below the probe membrane (denoted by the dashed line). Figure shows the image reconstructed from mean signal of nineteen MSOT frames using our model with total impulse response correction and spectral median taken over wavelength range 780–820 nm. A detail of the image containing the microsphere is shown in the inset magnification. Blue and orange arrows mark the ends of axial and lateral line profiles, respectively, which are plotted in panels b and c. **b,** Axial line profile of the imaged microsphere with 191 µm full width at half-maximum (FWHM). **c,** Lateral line profile of the imaged microsphere with 265 µm FWHM. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

4.2.3 Contrast-to-noise ratio computation

To quantify the contrast improvements, we compute contrast-to-noise ratio (CNR) as

$$CNR = 10 \cdot \log_{10} \frac{\mu_{OBJ} - \mu_{BKG}}{\sqrt{\sigma_{OBJ} + \sigma_{BKG}}}, \tag{4.2.5}$$

where μ and σ represent mean and variance, respectively. These statistics are computed within the object (OBJ), defined as the tissue up to 1.5 cm under the skin, and the background (BKG), defined as the area above the skin. When computing the statistics over a whole dataset, we evaluate the CNR on images at 880 nm (or 850–890 nm when considering the spectral median) reconstructed from signals pre-processed with the LCO_{700} band-pass filter (see section 5.1).

4.3 Results

Current handheld optoacoustic systems use simplistic image reconstruction methods [104] that fail to faithfully model their physical and electrical properties, leading to suboptimal resolution, spatial distortions, and imaging artifacts [162]. Our proposed 2G-OPUS image processing pipeline addresses these image quality problems by using a precise forward model with speed-of-sound correction, total impulse response correction, and compounding of motion-corrected frames, delivering high-resolution images with improved contrast, as showcased in Figs. 4.2 and 4.4. Figure 4.2 shows an image from a 75-year-old patient with an invasive lobular carcinoma appearing on the US as a hypoechoic mass (Fig. 4.2a; arrow) approximately 1.5 cm under the surface. The lesion is marked by a dashed line in all subsequent OA images. Fig. 4.2b shows that the US image (grayscale) is greatly enhanced by vascular structures upon superimposing the OA image (in red).

The superimposed OA image is the output of the image formation pipeline developed herein and described above.

We see a marked improvement when comparing the image reconstructed using a simple, uniform SoS model-based inversion (Fig. 4.2c; equivalent up to a non-negativity constraint to the reconstruction used in an earlier work [46]) to the image afforded by applying all the steps of the image formation pipeline proposed herein (Fig. 4.2d). The stepwise improvements achieved by the individual steps of our pipeline are shown in Fig. 4.2e (the displayed region of interest (ROI) is delineated in Fig. 4.2c,d with a dotted rectangle). Each step improves the image—minimizing background noise, the appearance of ring artifacts, and several other distortions—resulting in significantly higher image fidelity. The simple model-based inversion (and other simplistic methods as filtered back-projection) makes use of assumptions on sound propagation and detector properties that do not accurately capture the physical parameters of the experimental measurements. On the other hand, the use of a dual SoS model improves focusing and eliminates spatial distortions, correcting the position of the blood vessel in the middle of the ROI. Furthermore, the incorporation of the TIR, signal averaging, and spectral median correction reduces ring artifacts and improves the resolution, SNR, and CNR.

Fig. 4.2f summarizes the CNR improvements afforded by each step across all 22 images from our breast cancer study (see section 7.1); an overall mean CNR improvement of 3.6 dB is achieved. Marked resolution improvements were observed after TIR correction (Fig. 4.2g), resulting in a sharper appearance of blood vessels and other image features. As demonstrated on phantom measurements (Fig. 4.3), TIR correction can yield image resolutions approaching the limits of the detector hardware, or around 200 μm in the case of *Acuity Echo*®. That is 1.5x improvement over the simple model [176] and more than 2x improvement over the resolution attained by the *Imagio*® scanner [108]. Applying TIR correction also improved CNR by an average of ~2.6 dB.

Frame averaging (3-frames; see section 4.2.1) resulted in a further mean CNR improvement of 0.9 dB, using elastic image registration to reduce blurring due to motion by aligning successive reconstructed images prior to averaging. Ring-shaped artifacts caused by electrical noise (Fig. 4.2e,h; yellow arrows) were suppressed by spectral median processing, whereby every pixel was replaced by the median of its spectrum in a narrow wavelength range (see section 4.2.2). Spectral median correction exploits the premise that the absorption of endogenous chromophores varies smoothly with illumination wavelength, whereas noise appears as peaks at arbitrary wavelengths.

To further demonstrate the superiority of our image processing pipeline, Figure 4.4 provides a visual comparison with the image reconstruction approaches utilized in concurrent works—filtered back-projection and a uniform SoS model—on the eight cases displayed throughout this paper. For examples of image quality achieved by the *Imagio*® scanner, we refer the reader to relevant publications [55, 56, 108, 220].

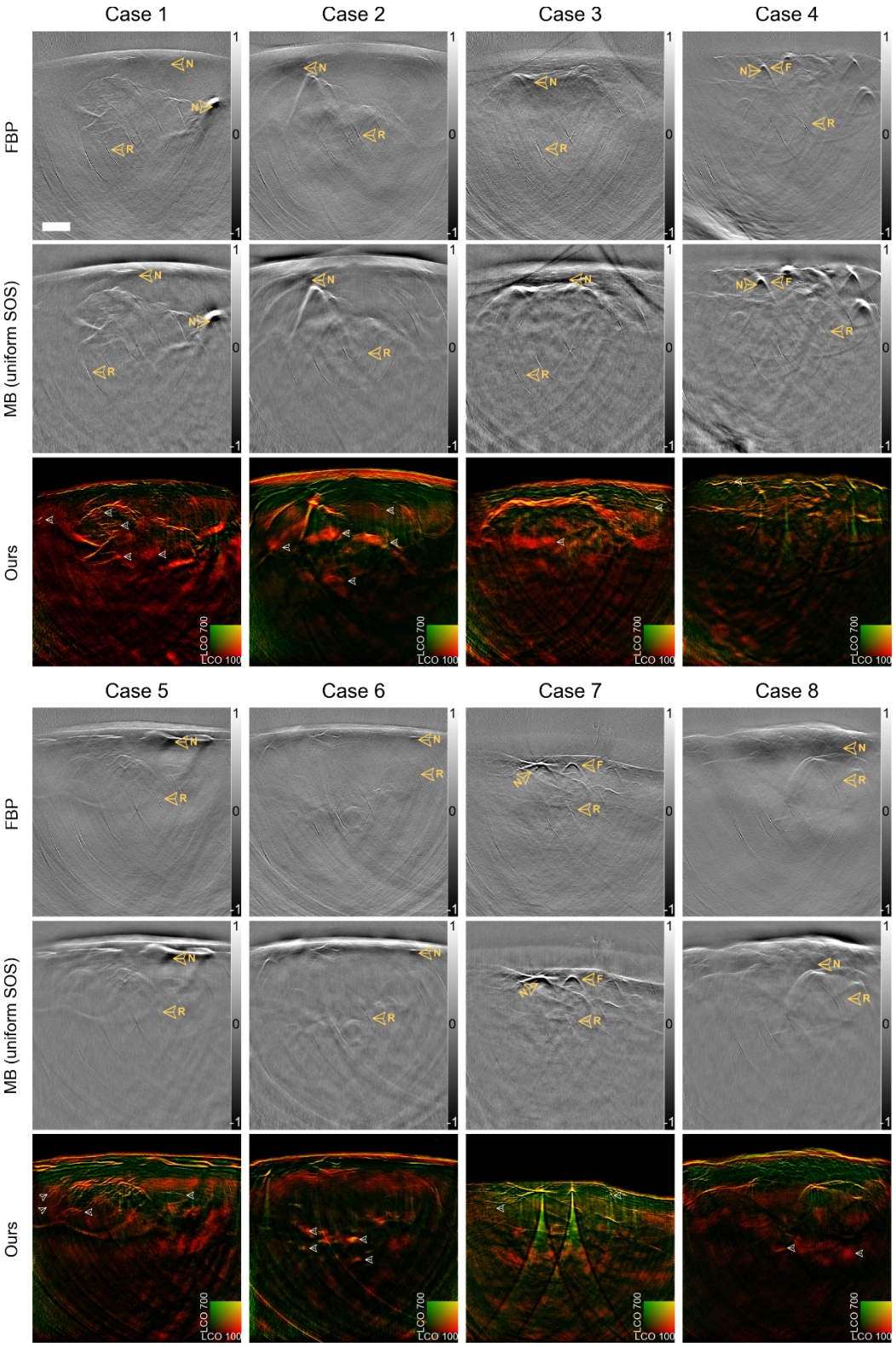

Figure 4.4: Visual comparison of images reconstructed using filtered back-projection (FBP), simple model-based inversion (MB), and our proposed pipeline with an improved model. Yellow arrows in the images denote some examples of image quality issues present in the FBP and simple MB reconstructions: negative values (N), bad focus (F), and ring-shaped noise artifacts (R). Furthermore, white arrows in our proposed images denote examples of fine details that cannot be properly distinguished in the images using FBP and simple MB. These include smaller blood vessels and deeper structures. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

4.4 Discussion

In this chapter, we have introduced a pipeline for processing of multispectral optoacoustic images called 2G-OPUS, which sets a new mark in the image quality of handheld optoacoustic imaging. Application of the proposed pipeline to a dataset of clinical breast cancer scans afforded 3.6 dB improvement in CNR and significant reduction in visible artifacts. Moreover, the combination of the proposed improvements facilitated the ability to resolve blood vessels with diameters as small as 200 μm at depths up to 2 cm (see section 7.1: Case 1—Fig. 7.4 and Case 7—Fig. 7.6). Comparable imaging quality has been so far reported only with dedicated bed-based scanners *SBH-PACT* [48] and *PAM-03* [47], which nevertheless yielded lower resolution (255 μm and 370 μm, resp.). Unlike stationary imaging systems, handheld optoacoustic imaging can be seamlessly integrated in routine breast ultrasound [221-223], improving the information obtained in the imaging session.

Currently, the 2G-OPUS image processing pipeline requires off-line computation which limits the benefits of the achieved image quality improvements during a radiological examination. Providing high-quality OA images to the operator in real-time would enable better localization of suspicious features and increase the efficacy of handheld OA examinations. Current efforts for acceleration of image reconstruction using deep learning show a promise towards achieving this goal [224].

CHAPTER 5

Image visualization

Optoacoustic tomography is—at present—primarily a qualitative modality. Until further advances in quantitative OAT enable precise measurements of chromophore concentrations, evaluation of OA images relies on detection and interpretation of visual patterns—*features*—by an expert observer. Detection of OA features depends crucially on an appropriate way of visualizing OA images, which is by no means a trivial task. An MSOT scan contains a huge amount of information in its spatial, spectral, and temporal dimensions. On top of that, the scan is accompanied by a stream of US images. A proper visualization needs to provide the right combination of information relevant for a certain clinical task and suppress the irrelevant scan content.

In the first part of this chapter, we present the *dual-band visualization* method to properly display features on different scales together. Standard visualization methods for OA imaging are biased towards larger features. The size of the imaged features is inversely proportional to the frequency of OA signals they generate: large objects emit low-frequency waves and small objects emit high-frequency waves [109], which are more strongly attenuated in the tissue [225]. Furthermore, since the pressure change is proportional to the total absorbed energy, which is proportional to the size of the absorber, the signal from smaller absorbers is weaker. Overall, the visibility of small absorbers in the image tends to be compromised in favor of large structures. The dual-band visualization utilizes two different band-pass filters on the OA signals to obtain two different OA images highlighting larger and smaller feature scales. The two images are then combined using a two-channel colormap, yielding a balanced representation of small- and large-scale features and offering a visual separation of tissue features from noise and out-of-plane signals.

In the second part of this chapter, we present *local contrast normalization* and other non-linear image filtering methods addressing the issue of an extremely skewed pixel intensity distribution in OAT images. The distribution of pixel intensities is skewed (Fig. 5.4) due to the light fluence decaying exponentially with depth. This long-tailed value range requires re-mapping into a 256-value range available on a computer screen to facilitate displaying relevant image features. Importantly, since the relevant local information on the absorption coefficient is convolved with the light fluence (primarily indicative of the depth), similar features may appear on completely different points on the intensity scale. Equalization of the visual appearance of similar image features is needed, yet intensity normalization methods applied globally, such as the power law normalization or histogram equalization, fail to achieve that goal. On the other hand, methods normalizing the contrast locally can adaptively enhance the dynamic range of the image and enable feature visualization across the whole intensity distribution. Since the optimal parameters of the presented filters vary across images, we develop an interactive user interface to aid in tuning the image appearance.

This chapter is partially based on the following publication whose parts appear verbatim without further notice: Kukačka, Jan, Stephan Metz, Christoph Dehner, Alexander Muckenhuber, Korbinian Paul-Yuan, Angelos Karlas, Eva Maria Fallenberg, Ernst Rummeny, Dominik Jüstel, and Vasilis Ntziachristos. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.

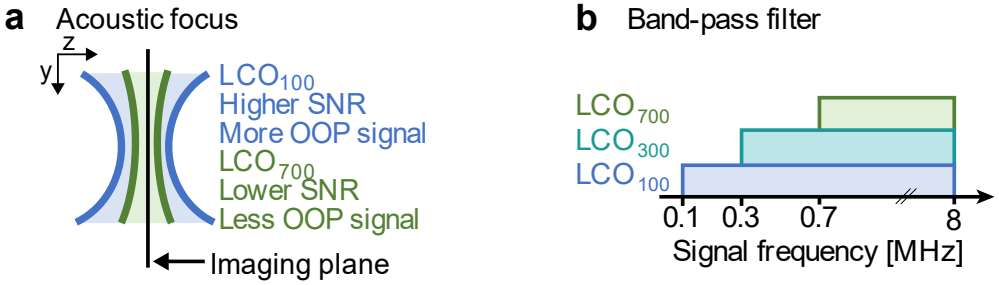

Figure 5.1: Frequency-dependent transducer sensitivity and band-pass filters. a, Schematic visualization of the acoustic sensitivity of broadband transducers. Higher frequencies, remaining in the signal after applying the LCO_{700} filter (shown in green), offer better definition of the imaging plane and are less affected by OOP signals, but suffer from reduced SNR. Lower frequencies, preserved by the LCO_{100} filter (shown in blue), correspond to a broader sensitivity field, resulting in more OOP signals. **b,** Schematic visualization of the frequency bands allowed through the three filters LCO100, LCO300, and LCO700. Whereas the upper cut-off is equal for all (8 MHz), the lower cut-offs are 100 kHz, 300 kHz, and 700 kHz, respectively. **Abbreviations:** LCO, lower cut-off; SNR, signal-to-noise ratio; OOP, out-of-plane. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

5.1 Dual-band visualization

Current image visualization methods for OAT do not adequately handle the broadband nature of OA signals and the spatial sensitivity of the acoustic detectors. The acoustic signals produced in response to pulsed-light illumination are broadband, since optical absorbers in tissue emit ultrasound pulses at wavelengths proportional to their size [109]. Small absorbers produce waves at higher frequencies which are predominantly attenuated in tissue [225]. Moreover, the magnitude of the optoacoustic pulses is proportional to the absorbed energy, which in turn grows with the absorber size. Smaller absorbers thus produce signals closer to the noise level and are harder to properly visualize along larger objects producing stronger signals.

Furthermore, the sensitivity field of focused linear detector arrays is frequency dependent, as shown in Fig. 5.1a. The relationship between the breadth of the sensitivity field around the imaging plane and the ultrasound frequency is inversely proportional: the sensitivity field grows as the sound frequency decreases [97]. Lower frequency parts of the OA signal contain thus larger contribution from *out-of-plane* (OOP) absorbers which cause image blurring and may be misinterpreted for in-plane features.

A potential way for removing the bias towards larger structures is applying a high-pass filter. A term representing such filter is indeed employed in the commonly used universal back-projection algorithm (cf. equation (2.6.4)). Another approach, utilized in optoacoustic mesoscopy and microscopy, is separation of the signals into disjoint frequency bands. For each band, a separate image is reconstructed, and a final compound image is formed using separate colors for different frequency bands [99, 226, 227]. However, removing parts of the frequency content from the signal yields data incompatible with the OA physical model. As a result, artifacts are present in images reconstructed from single-band data. A principled approach to disentangling the frequency content of OAT scans and reconstructing features on different scales in separate images has been recently proposed [228]. Resulting images show small features in greater detail than standard model-based reconstruction, whereas the reconstruction residual errors are lower compared to reconstructing individual frequency bands.

Suppression of OOP signals has also been attempted in several works. A method based on comparing images acquired with varying axial position of the transducer array has shown the ability to remove some of the OOP artifacts [229]. Instead of changing the axial position, the elevation of the detector can be modified to acquire multiple neighboring slices. An appropriate model-based method allows reconstructing those slices together while decomposing the signal contributions to the appropriate slices [230]. The elevation can also be altered via free-hand operation and a 3D volume can be reconstructed from the individual slices [231]. In such case, the OOP signals support the estimation of the unknown motion trajectory. The need for a dedicated hardware implementation and the increased scanning time—aggravating the sensitivity to motion—are limiting a broader applicability of these methods.

5.1.1 Method

There are clearly two contradictory objectives to be satisfied: Better acoustic focus on the imaging plane can be achieved by excluding the low frequencies and thus suppressing the OOP signals. This also leads to emphasizing smaller image features that would be otherwise obscured by stronger signals from bulky tissue. However, the OA signals are broadband by their nature and filtering makes the signal incompatible with the physical model. Removing low frequencies thus introduces artifacts and results in noisier images. Instead of finding a compromise by finding some "optimal" filtering threshold, we propose to reconstruct two different variants of the image and combine them to get the best of both worlds.

To that end, we process the recorded acoustic signals using the Butterworth band-pass filter. The higher cut-off value is set to match the transducer sensitivity (8 MHz for the iThera *Acuity Echo* scanner used in this experiment); for the lower cut-off (LCO) value we use three increasing thresholds (Fig. 5.1b): 100 kHz (LCO_{100}), 300 kHz (LCO_{300}), and 700 kHz (LCO_{700}). The order of the filter is set to 8$^{\mathrm{th}}$ order on the low-pass edge and 2$^{\mathrm{nd}}$ order on the high-pass edge.

Using the signals processed with the above defined filters, we reconstruct three variants of the images. The LCO_{100} version contains most of the signal and thus is less noisy but more affected by OOP absorbers. The LCO_{300} version represents a trade-off between less and more aggressive filtering. The LCO_{700} variant emphasizes small structures but inevitably also noise. Moreover, it introduces some artifacts stemming from violation of the physical model by discarding valid OA signals.

Finally, we generate the dual-band images by combining the LCO_{100} and LCO_{700} image variants using a two-channel color mapping. The color mapping is performed by applying the following procedure to every pixel.

Two-channel color mapping
Let $\boldsymbol{x} \equiv (x_1, x_2)$ be a tuple representing the intensities in one pixel of a two-channel image with each channel normalized between 0 and 1: $\boldsymbol{x} \in [0,1]^2$. The color mapping procedure converts $\boldsymbol{x}$ to a color value represented as a three-dimensional vector in the RGB color space[9] using a color mapping function $f_{\mathrm{cm}}\colon [0,1]^2 \mapsto [0,1]^3$. The output colors are controlled by a colormap. Let the colormap be defined by four anchor colors corresponding to the corners of the unit square. It is represented by a matrix $\boldsymbol{C} \in [0,1]^{3\times 4}$, whose columns are the anchor color vectors corresponding to the corners in the order $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. The output of the color mapping function is defined as a convex combination of the anchor colors, where the coefficients are given by an interpolation function $b\colon [0,1]^2 \mapsto [0,1]^4$:

[9] The method is also applicable to other color spaces than RGB. Other commonly used color spaces are HSV (hue–saturation–value), or perceptually more uniform CIELAB, $\mathrm{CIELCh}_{\mathrm{uv}}$.

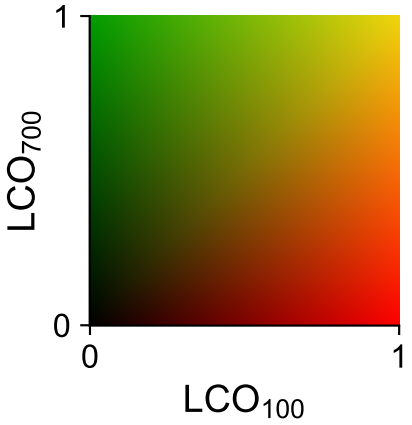

Figure 5.2: Dual-band visualization colormap. LCO_{100} and LCO_{700} represents images reconstructed using band-pass filters with lower cut-off frequencies set to 100 kHz and 700 kHz, respectively. Values from the LCO_{100} channel are plotted on the x-axis, values from the LCO_{700} channel are plotted on the y-axis. **Abbreviations:** LCO, lower cut-off.

$$f_{\mathrm{cm}}(\boldsymbol{x}; \boldsymbol{C}) = \boldsymbol{C} b(\boldsymbol{x}). \tag{5.1.1}$$

The interpolation function b maps the vector $\boldsymbol{x}$ to the coefficients by computing its barycentric coordinates relative to triangles $\triangle\, t_1 := (0,0), (0,1), (1,1)$ and $\triangle\, t_2 := (0,0), (1,0), (1,1)$ in the following way:

$$b(\boldsymbol{x}) = \begin{cases} \left(\lambda_1^{(t_1)}, \lambda_2^{(t_1)}, 0, \lambda_3^{(t_1)}\right)^T & \text{if } \boldsymbol{x} \text{ in } \triangle\, t_1 \ , \\ \left(\lambda_1^{(t_2)}, 0, \lambda_2^{(t_2)}, \lambda_3^{(t_2)}\right)^T & \text{if } \boldsymbol{x} \text{ in } \triangle\, t_2 \ , \end{cases} \tag{5.1.2}$$

where $\lambda_n^{(t)}$ represents the n-th barycentric coordinate w.r.t. a triangle $\triangle\, t := (a_1, a_2), (b_1, b_2), (c_1, c_2)$. For points on the diagonal, where both conditions are satisfied, the expressions are equal. The barycentric coordinates w.r.t. $\triangle\, t$ of a point (x_1, x_2) are obtained by solving the following linear system:

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}. \tag{5.1.3}$$

The third row of the above system ensures that $\sum_{i=1}^{3} \lambda_i = 1$ and the barycentric coordinates are unique.

The dual-band visualization uses the following colormap matrix, where the columns are colored to match the values which they represent in the RGB space:

$$\boldsymbol{C}_{\mathrm{db}} = \begin{pmatrix} 0 & 1 & 0 & 0.945 \\ 0 & 0 & 0.613 & 0.836 \\ 0 & 0 & 0 & 0.051 \end{pmatrix}. \tag{5.1.4}$$

Figure 5.2 visualizes the full unit square mapped using the $\boldsymbol{C}_{\mathrm{db}}$ colormap.

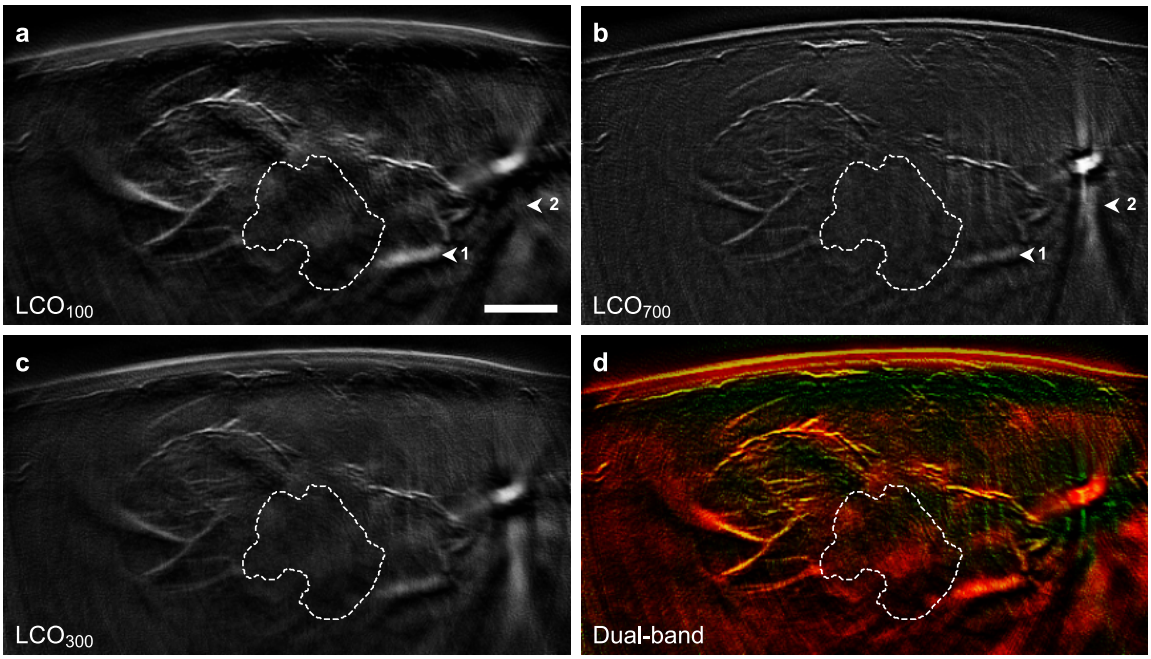

Figure 5.3: Effect of signal filtering on visualization of information contained in 2G-OPUS images demonstrated on a case of invasive lobular carcinoma. a–c, OA images reconstructed from signals filtered with band-pass filters with lower cut-off frequencies set to 100 kHz (LCO_{100}), 700 kHz (LCO_{700}), and 300 kHz (LCO_{300}), respectively. Arrows highlight differences between LCO_{100} (a) and LCO_{700} (b); arrow 1 marks an out-of-plane large blood vessel that is suppressed by LCO_{700}, while arrow 2 marks streak artifacts caused by limited view that are prominent in the LCO_{700} variant. **d,** Dual-band visualization, obtained by combining image variants LCO_{100} (a) and LCO_{700} (b) into a single image using a colormap shown in Fig. 5.2. All images show spectral median of 850–890 nm range. Dashed lines outline the tumor mass. The scalebar (a) represents 5 mm. **Abbreviations:** 2G-OPUS, second generation optoacoustic-ultrasound tomography; OA, optoacoustic; LCO, lower cut-off. Figure panels reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

5.1.2 Results

We present the benefits of using the proposed dual-band visualization on an example scan from our breast cancer study (see Chapter 7). Fig. 5.3 shows an image of an invasive lobular carcinoma (dashed line) from a 75-year-old patient located approximately 1.5 cm under the skin surface. Figs. 5.3a and 5.3b show OA images reconstructed from signals processed with band-pass filters LCO_{100} and LCO_{700}, respectively, differing in their lower cut-off thresholds. From the two images, the LCO_{100} variant exhibits noticeably less background noise and allows the reconstruction of larger-scale features. There is a higher contribution from OOP signals such as the larger blood vessel marked by the arrow 1. Moreover, the stronger low-frequency signals obscure the fine details visible in the LCO_{700} variant (Fig. 5.3b). The LCO_{700} variant appears crisper but exhibits more streak artifacts (arrow 2). Choosing a single cut-off level is thus a necessary compromise between these two aspects—Fig. 5.3c shows an image reconstructed using the LCO_{300} filter. It contains the drawbacks of both previous variants, albeit not as severe.

Fig. 5.3d shows the dual-band visualization of the same image, formed as a combination of the LCO_{100} and LCO_{700} variants displayed in Figs. 5.3a and 5.3b, respectively. The dual-band representation allows distinction of fine structures (yellow) from artifacts and noise (green), because the fine structures are present in both images, but artifacts stemming from the frequency filtering are only in LCO_{700}. This is an advantage over using disjoint frequency bands. Larger structures appear in red, hinting at possible OOP signals. Overall, the dual-band visualization

provides a better way of displaying OA information within the ultrasound bandwidth recorded by 2G-OPUS compared to using a single frequency band.

5.2 Local contrast normalization

Optoacoustic tomography images have an extremely broad intensity range which cannot be faithfully displayed on a computer screen. The MSOT scanner *Acuity Echo* (iThera Medical, Munich, Germany), employed in our research, records the acoustic signals with 12-bit resolution, yet the resolution of the human eye in the brightness range of computer screens is less than 10-bit [232]. The main reason for the broad range of intensity values is apparent from the eq. (2.5.15)—the initial pressure, which is recovered by standard reconstruction algorithms, is a product of the absorption coefficient and the light fluence. Whereas the distribution of the absorption coefficients is independent from the depth (up to the depth-dependent distribution of the biological tissue), the light fluence decays exponentially with the depth. The product of the two is also exponentially distributed. The implications are twofold: the image intensity distribution is extremely skewed towards small values and has a long tail of very high values. Fig. 5.4 shows the breast tumor image from Fig. 5.3a without any contrast adjustments and a distribution of its pixel intensities. Since most of the pixel intensities are concentrated at the bottom 10% of the value range, where the sensitivity of the human eye is lowest [233], most of the image details are imperceptible to an observer. To equalize the visual appearance of similar image features across the whole intensity range, a contrast normalization procedure is needed.

The simplest method for brightness adjustment is linear scaling to a narrower range (Fig. 5.5a). A drawback is that the values outside of the range are saturated. Linear scaling can be combined with a non-linear function, such as log or power law, to linearize the value range before adjusting the endpoints. An example of log-normalization is shown in Fig. 5.5b. Choosing the right parameters can be done manually or heuristically using statistics of the intensity distribution, such as certain quantiles. Another approach based on the image statistics is histogram equalization (Fig. 5.5c), where the distribution of the values is remapped to a uniform distribution [234]. A shared disadvantage of the named methods is that they operate on the image globally, i.e., the same intensity transformation is applied to all pixels. As such, they cannot avoid the value range compression and the consequent loss of details.

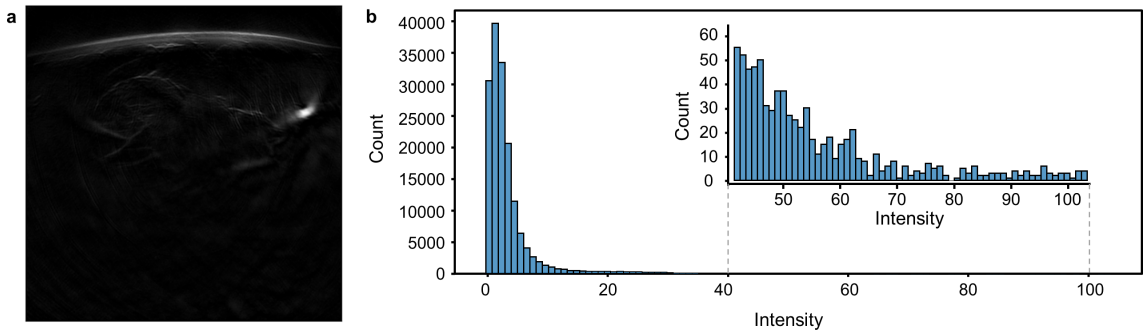

Figure 5.4: Distribution of intensity values in an unnormalized OAT image. a, An OAT image of a breast tumor before applying intensity normalization procedure (the normalized image is shown in Fig. 5.3a). **b,** A histogram of the intensity distribution of the 160 000 pixels forming the OAT image show in (a). Most pixels have intensities below 10, but a long tail of pixels with higher intensities exists (inset). **Abbreviations:** OAT, optoacoustic tomography.

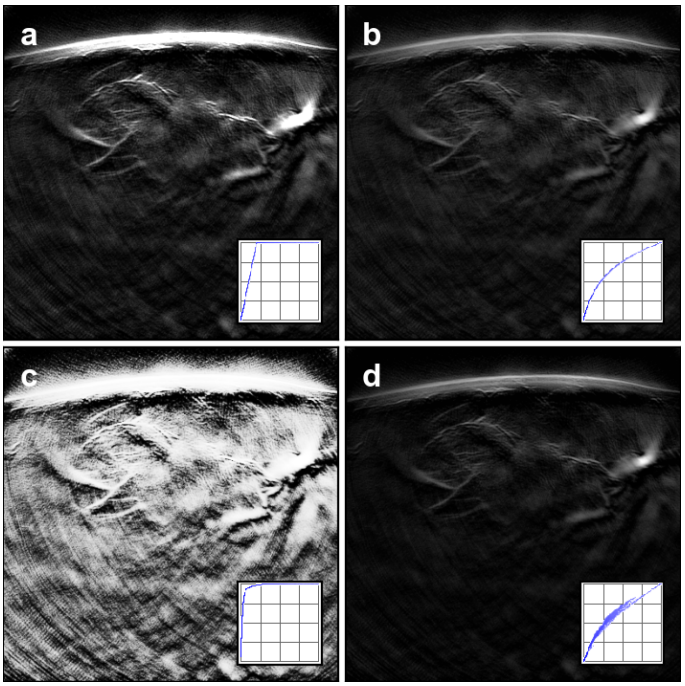

Figure 5.5: Intensity normalization techniques demonstrated on an image of a breast tumor. a, Linear scaling with cut-off at 98th percentile. **b,** Logarithmic scaling. **c,** Histogram equalization. **d,** Adaptive histogram equalization. Inset axes show the intensity mapping for each technique.

Alternative to global contrast adjustments are methods that operate locally, such as adaptive histogram equalization [235-237], shown in Fig. 5.5d. These methods apply a different transformation to each pixel depending on its neighborhood. Adaptive methods are not bijective and the correspondence between the normalized values and the original values is lost. Since the new intensity values are detached from the original reconstructed values, images processed with adaptive methods cannot be used for any meaningful quantitative analysis. In exchange for invertibility, adaptive methods provide *dynamic contrast*, since the available value range can be used to display different sub-ranges of the original range as needed to deliver good contrast in all parts of the image. A drawback of adaptive histogram equalization is the potential amplification of noise in regions without any other features. Contrast-limited variant of the algorithm (CLAHE) has been proposed to limit the noise amplification [237]. It has been applied for contrast enhancement in pre-clinical MSOT [238].

The above-named methods are standard image processing procedures that disregard the physical origin of the contrast imbalance—the varying light fluence. On the other hand, numerous methods have been proposed for normalization of the light fluence in OAT. Elimination of the light fluence component from the OA images would ensure improved contrast, yet the problem has been notoriously difficult to solve. Park *et al.* have utilized a simple fluence model based on the Beer-Lambert law[‡] with adaptive estimation of the effective attenuation coefficient in 3D OAT [239]. They have demonstrated superior results to CLAHE on both simulated and experimental images of breast tissue. The accuracy of the method is limited by the assumption of optically homogenous tissue. Furthermore, over-amplification of noise in deeper regions of the image is also an issue. Another simple adaptive method for compensation of depth-dependent effects of light fluence and ultrasound attenuation has been utilized by the *Imagio*® handheld system [108]: Strong absorbers such as blood vessels and solid tumors are first segmented and excluded from the computation of the normalization factor. Then, each pixel intensity is divided by the mean of the pixel intensities in its row. Unlike the previous simplified physical model, this approach can adapt to non-uniform optical properties in the tissue, such as layers of subcutaneous fat over muscle tissue, yet it still leads to amplification of noise and may be sensitive to signal variations between image rows. Using

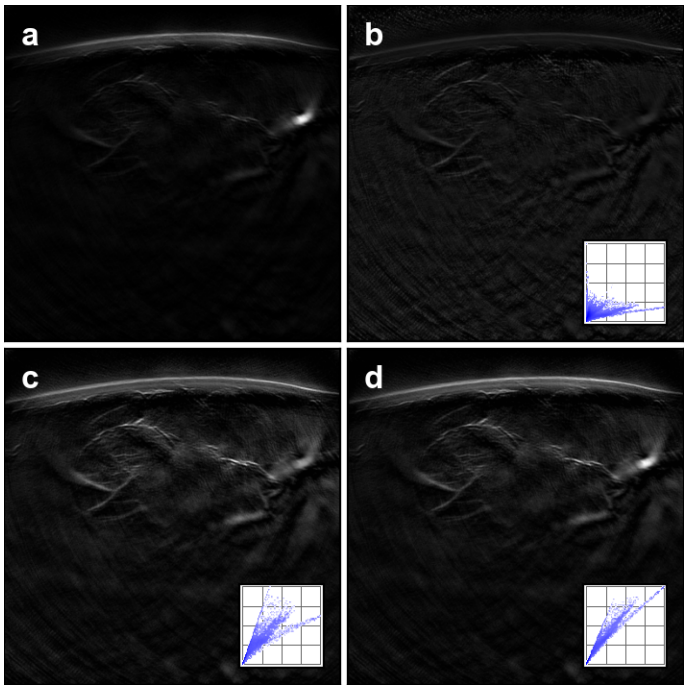

Figure 5.6: Variants of local contrast normalization. a, Unnormalized OA image of a breast tumor. **b,** Normalization I_1 with Gaussian kernel size $s = 7$. **c,** Normalization I_2 with Gaussian kernel size $s = 7$ and cut-off parameter $\alpha = 0.22$. **d,** Local contrast normalization ($s = 7$, $\alpha = 0.22$). Inset axes show the histogram of the intensity mapping of the respective techniques.

a tissue model based on segmentation of a speed-of-sound map obtained from US data and a-priori known tissue properties, Pattyn *et al.* have simulated the light fluence with Monte Carlo (MC) approach [240]. Their method has the promise to account for heterogenous composition of real tissue, but its performance has only been validated on simple phantoms. Moreover, MC methods are notoriously computationally intensive and prohibit real-time usage. Besides estimating the light fluence directly, iterative methods have been proposed as well. Cox et al. have developed a model-based fluence estimation utilizing the diffusion model of light transport [241]. This approach can be extended by using the more precise radiative transfer light transport model instead of the diffusion approximation [242], or incorporation of tissue segmentation into a regularizer to constrain optically homogenous regions [243]. A regularization parameter can also be used to control the noise amplification in low-fluence regions. To alleviate the noise amplification issue, Bu *et al.* have incorporated a precomputed fluence estimate directly into an iterative model-based image reconstruction procedure [187]. Their fluence map was computed via an MC simulation relying on assumed homogenous optical properties. Finally, a body of work has been developed exploiting the fact that the light fluence distribution is smooth and thus resides in a low-dimensional space which can be modelled using few basis functions. Rosenthal *et al.* have modelled the fluence as a sparse vector in the Fourier basis, avoiding errors from invalid light propagation modelling [244]. On the other hand, Tzoumas *et al.* have used simulations to identify the basis (eigen)spectra that suffice to accurately model the fluence in depth [192]. Overall, the existing fluence correction methods have shown impressive results considering the difficulty of the problem. However, they are still computationally demanding, require system-specific modelling, rely on a-priory assumed tissue properties, or suffer from noise amplification in the low-fluence region.

5.2.1 Method

Here, we introduce *local contrast normalization* (LCN)—a method for contrast adjustment in OA images for optimal visualization of features to facilitate qualitative analysis. Without the ambition to provide quantitative information, our method does not require any prior assumptions on the

tissue properties nor light propagation modelling. This leaves the proposed method simple, fast, and universally applicable.

As explained above, methods applying the same intensity transformation across the whole image are unsuitable for contrast enhancement in OA images and a localized adjustment is needed. The local adjustment can be based on the light fluence estimate or the neighborhood intensities. In fact, one can interpret the neighborhood-based adjustment used in adaptive methods as a form of local fluence estimation. In the proposed LCN, we normalize the intensities by dividing each pixel by a weighted average of its neighborhood. The weights are given by a Gaussian of the L_2-distances from the normalized pixel. This operation can be efficiently computed by dividing the image by itself blurred by a Gaussian kernel.

Formally, we denote the image as a function defined on a set of pixels in the 2D space, $I\colon \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R}$. Furthermore, let G_s be a Gaussian kernel with the scale s. Then the above-described normalization scheme produces a new image, I_1:

$$I_1 = \frac{I}{I * G_s} \,. \tag{5.2.1}$$

Fig. 5.6b shows an example of the I_1 normalization. It is apparent that I_1 suffers from noise amplification in regions of the image where the intensities are low and the denominator is small. To avoid that, we clip the minimum value of the denominator at a cut-off value given as a fraction of the supremum of the denominator:

$$I_2 = \frac{I}{\max(I * G_s, \alpha \cdot \sup(I * G_s))} \,. \tag{5.2.2}$$

The value of α acts as a regularization parameter controlling the noise amplification. With $\alpha = 0$, the eq. (5.2.2) reduces to eq. (5.2.1) and the noise amplification is not suppressed. At $\alpha = 1$, the whole image is divided by a constant value and no adaptive normalization is performed.

Fig. 5.6c shows an example of the I_2 normalization. One drawback of the I_2 normalization scheme is that high-intensity features, often corresponding to important blood vessels, have a very high denominator and may be overly suppressed in the normalized image. This is apparent from the mapping distribution shown in the inset of Fig. 5.6c: the highest values in the original image are mapped below 75% of the maximum in the normalized image. To avoid that, we average the original and the normalized image. This way, the high-intensity features remain unchanged but low-intensity features are emphasized. Overall, our proposed LCN is defined as:

$$I_{\text{lcn}} = \frac{1}{2}\left(\frac{I}{\sup(I)} + \frac{I}{\max(I * G_s, \alpha \cdot \sup(I * G_s))}\right). \tag{5.2.3}$$

Fig. 5.6d shows the effect of LCN applied to a breast tumor image (cf. other methods in Fig. 5.5). A Python implementation of LCN is shown in Listing 5.1.

Besides LCN, we utilize two additional transformations for visualization of OAT images: sigmoid normalization and unsharp masking.

Sigmoid normalization (SN) is a flexible, non-linear intensity adjustment exploiting the S-shape of the logistic sigmoid function, $\sigma(x) = 1/(1 + e^{-x})$. The flexibility stems from the fact that before applying the logistic sigmoid, the input intensities are rescaled to ensure that the input values from a range (x_1, x_2) will be mapped to a range (y_1, y_2). That means that the values from the range $(-\infty, x_1)$ will be compressed to the range $(0, y_1)$ and the values from the range (x_2, ∞) will be compressed to the range $(y_2, 1)$ by extension of the sigmoid. Fig. 5.7 demonstrates on four example

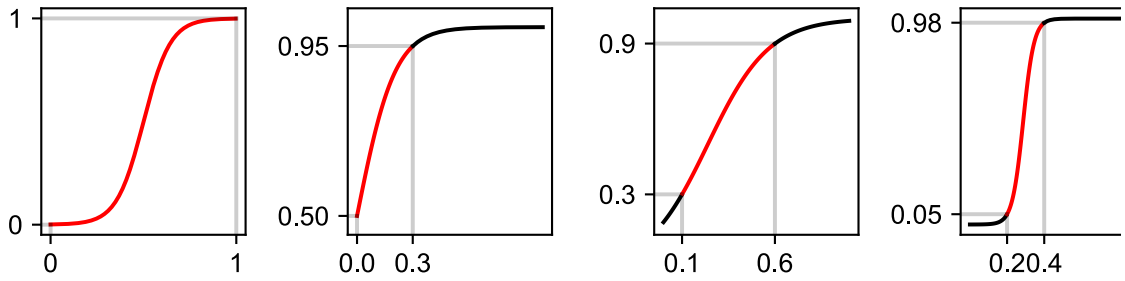

Figure 5.7: Sigmoid normalization examples. Depending on the selection of the control points (marked on the axes), sigmoid normalization can represent a whole range of non-linear functions. The range between points x_1 and x_2 is mapped to the part of the sigmoid corresponding to the output range between y_1 and y_2 (highlighted red). Input values outside the (x_1, x_2) range are mapped to the extension of the sigmoid outside (y_1, y_2).

values of the control points that the rescaling effectively gives the user control over which part of the S-shape will be applied to the data. Formally, SN applies the following function:

$$f(x; x_1, x_2, y_1, y_2) = \sigma\left(\left(\sigma^{-1}(y_2) - \sigma^{-1}(y_1)\right)\frac{(x - x_1)}{x_2 - x_1} + \sigma^{-1}(y_1)\right). \tag{5.2.4}$$

Here, σ^{-1} represents the inverse sigmoid function, defined as

$$\sigma^{-1}(x) = \ln\left(\varepsilon + \frac{x}{1 - x + \varepsilon}\right), \tag{5.2.5}$$

where $\varepsilon = 10^{-3}$ is added for numerical stability. A Python implementation of SN is shown in Listing 5.2.

Unsharp masking (UM) is a standard method for sharpening the details in the image. It is defined as:

Listing 5.1: Python implementation of the local contrast normalization.

```
1. def local_contrast_norm(img, kernel_size, cutoff_fraction):
2.     ## Compute input range
3.     img_min, img_max = img.min(), img.max()
4.
5.     norm = gaussian_filter(img, kernel_size)
6.     cutoff = np.max(norm) * cutoff_fraction
7.     norm_img = img / np.maximum(norm, cutoff)
8.     norm_img = np.nan_to_num(norm_img)
9.
10.    ## Ensure norm_img has same scale as input to enable averaging
11.    norm_min, norm_max = norm_img.min(), norm_img.max()
12.    scale = (img_max-img_min)/(norm_max-norm_min)
13.    norm_img = img_min + (norm_img-norm_min)*scale
14.    result = img+norm_img
15.
16.
17.    ## Ensure output has same range as input
18.    res_min, res_max = result.min(), result.max()
19.    scale = (img_max-img_min)/(res_max-res_min)
20.    result = img_min + (result-res_min)*scale
21.
22.    return result
```

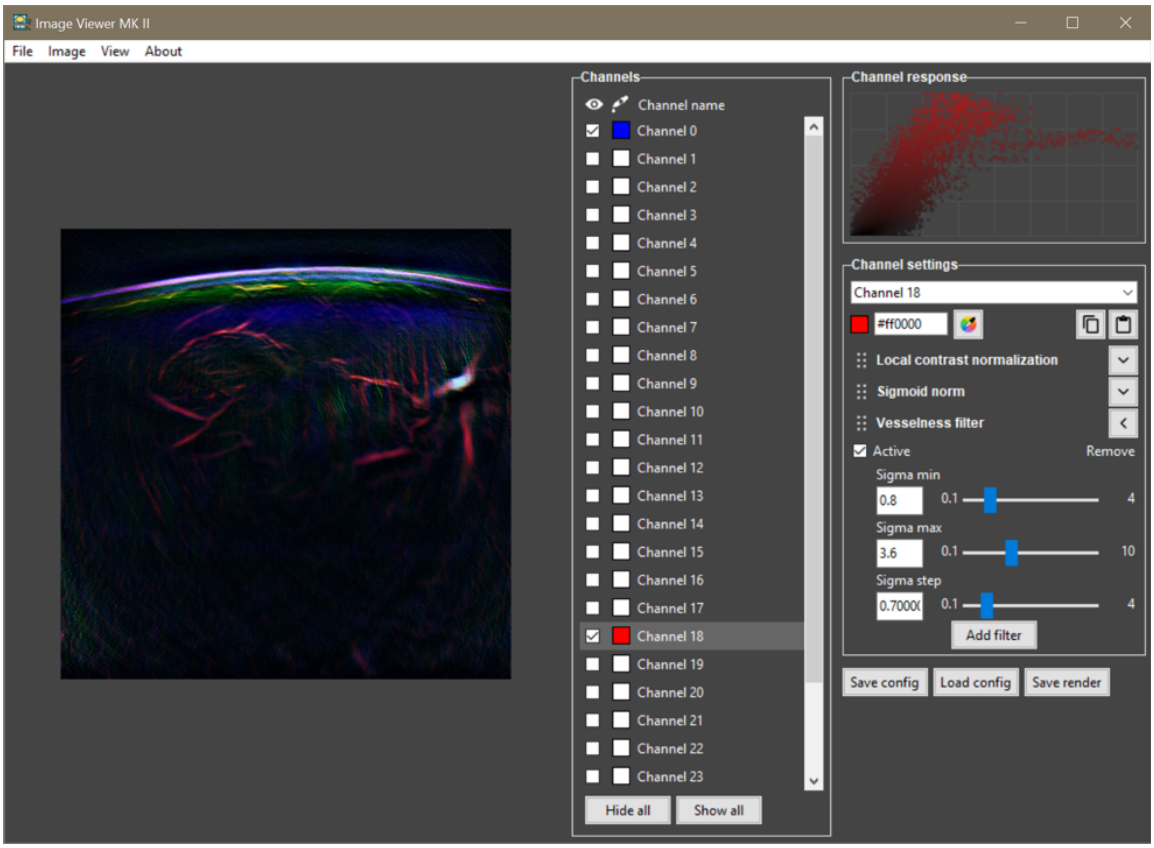

Figure 5.8: Screenshot of the interactive image viewer for manual tuning of contrast enhancement parameters.

$$I_{\text{um}} = I + \beta(I - G_s * I). \tag{5.2.6}$$

Here, β controls the strength of the sharpening effect and the scale s of the Gaussian kernel controls the size of the sharpened features.

For optimal performance, the values of parameters α, x_1, x_2, y_1, y_2, β, and kernel sizes s need to be selected manually for each image. In case of multi-channel images, these transformations are applied to each channel individually. We have implemented an interactive image viewer for selecting the transformation parameters and made it available in a public code repository.[10] A screenshot of the user interface is shown in Fig. 5.8.

Listing 5.2: Python implementation of the sigmoid normalization.

```
def sigmoid_norm(img, x1, x2, y1, y2):
    img_min, img_max = img.min(), img.max()
    eps = 1e-3

    new_low = np.log(eps + y1/(1-y1))  # eps to avoid log(0)
    new_high = np.log(y2/(1-y2+eps))   # eps to avoid division by 0
    norm_img = (new_high-new_low) * (img-x1)/(x2-x1+eps) + new_low
    norm_img = 1/(1+np.exp(-norm_img))

    norm_min, norm_max = norm_img.min(), norm_img.max()
    scale = (img_max-img_min) / (norm_max-norm_min)
    return img_min + (norm_img-norm_min) * scale
```

[10] www.github.com/jankukacka/image_viewer_mk2

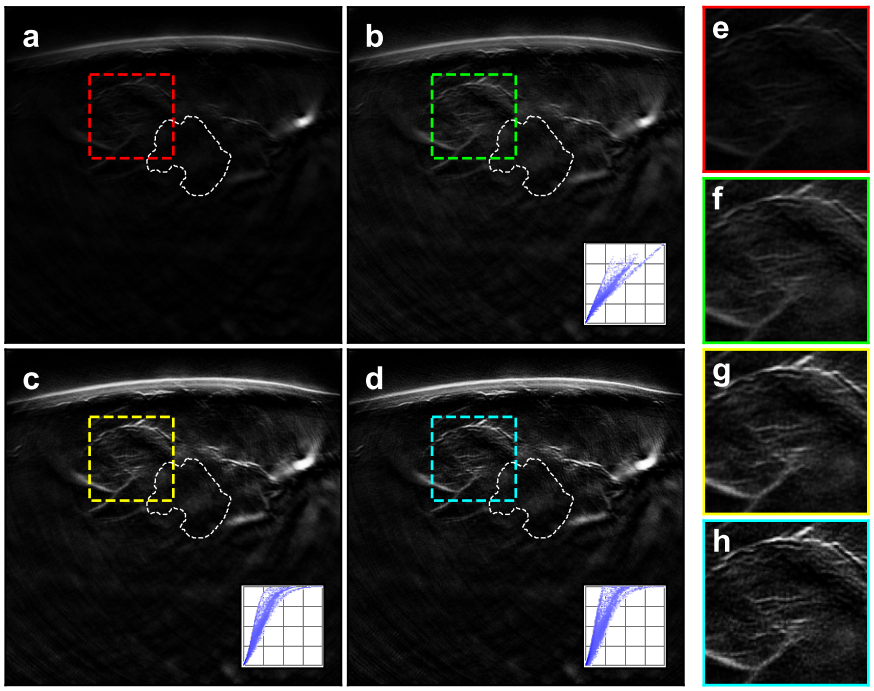

Figure 5.9: Demonstration of proposed contrast enhancement methods on a scan of breast tumor. **a,** Unnormalized image. **b,** Local contrast normalization (LCN; $s = 7$, $\alpha = 0.22$). **c,** LCN as in (b) with sigmoid normalization on top (SN; $x_1 = 0.06$, $x_2 = 1$, $y_1 = 0.145$, $y_2 = 0.995$). **d,** LCN and SN as in (c) with unsharp masking ($\beta = 1.5$, $s = 0.5$). **e–h,** enlarged cut-outs of images shown in (a)–(d) (marked by dashed squares of the respective color) showing details of small peritumoral vasculature. White dashed contours delineate the tumor core. Inset axes in (b)–(d) show the histogram of the intensity mapping of the respective contrast enhancement techniques.

5.2.2 Results

The proposed contrast enhancement methods are effective at visualization of detailed features in OAT images. Fig. 5.9 demonstrates on a scan of a breast tumor the incremental image improvements afforded by our proposed local contrast normalization, sigmoid normalization, and unsharp masking. Cut-outs displayed in Figs. 5.9e–h show a detail of a peritumoral region (tumor delineated by white contour) with numerous small blood vessels. Whereas the original image has poor contrast preventing observing any details, the application of LCN ($s = 7$, $\alpha = 0.22$) improves the visibility of small blood vessels. At the same time, it does not amplify the noise in the lower half of the image where the signal is weak. The application of SN ($x_1 = 0.06$, $x_2 = 1$, $y_1 = 0.145$, $y_2 = 0.995$) improves the contrast further by increasing the intensity values in the lower-mid-range. Thanks to the utilized non-liner S-shape, SN avoids saturation in the high-intensity region and preserves dark values in the low-intensity region. The application of UM ($\beta = 1.5$, $s = 0.5$) leads to additional enhancement of the visibility of small image features, although it also amplifies noise patterns.

Fig. 5.10 shows the histograms of the image before (Fig. 5.9a) and after the contrast enhancement (Fig. 5.9d). The intensity distribution after the contrast enhancement is still skewed towards low intensities, which is not surprising given the FOV contains both deep tissue without OA signal as well as a region above the skin. However, the number of very dark pixels is reduced by a half, compared to the original image. Moreover, the mid- and high-intensity region is much more utilized in the contrast-enhanced image. From an information-theoretic point, we observe a 39% increase in the histogram entropy (from 2.073 nat to 2.884 nat), which can be seen as an increase in the information carried by the individual intensity values.

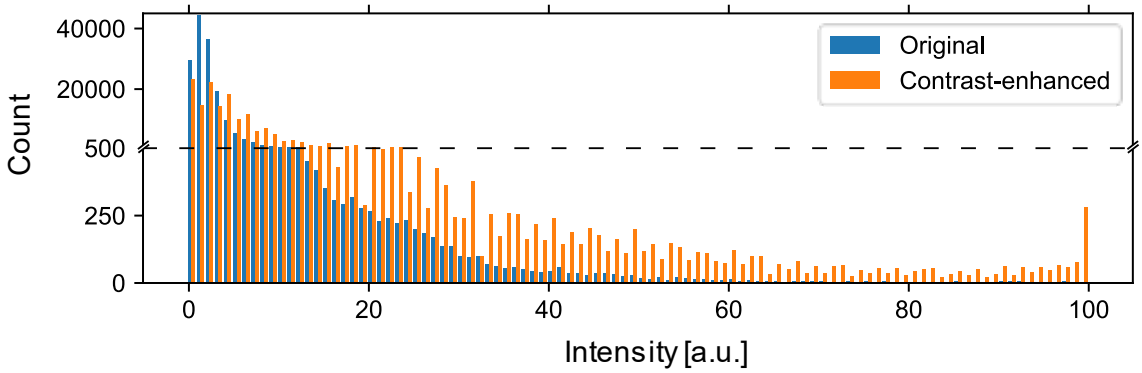

Figure 5.10: Intensity distribution before and after contrast enhancement. The histograms were evaluated on breast tumor images shown in Fig. 5.9a (before) and Fig. 5.9d (after). The scale of the y-axis changes at 500 to better show the distribution of higher intensities. Abbreviations: a.u., arbitrary unit.

5.3 Discussion

Hybrid 2G-OPUS imaging is an information-rich modality and proper visualization is needed to support qualitative evaluation of the images. In this chapter, we have presented dual-band visualization and contrast enhancement tools that together enable displaying of the smallest vascular features on par with large blood vessels and the epidermis layer—the only features visible in unprocessed images. This enables qualitative evaluation of the peri-tumoral microvasculature, which is an important indicator of many tumor parameters relevant for tumor staging, personalized therapy design, and lesion excision planning [29, 245].

The dual-band visualization is an effective technique for emphasizing small-scale image details and differentiating OOP features. It is a precursor to the frequency-band decomposition concept developed later in our group by Longo *et al.* [228]. The latter method has a better theoretical foundation and introduces less artifacts due to its more flexible signal decomposition. In that sense, the dual-band visualization has already been surpassed by a follow-up method at the time of writing this thesis. Furthermore, the dual-band visualization displays single-wavelength images that are suitable for showing a particular structure with a specific absorption peak. The example in Fig. 5.3 uses 880 nm wavelength to show blood vessels. Wavelengths in the range 700–750 nm could be used for deoxy-hemoglobin-specific contrast, whereas 930 nm is the optimal wavelength for showing lipid contrast. However, the single-band images of the newer method by Longo are suitable for spectral unmixing and could be used for visualization of multiple chromophores in a single image.

The contrast enhancement pipeline, proposed in the second part of this chapter, has proved to be effective at suppressing the negative effects of the exponential variation of light fluence with depth. Avoiding explicit modelling of the light transport in the tissue, our proposed method is computationally very fast and avoids errors stemming from simplifying assumptions. In its data-driven approach to intensity normalization, our method is similar to the scheme implemented by *Imagio* ® [108], yet we do not give any false promises to provide quantitative images.

The proposed local contrast normalization produces images visually similar to the adaptive histogram equalization (see Figs. 5.5d and 5.6d), yet the inspection of the mapping histograms (figure insets) shows that LCN maps the mid-range values to a broader intensity region and achieves thus a higher dynamical contrast. At the same time, LCN has an effective control over noise exaggeration in deeper regions. Using the presented graphical interface, it is easy to achieve a good visualization of individual OA images.

Overall, the importance of the methods presented in this chapter lies with the specific clinical scenarios which they enable. Chapter 7 is dedicated to clinical studies which we have performed within this thesis and a separate discussion of the clinical observations, enabled by our visualization methods, is deferred there.

‡ Despite what its name might suggest, the only relation of Beer-Lambert law to beer the beverage is that the law can indeed be used to model the loss of light intensity upon its propagation through liquid media including beer and other beverages. The law owns its name to scientists Johann Heinrich Lambert and August Beer (pronounced [beːɐ̯], not [bɪːə]) and the reader is cordially invited—encouraged even—to make a toast to honor them upon the first suitable occasion.

CHAPTER 6

Image segmentation

Semantic image segmentation (SIS) is a fundamental task of advanced image analysis. The goal of SIS is separation of the image into coherent regions (segments) and identification of their contents (semantics). Alternatively, SIS can be thought of as classification of every single pixel of the image according to the object it belongs to. For example, an OAT image might be segmented into blood vessels, skin layers, fatty tissue, the probe membrane, and other targets [246]. In that case, the segmentation would yield a map of tissue distribution within that image.

Knowledge of the tissue distribution in the FOV would benefit OA imaging on multiple levels. First, the image reconstruction can be improved by incorporating a speed-of-sound distribution estimated from the tissue segmentation [247, 248]. Second, quantitative OAT can be facilitated through a correction of the light fluence using a segmentation-based tissue model. The tissue model can be either used to directly simulate the light propagation [249, 250], or to regularize an optimization-based fluence correction scheme [243, 251]. Finally, further quantitative biomarkers can be derived from the morphology of the segmented anatomical features, such as parameters of the vascular network [252-255].

Semantic segmentation of OA images has been attempted in various settings. In pre-clinical OAT, localization of the animal body in the FOV belongs to the most studied segmentation problems [250, 256]. Localization of the mouse body enables derivation of a speed-of-sound model for an improved image reconstruction. In pre-clinical optoacoustic microscopy (OAM), automatic segmentation of blood vessels has been demonstrated while studying wound healing mechanisms [254] and the vasculature of a prostate tumor model [255]. Raumonen and Tarvainen have devised a multi-step algorithm for blood vessel segmentation in pre-clinical OAT [257]. In clinical imaging of blood vessels, segmentation using deep learning has been demonstrated on healthy volunteer data from OAM [258] and OAT images [259]. Finally, the feasibility of multi-label tissue segmentation in hybrid MS-OPUS clinical scans of healthy volunteers using deep learning has been demonstrated by Schellenberg *et al*. [246]. Their study is the most advanced effort towards obtaining a precise tissue segmentation in clinical OAT so far.

Despite the above-listed efforts, the problem of tissue segmentation in OAT is far from being solved. On one hand, the existing results suggest that—similarly to many other fields [260]—deep learning is currently the superior approach towards semantic segmentation in OAT. On the other hand, it is well known that the key to unlocking the potential of deep learning lies in the training data. While the above studies serve as a proof-of-concept of applicability of the deep learning methodology to the OAT domain, they lack in terms of the available training data. Schellenberg *et al.* acquired a dataset of 288 images from 16 healthy volunteers [246], Chlis *et al.* acquired 164 images from 6 healthy volunteers [259], while Ly *et al.* acquired three volumetric scans from a

This chapter is partially based on the following manuscript whose parts appear verbatim without further notice: Kukačka, Jan, Anja Zenz, Marcel Kollovieh, Dominik Jüstel, and Vasilis Ntziachristos. "Self-Supervised Learning from Unlabeled Fundus Photographs Improves Segmentation of the Retina." arXiv preprint arXiv:2108.02798 (2021).

single healthy volunteer [258]. Such datasets are too small to be representative of the tissue heterogeneity and larger datasets of annotated clinical OA images are unavailable.

Collection of annotated datasets of clinical images is a challenging task in general [261]. Besides the patient privacy and ethical concerns, manual annotations by medical experts are also particularly costly to obtain, especially in the labor-intensive process of manual segmentation. Moreover, the manual segmentation is affected by inter-observer variability and the collected labels carry unknown uncertainty [262].

Apart from the general difficulty of collection of clinical datasets for deep learning, OAT imaging faces two specific challenges. First, no clinical standard for interpretation of OAT images exists [263]. Whereas established imaging modalities have been studied for decades, OAT is an emerging field and precisely understanding the meaning of the obtained images is a focus of active research. To that end, several studies have correlated OAT images with histological findings and other imaging modalities. Dual-wavelength OPUS images of *ex vivo* breast cancer specimens were examined for correspondences with histology [264] and X-ray images [265] to identify the appearance of adipose tissue, larger blood vessels, and microcalcifications. While offering a precise spatial co-registration of the images, *ex vivo* specimens do not realistically portray the appearance of the tumor tissue *in vivo*, primarily due to alteration of the hemoglobin contrast (or lack thereof). Images of breast cancer *in vivo* were correlated with the matching histological sections of the excised tumors in two studies demonstrating the relation between the detected OA features and the true distribution of blood vessels in the tumors [266, 267]. In pre-clinical setting, the location of blood vessels has been validated using golden nanorods as a contrast agent [268]. Besides imaging real tissues, the OA contrast may be also studied through realistic tissue phantoms [269]. In all cases, the interpretation of OAT images is further complicated by presence of various imaging artifacts caused by acoustic reflections [270, 271], detector sparsity and limited angular coverage [272]. Characterization and suppression of those artifacts is also important for further progress. Overall, the lack of consensus about correct labeling can be visualized on the result of an experiment performed by Schellenberg *et al.*, where segmentations of blood vessels from five human annotators were compared to a baseline segmentation provided by another human expert. The resulting mean Dice coefficient was only 0.66±0.09, indicating a considerable level of disagreement [246].

The second challenge stems from a significant variability of the produced image appearance and quality between various implementations of OAT by various academic and industrial manufacturers [104]. Thus, datasets collected from one type of scanner are useless to users of another type. The competition among different scanner designs and the incompatibility of the collected data obstructs large-scale dataset-sharing collaborations between research groups. Moreover, the scanners and image processing pipelines evolve at a fast pace and considerably improved designs are presented every few years. This renders previously collected data obsolete and hinders the efforts towards collection of large, high-quality datasets [273].

For these reasons, researchers often resort to using simulated data [273]. However, a considerable domain gap between the simulated and *in vivo* data exists, and models trained on simulated data fail to perform well on experimental datasets.

In our work, we focus on validation of deep learning methods that can operate under the constraints outlined above. We assume that obtaining a relatively large, unstructured database of OAT images from one or more scanners is feasible, whereas collecting a large, well-structured, reliably annotated dataset is not. Following this assumption, we propose to utilize self-supervised learning, where a neural network is trained from unlabeled data to discover patterns and learn a high-level representation of the training data. The obtained high-level representation can then be exploited to simplify learning of the segmentation task even if annotated data are scarce.

Facing the lack of available OAT data, we examine the assumed scenario using fundus photographs (retina images) instead. Similarly to OA, fundus photography also uses optical contrast to visualize blood vessels, making it a natural experimental substitute. Owing to the availability of numerous high-quality research datasets, fundus images have indeed been utilized in OA research previously [274, 275]. We make use of a large publicly available dataset of fundus images, EyePACS [276], as an unstructured data corpus to learn a general representation from. Then, we use various smaller annotated datasets to evaluate the benefits of using that pre-trained representation for retinal segmentation tasks.

Our results show that through self-supervised learning, our network—without being provided any labels—learns to recognize distinct anatomical and pathological features of the retina, such as blood vessels, optic disc, fovea, and various lesions. Furthermore, when the network is fine-tuned on a downstream blood vessel segmentation task, it generalizes well and achieves a state-of-the-art performance on images from different datasets. Additionally, the pre-training also leads to shorter training times and an improved few-shot performance on both blood vessel and lesion segmentation tasks. Altogether, our results showcase the benefits of self-supervised pre-training which can play a crucial role in real-world clinical applications requiring robust models able to adapt to new devices with only a few annotated samples.

6.1 Methods

An emerging type of learning neural networks from unlabeled data is self-supervised learning, where the network learns to solve an auxiliary *pre-text task*. Pre-text tasks are designed such that targets can be easily generated from unlabeled data, but and solving them requires learning a non-trivial semantic representation of the data. A network pre-trained on the pre-text task can be further fine-tuned in the usual supervised manner on a *downstream task*. Self-supervised pre-training employing an image restoration pre-text task has been demonstrated for radiological images, outperforming generic representations transferred from ImageNet classification [277]. In the domain of fundus images, several self-supervised pre-training strategies have been shown to improve the accuracy of diabetic retinopathy classification, such as generic visual tasks performed on unlabeled images [278] or depth prediction from matching optical coherence tomography (OCT) images [279]. A multi-modal approach using fluorescein angiography images has led to enhanced optic cup and disc segmentation [280].

Recently, using the pre-text task of contrastive instance discrimination has become the primary method for self-supervised learning of general visual representations from large datasets such as the ImageNet [281-283]. In this type of learning, a neural network is trained to map randomly transformed variants of input images to a new representation, where samples originating from the same image are similar to each other and dissimilar to samples from other images. Despite its success in the general vision domain, the applicability of contrastive self-supervised learning to retinal imaging has not been demonstrated yet. We hypothesize that when combined with a rich, large dataset of retinal photographs such as the EyePACS dataset [276], the method can be used to learn a useful domain-specific representation from unlabeled images. Such a representation could improve the performance on downstream segmentation tasks.

6.1.1 Experiments

To evaluate our hypothesis of the benefits of contrastive self-supervised learning for retinal segmentation, we performed three experiments.

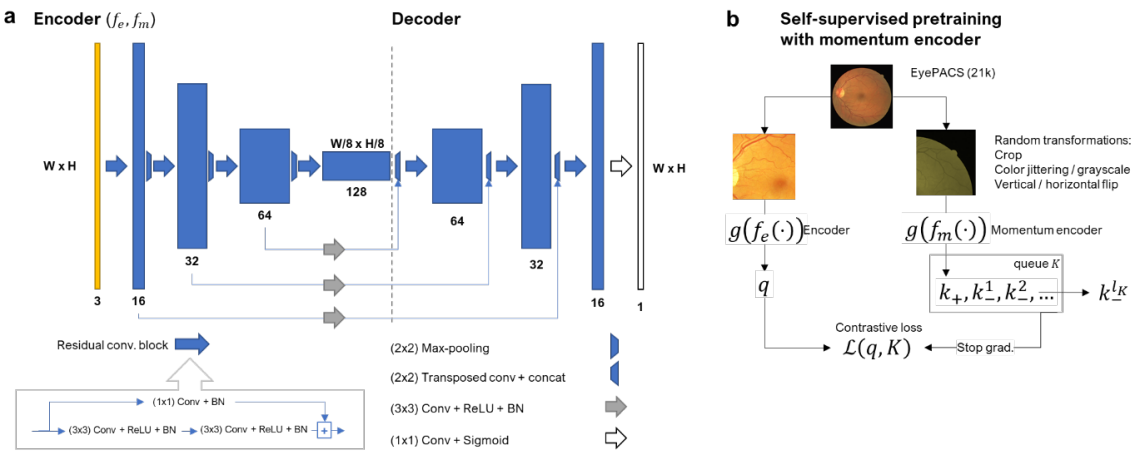

Figure 6.1: **Method overview. a**, Architecture of the used U-Net variant. **b**, Scheme of the momentum contrast (MoCo) self-supervised pre-training approach.

Representation learning experiment. We trained a convolutional encoder in a self-supervised manner on unlabeled images from the EyePACS dataset. Then, we inspected the learned representation by correlating the output feature activation maps to ground truth segmentation of retinal anatomical and pathological targets in the DRIVE, IDRiD, and HRF datasets to identify which targets are recognized by the network. This way, we could characterize the representation obtained through contrastive self-supervised learning.

Image segmentation experiment. To the encoder pre-trained in the representation learning experiment, we appended a decoder with skip connections to form a U-Net network for image segmentation. We fine-tuned this network on DRIVE, HRF, and CHASE-DB retinal vasculature segmentation datasets and IDRiD retinal lesion segmentation dataset. Then, to evaluate the benefits of using the pre-trained representation, we compared the performance of the fine-tuned networks to baselines trained from a random initialization. We also examined the benefits of pre-training in limited data scenarios by repeating this experiment with decreasing numbers of training images down to a single one. We compared the two approaches in terms of segmentation quality and the number of training epochs needed to reach the optimal performance. To increase the sample size for the latter comparison, we pooled together experiments with varying numbers of training images.

Domain transfer experiment (cross-training). To evaluate if pre-training improves the robustness of the model to domain shift, we repeated the cross-training experiment of Galdran *et al.* [284]. We trained a network on the DRIVE dataset and evaluated its performance on five other retinal vasculature segmentation datasets. As in the image segmentation experiment, we compared the performance of a pre-trained network with a baseline trained from scratch.

Domain-transfer error analysis. To analyze how much of the performance drop accompanying model transfer can be attributed to domain shift and how much is caused by labeling mismatch, we tested three modifications: selecting the model checkpoint according to its performance on the target dataset, selecting the classification threshold on the target dataset, and a combination of both. This way, we compared the optimal performance that could be achieved with a model trained on a distinct source dataset and the performance that is achieved in practice due to labeling mismatch.

6.1.2 Network architecture

Previous research on retinal segmentation has demonstrated that a simple U-Net [285] can outperform various other, more intricate networks [284]. Based on that, we utilized a variant of U-Net architecture in our experiments, shown in Fig. 6.1a. U-Net is a type of a network comprising of a fully convolutional encoder-decoder pair with additional skip connections. The used encoder

has 4 levels, each composed of two blocks of 3x3 convolution, ReLU activation, and batch normalization (BN). Additionally, each level has an additive residual connection with 1x1 convolution and BN. Encoder levels are separated by 2x2 max-pooling layers. The encoder is followed by 3 decoder levels composed of the same blocks as the encoder levels. Before every decoder level, a transposed 2x2 convolution with stride 2 is used, whose output is concatenated to a skip connection from the encoder level with corresponding resolution. Optionally, the skip connections contain a 3x3 convolution, ReLU, and BN. Finally, a 1x1 convolution with sigmoid activation serves as the final binary classifier. The first encoder level uses 16 filters; this number is doubled after each pooling operation and halved by each transposed convolution. One exception is the domain transfer experiment where we use a constrained decoder with 16-8-4 features to reduce overfitting.

6.1.3 Self-supervised pre-training

Fig. 6.1b outlines the approach called *momentum contrast* (MoCo) [286] which we used in our experiments for self-supervised encoder pre-training. MoCo is a representation learning method based on the pre-text task of contrastive instance discrimination: randomly transformed variants of training images are encoded by the network to a latent space, where samples originating from the same image should be close to each other and far from samples from other images.

Specifically, images x_i from the training set are transformed by a stochastic transformation $\tilde{t}$ (data augmentation) and encoded to a latent space by an encoder f_e with a non-linear projection head g to obtain a set of keys $k_i \in K$. The same procedure is applied to generate a query sample q_j from an image x_j. Cosine distance is used to compute the similarity of q to keys in K and should be small for a positive key $k_+ \equiv k_{i=j}$ and large for other keys $k_{i\neq j}$. Assuming that vectors q and k are ℓ_2-normalized, this objective can be expressed via the InfoNCE loss function [287, 288] with a temperature hyperparameter τ:

$$\mathcal{L}(q, K) = -\log \frac{\exp(q \cdot k_+/\tau)}{\sum_{k\in K} \exp(q \cdot k/\tau)} \,. \tag{6.1.1}$$

Having a large set of negative keys is essential for good learning, but recomputing embeddings and gradients of the whole training dataset after each update of the encoder is inefficient. To bypass this issue, MoCo uses small batches but maintains the keys in a queue of length l_K. This enables efficient contrastive learning since only a small set of query points is passed through the encoder in each iteration, but a large set of negative keys is available in the queue. A necessary requirement for using such a queue is that the latent representation is not changing too rapidly. To ensure this, MoCo employs a separate *momentum encoder* f_m to embed the keys, whose weights θ_m slowly follow the query encoder weights θ_e via the momentum update $\theta_m \leftarrow \alpha\theta_m + (1-m)\theta_e$, where m is the momentum parameter.

Projection head. Using a non-linear projection head on top of the encoder improves the representations learned by contrastive self-supervised learning as it avoids enforcing invariance to transformations $\tilde{t}$ [283, 286]. We utilized a shared non-linear projection head g on top of both f_e and f_m consisting of global average pooling and two fully connected layers per 128 units with ReLU activation after the first one.

6.1.4 Datasets

Table 6.4 contains a summary of datasets used in our work. For pre-training, we used 21 072 images from the EyePACS dataset [276]. The dataset was collected with different fundus cameras at

multiple centers in the US and is representative of a large variety of medical conditions and imaging artifacts. Although the dataset comes with labels for diabetic retinopathy classification, we did not use them.

For blood vessel segmentation, we utilized several datasets: DRIVE [289], HRF [290], CHASE-DB [291], STARE [292], LES-AV [293], AV-WIDE [294], and DR HAGIS [295]. For image segmentation experiments, we followed train/test splits proposed by [284]. For domain transfer experiments, we used the training split of the source dataset and tested on all data from the target dataset. We repeated experiments with different train/validation splits.

Furthermore, we utilized the IDRiD dataset for lesion segmentation evaluation [296]. The dataset comes with segmentations of hard exudates (EX), soft exudates (SE), hemorrhages (HE), microaneurysms (MA), and optic disc (OD). Additionally, the location of the fovea (FO) is provided. The train/test splits are provided by the authors. We reserved 20% of the training data for validation.

6.1.5 Training setup

For pre-training of the encoder, we first resized and cropped the images to a uniform size 512×512 px. Then, we generated two input samples from each image in the mini-batch by applying data augmentation, with one fed to the encoder and the other to the momentum encoder. Samples processed by the momentum encoder were stored in a queue of length 4 096 and used to compute the loss. In total, we pre-trained the network for 600 epochs.

For training the whole U-Net on segmentation tasks, we resized the images from HRF and IDRiD datasets to a width of 1024 px, images from DRIVE and CHASE-DB datasets were used in their original resolution. For IDRiD, we trained networks for each segmentation target separately (one-vs-all). In total, we trained the networks for 1500 epochs, saved checkpoints every 10 epochs, and finally selected the checkpoint with the best Dice score on the validation set. The validation sets were created by reserving 20% of the training data (rounded down) before the training. In experiments with reduced training set, the validation sets were still selected as 20% of the whole training set, not the reduced one. No pre-processing of the images or post-processing of the segmentation masks other than resizing was used.

Table 6.1: Training hyperparameters for self-supervised encoder pre-training.
Abbreviations: LR, learning rate.

Dataset	EyePACS (first 21 072 images)
Data processing and data augmentation	Resize + crop (512x512 px) Random crop (128x128 px) Color jitter (probability=80%, brightness 0.4, contrast 0.4, saturation 0.4, hue 0.1) Grayscale (p=20%), Horizontal flip (p=50%), Vertical flip (p=50%)
Batch size	64
Optimization	Adam, LR cosine schedule [10^{-2}, 10^{-8}] with restarts after 50 epochs
Weight decay	10^{-4}
Training length	600 epochs
Queue length (l_K)	4096
InfoNCE temperature (τ)	0.07
Momentum (m)	0.999

Table 6.2: Training hyperparameters for segmentation network fine-tuning and training from random initialization. Abbreviations: LR, learning rate.

Dataset	DRIVE	HRF	CHASE-DB	IDRiD
Data processing	Resize (682x1024 px)			Resize (680x1024 px)
Data augmentation	Random rotation [−45°,45°] or scaling [0.95, 1.2] or horizontal translation [−5%, 5%] Color jitter (brightness 0.25, contrast 0.25, saturation 0.25, hue 0.1) Horizontal flip (p=50%), Vertical flip (p=50%)			
Batch size	4			
Optimization	Adam, LR cosine schedule [10^{-2}, 10^{-8}], 50 epoch restarts			Adam, LR=10^{-3}
Weight decay	0			
Training length	1500 epochs			
Convolutional skip connections	No			Yes

Table 6.1 summarizes the hyperparameters used during self-supervised pre-training (representation learning experiment). The temperature hyperparameter τ and the momentum m were set according to previous reports [282, 297]. Table 6.2 summarizes hyperparameters used for training of image segmentation networks (image segmentation experiment). Table 6.3 summarizes parameters used for cross-training (domain-transfer experiment). Weight initialization followed the scheme proposed by He *et al.* [298] in all experiments.

Table 6.3: Training and testing hyperparameters for cross-training experiments. Abbreviations: LR, learning rate.

	Training					
Dataset	DRIVE					
Data processing	Resize (512x529 px)					
Data augmentation	Rand. rotation [−45°,45°] or scaling [0.95, 1.2] or horiz. transl. [−5%, 5%] Color jitter (brightness 0.25, contrast 0.25, saturation 0.25, hue 0.1) Horizontal flip (p=50%), Vertical flip (p=50%)					
Batch size	4					
Optimization	Adam, LR cosine schedule [10^{-2}, 10^{-8}], 50 epoch restarts					
Weight decay	10^{-4}					
Training length	1500 epochs					
Convolutional skip connections	No					
Decoder layers	16-8-4					
	Testing					
Dataset	HRF	DRHAGIS	AV-WIDE	LES-AV	CHASEDB	STARE
Data processing	Resize 682x1024px	Resize 682x1024px	Resize w=1024px	Resize w=512px	Resize w=512 px	Resize w=512px
Augmentation	Horizontal and vertical flipping					

6.1.6 Evaluation protocol

To ensure good reproducibility of our results, we adopted the evaluation protocol described by Galdran *et al.* [284]. In short, we followed the same train/test splits (Table 6.4). We selected the final classifier threshold which maximizes the Dice coefficient on the training set predictions. Then, we evaluated the final performance on all pixels of all testing images (except masked areas outside of the field-of-view) at their original resolution together (contrary to computing mean performance over individual images). Furthermore, we also utilized test-time augmentation by averaging the predicted segmentations for all four possible horizontal and vertical flip combinations.

For blood vessel segmentation tasks, we evaluated the predictions using the Dice coefficient (F1 score). For lesion segmentation, we computed the area under precision-recall curve (AUPRC) to match the evaluation used in the IDRiD challenge.

6.1.7 Statistical analysis

We performed statistical testing of the following hypotheses: in the image segmentation experiment, we tested whether 1) self-supervised pre-training leads to a higher Dice coefficient than the baseline, 2) pre-trained models converge in a lower number of epochs than the baseline, and in the domain transfer experiment, we tested the hypothesis that 3) self-supervised pre-training leads to a higher Dice coefficient on a transfer dataset than the baseline.

We repeated the experiments several times (n=4 for image segmentation, n=12 for domain transfer) and paired the results from matching training/validation splits. To test the hypotheses, we computed one-sided 95% *t*-confidence intervals (CI) on the differences and declared the difference significant if zero was outside the CI. This test assumes normality of the distribution of differences, which was validated by the Shapiro-Wilk test.

The analysis was performed using SciPy (v. 1.6.3) [208] and MS Excel (v. 2104).

Table 6.4: **Overview of datasets utilized in our experiments.** For DRIVE, HRF, and CHASE-DB, we used the same test splits as proposed by [284]. For IDRiD, we used the official test splits.
Abbreviations: SE, soft exudates.

Dataset	Train / Val / Test	W × H	Labels
EyePACS [276]	21 072	400–5184 × 289–3456	None
DRIVE [289]	16 / 4 / 20	565 × 584	Blood vessels
HRF [290]	12 / 3 / 30	3504 × 2336	Blood vessels
CHASE-DB [291]	6 / 2 / 20	999 × 960	Blood vessels
STARE [292]	20	605 × 700	Blood vessels
LES-AV [293]	22	1144 × 1620 1958 × 2196	Blood vessels
AV-WIDE [294]	30	2816 × 1880 1500 × 900	Blood vessels
DR HAGIS [295]	39	2816 × 1880 4752 × 3168	Blood vessels
IDRiD IDRiD (SE) [296]	43 / 11 / 27 21 / 5 / 27	4288 × 2848	Retinal lesions, optic disc, fovea

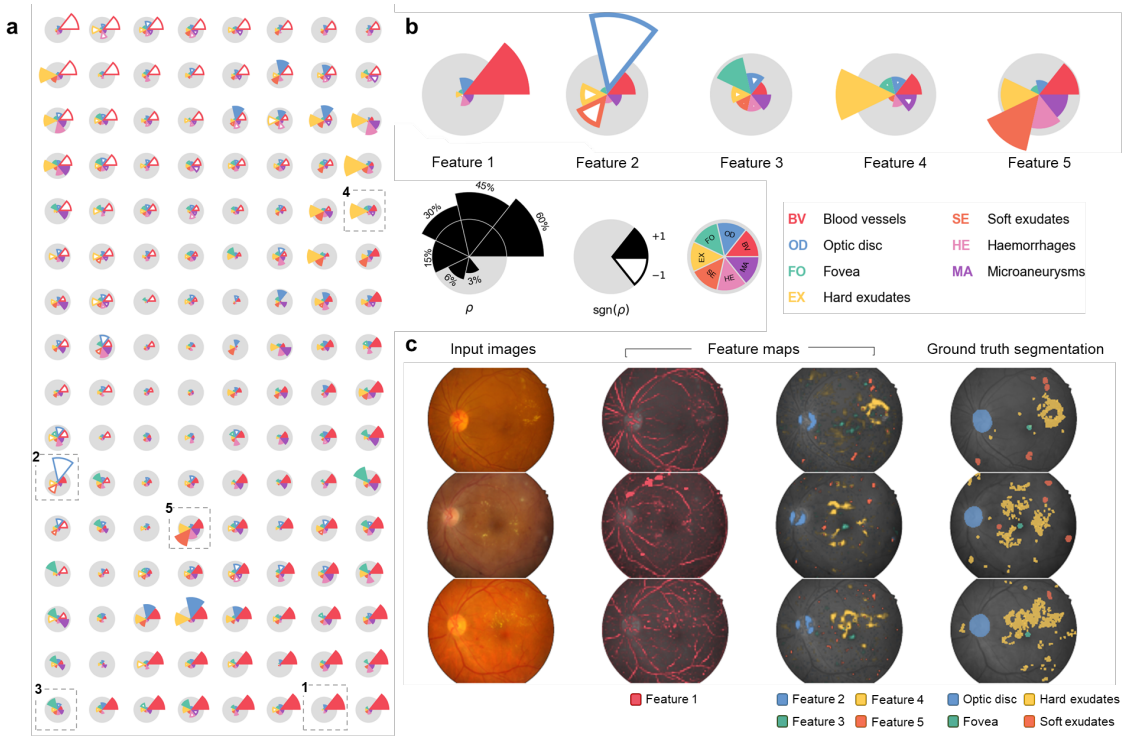

Figure 6.2: **Visualization of the features learned by the self-supervised training. a)** Diagram of Pearson correlation coefficients between the 128 learned features and localization of seven anatomical and pathological targets: blood vessels, optic disc, fovea, hard exudates, soft exudates, hemorrhages, and microaneurysms. **b)** Upper row: Enlarged diagrams for five features marked in panel a. Lower row: Diagram legend. **c)** Three representative images from the IDRiD dataset (left), corresponding activations maps of features 1–5 (middle two columns), and ground truth segmentations of optic disc, fovea, hard exudates, and soft exudates (right). Visual comparison of columns 3 and 4 shows a good spatial match between activations of these selected units with true segmentation.

6.2 Results

6.2.1 Representation learning experiment

Our first experiment revealed that despite using no labels, the representation obtained by self-supervised learning contains features that are specialized in recognizing various retinal structures. Specifically, we evaluated the spatial correlation of seven anatomical and pathological structures with the activation maps in the last layer of an encoder trained on unlabeled images from the EyePACS dataset. Fig. 6.2a shows that many of the 128 units developed a strong correlation with blood vessels, but some specialized in detecting other structures. Fig. 6.2b shows a detailed correlation of five selected units that respond strongly to blood vessels, optic disc, fovea, hard exudates, and soft exudates. Fig. 6.2c displays the spatial agreement of the activations of these five units with the respective targets. Good preservation of spatial information is an essential feature of the self-supervised representation for precise downstream segmentation. Additionally, the features are detected in images of a different population from a separate dataset (USA vs India), demonstrating the robustness of the representation to domain shifts.

6.2.2 Image segmentation experiment

Our second experiment confirmed that using the self-supervised representation as an initialization of the encoder of a U-Net benefits the downstream segmentation tasks and leads to higher segmentation accuracy, especially in the few-shot regime, and faster convergence.

Table 6.5: Performance (AUPRC, %) **of the best methods for retinal lesion segmentation from the IDRiD Grand challenge leaderboard** (sub-challenge 1). Top two rows show test set performance of our U-Net trained with random initialization (baseline) and with self-supervised pre-training. Bold highlights highest performance in each column. Our experiments are reported as mean ± standard deviation of four runs with different training/validation splits. For performance of other methods, we report scores from the challenge leaderboard evaluated on the same test set as our results [296]. **Abbreviations:** EX, hard exudates; SE, soft exudates; HE, hemorrhages; MA, microaneurysms.

Method	EX	SE	HE	MA
Random initialization	88.33 ± 1.46	67.43 ± 1.82	64.73 ± 1.12	**50.70** ± 1.37
Self-supervised pre-training	**89.42** ± 0.60	**69.98** ± 5.46	66.37 ± 1.72	48.42 ± 1.87
PATech	88.50	-	64.90	47.40
VRT	71.27	69.95	**68.04**	49.51
iFLYTEK	87.41	65.88	55.88	50.17
LzyUNCC-I	76.15	66.07	-	-
LzyUNCC-II	82.02	62.59	-	-
SAIHST	85.82	-	-	-
SOONER	73.90	53.69	53.95	40.03

First, we compared the performance of a pre-trained U-Net to a randomly initialized baseline on the IDRiD retinal lesion segmentation dataset and observed an improvement in three out of four lesion types. Table 6.5 shows that the pre-trained model achieved on average a higher AUPRC for hard exudates (EX; 1.09 percentage points, one-sided 95% CI [0.05, +∞]), soft exudates (SE; 2.55 pp), and hemorrhages (HE; 1.64 pp, [0.29, +∞]), but not for microaneurysms (MA; −2.27 pp [−5.24, +∞]). Furthermore, the table lists reported results of the best performing methods from the IDRiD Grand challenge [296]. For EX and SE, the pre-trained model outperformed the challenge winners, whereas for HE and MA its mean performance was within 1 standard deviation of the best method. To provide a qualitative illustration of predictions by the pre-trained model, Fig. 6.3a shows a representative example of the segmentation.

Fig. 6.4 shows that the pre-trained model outperformed the baseline even more when we reduced the number of training images. The improvements were statistically significant for hard exudates

Table 6.6: Performance (Dice coefficient, %) **of retinal vasculature segmentation on DRIVE, CHASE-DB, and HRF.** Our results are reported as mean ± std. deviation of four runs with different training/validation splits. For other methods we list results reported by their authors. We followed exactly the same evaluation procedure, including testing split, described in [284]. Methods using different training/testing split or different score computation procedure are marked by an asterisk (*).

Method	DRIVE	CHASE-DB	HRF
Random initialization	82.66 ± 0.20	80.97 ± 0.44	81.48 ± 0.22
Self-supervised pre-training	82.87 ± 0.13	80.72 ± 0.71	81.38 ± 0.35
Little W-Net [284]	82.82	81.55	81.04
M-GAN [299]	83.17 ± 0.02	81.10	79.72*
HAnet [300]	82.93	81.91*	80.74
M2U-Net [301]	80.30 ± 1.42*	80.22 ± 1.93*	78.00 ± 5.74*
VGN [302]	82.63*	80.34*	81.51*
DEU-Net [303]	82.70	80.37*	
DUNet [304]	82.37	78.83*	

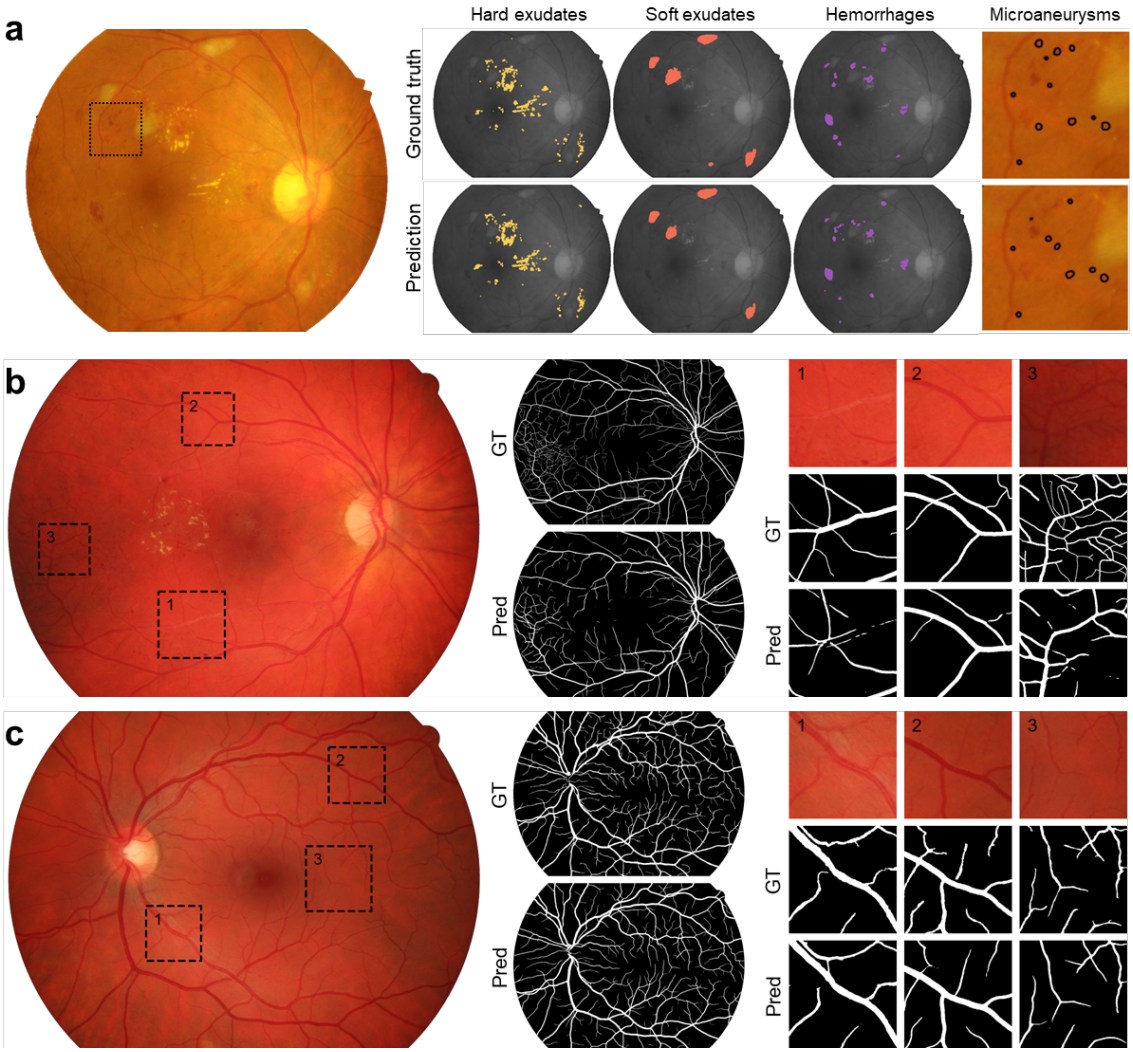

Figure 6.3: Examples of lesion and blood vessel segmentation quality achieved by the pre-trained models. a) Image #59 from IDRiD test set, ground truth segmentations, and segmentation predictions. Microaneurysms are displayed on an enlarged cut-out due to their small size (enlarged region marked in the left image). **b)** Diabetic retinopathy patient #6 of HRF dataset (worst case, Dice 72.84%). **c)** Healthy patient #12 of HRF dataset (best case, Dice 88.30%). Cut-outs showing arteries (1), veins (2), and microvessels (3) with corresponding ground truth and predicted segmentation. Locations of the cut-outs are delineated in the images on the left in panels b and c. **Abbreviations:** GT, ground truth; Pred, prediction.

(21 images: 7.69 pp [2.74, +∞], 10 images: 5.02 pp [3.97, +∞], 5 images: 9.31 pp [3.77, +∞], 2 images: 12.25 pp [2.83, +∞]) and hemorrhages (10 images: 16.56 pp [0.91, +∞]). However, for microaneurysms, the pre-trained model performed better only in the case of 2 training images.

Second, we compared the performance of a pre-trained U-Net to a randomly initialized baseline on the three retinal vessel segmentation datasets, but in this case, we did not observe any significant differences. Table 6.6 shows that the Dice score on DRIVE improved slightly (0.21 pp, two-sided 95% CI [−0.21, 0.63]) and decreased a little on CHASE-DB (−0.25 pp [−1.13, 0.64]) and HRF (−0.11 pp, [−0.40, 0.19]). As the confidence intervals show, neither of the differences is statistically significant. Additionally, the table lists the performance reported by several state-of-the-art methods (asterisks mark deviations from our evaluation protocol; see section 6.1.6) [299-304]. Our results confirm that a simple U-Net achieves comparable results to more complicated, specialized architectures.

To provide representative segmentation examples, Fig. 6.3b and c show the worst and the best images from the HRF dataset respectively, with detailed segmentations of arteries (1), veins (2),

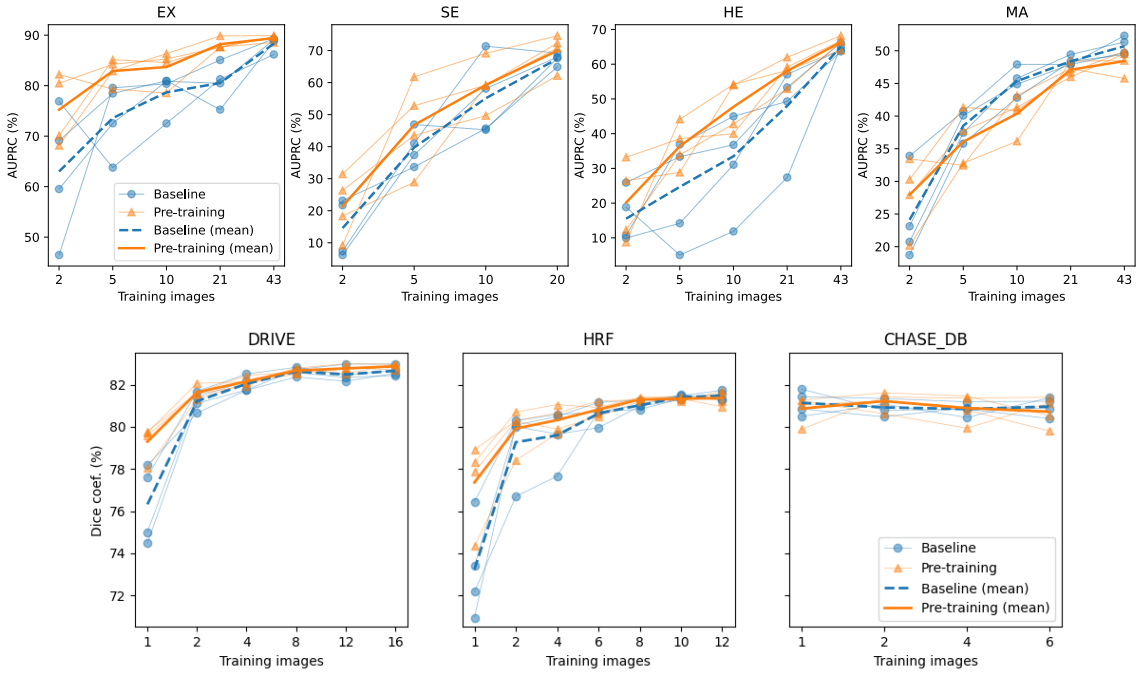

Figure 6.4: **Performance comparison of pre-trained and randomly initialized U-Net** on IDRiD lesion segmentation dataset (upper row) and retinal vasculature segmentation datasets DRIVE, HRF, and CHASE-DB (bottom row) for decreasing amount of training data. Plots in the bottom row have a shared y-axis. The experiments were repeated four times with different training/validation splits (individual runs are plotted as thin lines, thick lines represent the means). **Abbreviations:** EX, hard exudates; SE, soft exudates; HE, hemorrhages; MA, microaneurysms.

and microvessels (3). Fig. 6.3b shows a diabetic retinopathy case where the model achieved a 72.84% Dice score. Here, the model correctly segmented the veins but failed to detect arteries with strong central reflex lending them a whitish appearance and atypical retinopathic neovascularization in a dark region of the image. Fig. 6.3c shows a healthy case where the model provided a correct segmentation for veins, microvessels, and arteries even in the presence of central reflex, and achieved a Dice of 88.30%.

In the few-shot regime, the pre-trained model outperformed the baseline on the DRIVE and HRF datasets but not on CHASE-DB. Fig. 6.4 shows that the gap between the models grows as we reduce the number of available training images. For DRIVE, we observed significant improvement

Table 6.7: **Mean training length improvement of pre-trained network over a randomly initialized U-Net** (epoch count). The number of experiment runs considered is listed under n. **Abbreviations:** CI, confidence interval.

Lesions	**n**	**Difference**	**95% CI**
Hard exudates	20	232	64, +∞
Soft exudates	16	205	-139, +∞
Microaneurysms	20	292	75, +∞
Hemorrhages	20	283	133, +∞
Blood vessels			
DRIVE	24	173	99, +∞
HRF	28	180	50, +∞
CHASE-DB	16	289	154, +∞

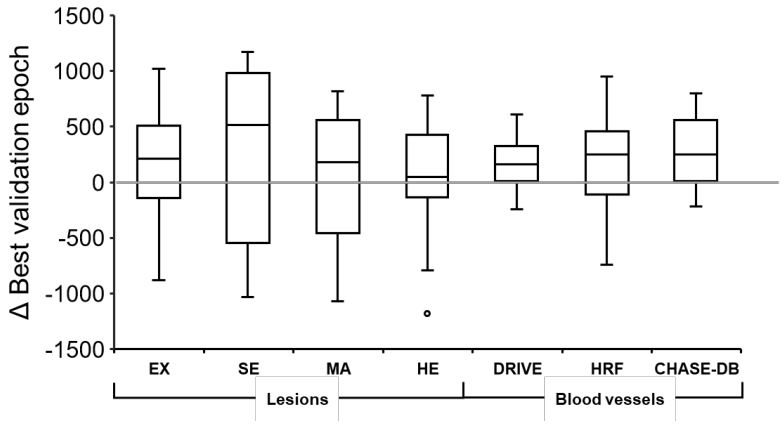

Figure 6.5: **Training length improvement of pre-trained network over a randomly initialized U-Net. Abbreviations:** EX, hard exudates; SE, soft exudates; HE, hemorrhages; MA, microaneurysms.

over baseline in case of one training image (2.98 pp, one-sided 95% CI [0.11, +∞]) and two training images (0.42 pp [0.01, +∞]). For HRF, the improvement was significant in case of one training image (4.12 pp [0.90, +∞]) and eight training images (0.28 pp [−0.03, +∞]). The segmentation task for the CHASE-DB dataset appears easier than in DRIVE and HRF and reducing the number of training images down to a single one does not alter the performance of either model.

Besides improved segmentation performance, the pre-trained models were also converging faster. Fig. 6.5 shows the distribution of differences in the number of training epochs needed by the baseline and the pre-trained model to reach the best validation error across the lesion and blood vessel segmentation tasks. Table 6.7 summarizes the exact differences and shows that, except for soft exudates and microaneurysms, the improvements are statistically significant.

6.2.3 Domain transfer experiment (cross-training)

Our final experiment demonstrated that pre-trained models are also more robust to domain shift and thus better suited for domain transfer. Table 6.8 shows a performance comparison of models trained on DRIVE dataset and tested on other retinal vasculature segmentation datasets. In four of the six cases (STARE, LES-AV, DR HAGIS, AV-WIDE), the pre-trained model outperformed the baseline. The improvement was statistically significant for DR HAGIS (0.62 pp, [0, +∞]) and AV-WIDE (1.89 pp, [0.14, +∞]). Furthermore, the table shows that in three cases (HRF, DR HAGIS, and AV-WIDE), our approach also outperformed Little W-Net [284], a state-of-the-art approach in retinal segmentation domain transfer using pseudo-label re-training on the target dataset.

6.2.4 Domain-transfer error analysis

We observed that although the self-supervised pre-training improves the model transfer, training even on a single image from the target dataset can yield superior performance to transferring a model from another, larger dataset (cf. Fig. 6.4). This performance gap is partially caused by

Table 6.8: **Cross-training evaluation.** Dice coefficients (%). Mean ± standard deviation of 12 runs is reported. For Little W-Net, results reported by its authors are shown.

Method	CHASE-DB	HRF	STARE	LES-AV	DR HAGIS	AV-WIDE
Random init.	75.91 ± 0.61	75.93 ± 0.41	78.47 ± 0.62	77.12 ± 1.12	70.91 ± 1.16	67.43 ± 3.70
SSL pre-training	75.79 ± 0.52	75.84 ± 0.51	78.77 ± 0.65	77.31 ± 0.79	71.54 ± 1.09	69.32 ± 1.51
Little W-Net [284]	76.49	71.12	79.76	77.93	68.67	62.46

domain shift, in which case the network is unable to recognize patterns in images from a different dataset, and can be potentially remedied by learning a more robust representation. Additionally, the performance is reduced due to labeling mismatch, in which case the network can recognize patterns, but it was trained to label them differently (e.g., small vessels). This mismatch is expected in fundus photographs, since the presence of retinal blood vessels cannot be exactly represented as a binary mask and there is a considerable inter-observer variability regarding how small vessels and vessel edges are handled [305]. The relevance of labeling-related performance drop for medical segmentation applications has been questioned in the past [306].

We estimate that about 2.5 pp are caused by labeling mismatch and expect that the remaining ~4 pp could be reduced by further improvements to representation learning. To characterize the contribution of labeling mismatch to the performance gap, we used the target dataset for selection of the optimal checkpoint and classification threshold instead of the source dataset. These changes reduce the error due to labeling error but do not alter the training data and thus do not affect the domain shift problem. Table 6.9 lists the performance improvements achieved by these modifications.

6.3 Discussion

The discussion of this chapter is split into two parts. First, we discuss the results of our experiments from the perspective of retinal imaging. Later, we discuss the implications of this work on the segmentation of optoacoustic images.

6.3.1 Retinal segmentation

In this work, we have demonstrated for the first time the feasibility of learning robust representations of fundus images capable of recognizing important anatomical and pathological features in a completely label-free manner. Furthermore, we have identified benefits of using this representation for downstream image segmentation tasks, notably, an improved performance in few-label scenarios, shorter training times, and improved robustness to domain shifts. Our results suggest contrastive self-supervised learning as an effective way to exploit unlabeled fundus images and advance automated retinal diagnostics.

Various methods to obtain useful representations of fundus images have been considered in the past. Unsupervised methods relying on handcrafted features have long been used to segment retinal vasculature [289, 290, 292]. It was recently shown that weakly supervised convolutional neural networks implicitly learn to detect retinopathic lesions [307, 308]. Our work is the first demonstration that a conv-net can learn to recognize retinal vasculature and lesions in a fully data-driven manner without any labels. By learning directly from the data, the resulting representation

Table 6.9: Influence of checkpoint selection and binarization threshold of models trained on DRIVE and tested on HRF. Means of n=4 runs are reported.

Training set	Checkpoint	Threshold	Dice (%)
DRIVE	DRIVE	DRIVE	75.06
	DRIVE	HRF	75.57
	HRF	DRIVE	77.11
	HRF	HRF	77.47
HRF	HRF	HRF	81.38

is more robust than handcrafted approaches. Moreover, requiring no labels, the self-supervised approach can be easily applied to unannotated datasets, which are cheaper to obtain.

Our work provides insight into the features learned during self-supervised pre-training. Currently, self-supervised pre-training is mostly used in other domains as a black-box method, and there is very little insight into the properties and quality of the learned representation. It is well recognized that the representation depends on numerous hyperparameters, such as stochastic transformations used for generating training samples and the encoder architecture, but their immediate effect on downstream tasks is not clear [281, 309, 310]. Only few studies have inspected the qualitative properties of the self-supervised representations, such as exploring nearest neighbors in the embedding space [310, 311], visualizing saliency maps [310], reconstructing inputs from the embeddings [312], or studying the intrinsic dimensionality of the embedded dataset [311]. Complementing these approaches, our feature correlation analysis (Fig. 6.2**6.2**) allows to easily assess the impact of changing the hyperparameters on the features that the network recognizes without lengthy fine-tuning on a downstream task.

Furthermore, we have identified specific benefits of using the self-supervised representation for downstream segmentation tasks. First, it achieved better segmentation performance in a few-shot learning regime than a baseline trained from scratch. In vessel segmentation tasks, this improvement diminished when we sufficiently increased the number of labeled images (Fig. 6.4). For more difficult, lesion segmentation tasks (EX, SE, HE), we did not observe such saturation even when the full training dataset was used (Fig. 6.4). The same saturation pattern was also recently reported on simulated data [313]. An interesting exception was performance on the CHASE-DB dataset, where the performance with full training dataset was comparable to performance of using a single training image for both baseline and pre-trained models. This task requires segmenting only large vessels and appears to be straightforward. Another exception was segmentation of microaneurysms where the pre-trained model performed consistently worse than the baseline. We assume this is because the pre-trained model does not contain any features that correlate very well with MA (Fig. 6.2) and cannot adapt to new targets from its initial configuration. Using different data augmentation enhancing their contrast could improve this situation.

The second observed benefit was faster convergence of pre-trained models. This is not surprising since the pre-trained model has already learned necessary low-level feature detectors and similar results were reported in other studies [278, 314]. In practice, faster convergence can lead to large computation savings for hyperparameter tuning on the downstream task and serves as amortization of the pre-training computation costs.

The third benefit was improved robustness to domain shifts—probably the most unique and important advantage of using self-supervised pre-training (Table 6.8). Self-supervised pre-training outperforms the pseudo-label method of Little W-Net in three out of six cases [284] but is generally worse than training on a combination of multiple datasets [301]. As these three approaches are independent, we expect that using them in combination could lead to superior results. We also remark that in cases of HRF, DR HAGIS, and AV-WIDE, we resized the target images to a width of 1024 px instead of 512 px used by the Little W-Net authors, which granted quite large improvements (observable in our baseline performance, Table 6.8). We conclude that attention must be paid to proper resolution matching between source and target images, considering the field-of-view of the cameras and aspect-ratio of the image files, since a resolution mismatch can easily hinder improvements granted by advanced machine learning techniques.

The advantages of self-supervised pretraining are essential for utilization in fast-evolving, low-cost, smartphone-based fundus cameras [315-317], which exhibit large variability in imaging conditions and require methods that can easily adapt to new devices without large, annotated

datasets. Additionally, unlike large-scale self-supervised visual models, our approach can be easily used with consumer-level GPUs (pre-training for 600 epochs took 7 hrs on a single NVidia RTX 3090).

Whereas our experiments show promising results, open questions remain. It has been shown that large self-supervised networks can learn strong representations from huge datasets with billions of images and outperform their fully supervised counterparts [283]. Conversely, the representation learned by our small network did not improve beyond 21k images (only 1/4 of the available data). Using a larger encoder might enable learning even better representations. On the other hand, huge networks often perform poorly on small downstream datasets typical for the medical domain, and careful layer freezing [314] or knowledge distillation [283, 318] might be necessary to prevent overfitting.

Overall, this work demonstrates how contrastive self-supervised learning can be applied to segmentation of fundus photographs. We identified specific benefits, such as better few-shot performance, faster convergence times, and improved domain transfer. These benefits are relevant for deploying automated models into screening programs based on fundus photography, which is essential to reduce the workload on individual ophthalmologists and increase the capacity of current healthcare systems.

6.3.2 Optoacoustic tomography segmentation

Although we did not perform experiments on OAT data, conclusions related to OAT segmentation can still be drawn from the presented work.

First, we observed that the contrastive learning led to emergence of features responding to various anatomical features and abnormalities encountered in the unlabeled training data. We hypothesize that similar features could also be learned from OAT images in a label-free manner. Potential targets are blood vessels, subcutaneous fat, skin, but also frequently appearing artifacts such as streaks related to limited view.

The image segmentation experiment has shown that a high accuracy of blood vessels segmentation in fundus photographs could be achieved even with a modest number of annotated images, and the pre-trained network did not improve this accuracy further. Since the detection of large blood vessels, adipose tissue, and skin is also relatively easy in OAT images, we expect that a pre-trained representation would have only a small benefit for segmentation of those targets. Indeed, Chlis *et al.* have achieved a very high Dice coefficient of 90% at segmentation of large blood vessels in MSOT scans [259]. On the other hand, the improvements on the harder, lesion-detection task were significant. Accordingly, a pre-trained model for OAT could also provide larger benefits for targets that are harder to detect, such as disease-related pathologies.

Besides benefits to automated segmentation, a pre-trained representation sensitive to image artifacts could be exploited for image quality improvement. For example, penalization of activation of artifact-related features could serve as a regularizer in a model-based image reconstruction procedure. Alternatively, the activations could be used to mask-out artifact-related features as a post-processing step or highlight the features as uncertain during a clinical evaluation.

To examine which features emerged from the contrastive learning, we proposed a feature interpretation method based on correlations between the feature map activations and the manual segmentations. The method can be used to evaluate if the representation contains the desired features (e.g., limited-view artifacts) and adjust the pre-text task as needed. Specifically, the combination of data augmentation transformations used to generate queries and keys play a crucial role in the representation obtained by a contrastive instance discrimination [281]. The amount of manually segmented data needed to compute the feature activation correlations is small compared

to the amount of data needed for training a neural network and should not present a practical limitation on the usability of the method.

The EyePACS dataset contains images collected from multiple centers and multiple cameras. Our (unreported) examination of the embeddings revealed clusters corresponding to trivial image differences based on the FOV and the size of the black margins. It is a remaining open question if pooling the data from multiple devices improves the learned representation or hinders the learning due to trivial distinction between images from various sources. In the latter case, increased robustness of the representation could not be achieved by pooling data from multiple scanner types. It is more likely that a benefit could be observed when pooling data from a single scanner reconstructed with different variants of the reconstruction algorithms or different reconstruction parameters.

A possible direction to utilize pooled datasets is through simulation of appearance of images from a certain scanner with image style transfer approaches [319, 320]. Specifically, the style transfer could be used as an image transformation in the pre-text augmentation to force the network to focus on image features instead of image appearance. Furthermore, as different scanners capture images at different wavelengths, image style transfer methods could be used, to a certain extent, to simulate the appearance of an image at a wavelength that was not obtained. Image style transfer approach has been successfully used for virtual staining of histology specimens from autofluorescence images [321] and UV-OAM [322, 323]. In OAT, style transfer has been utilized to reduce the domain gap between real and simulated data [324]. However, the feasibility of a simulation of the appearance of images from a different scanner has not been studied yet.

Overall, this work supports our initial hypothesis that unlabeled OAT data could be utilized by contrastive learning to produce a neural representation with multiple benefits for downstream image segmentation.

CHAPTER 7

Clinical applications

The technical improvements to the image processing and visualization presented in this thesis—collectively termed second-generation optoacoustic-ultrasound, or 2G-OPUS—empower OAT with new capabilities to visualize tissue features. In this chapter, we present the results of two clinical studies utilizing 2G-OPUS for *in vivo* imaging. In the first study, we characterized features observed in breast tumors, including tumor rim enhancement and hypoxia, centripetally arranged blood vessels resolved to a level of detail unprecedented in handheld OAT, as well as cysts and periductal inflammation. The second study explored the possibility of non-invasive detection of metastases in cervical lymph nodes with the aid of tumor-specific optical contrast agent. Unable to detect the tracer, we assess the limitations of our approach and recommend solutions to be adopted in future studies. We also characterize the differences in observed hemoglobin distribution in healthy and metastatic lymph nodes. Overall, the results of these two studies highlight the power of multimodal handheld OPUS imaging in clinical applications.

7.1 Breast cancer

The seriousness of breast cancer and the role of OA imaging in its clinical management have been already outlined at the very beginning of this thesis. Despite the long-standing promise of OAT to aid diagnosing breast cancer and monitoring its treatment, existing handheld solutions have been delivering images of insufficient quality. The 2G-OPUS processing pipeline pushes the boundaries of image features that can be resolved and may facilitate new clinical applications. We conducted a clinical study involving 22 patients to examine the breast cancer features that can be detected with the 2G-OPUS technique. We present a detailed analysis of eight representative cases to characterize the observed patterns. We showcase both anatomical and spectral features from selected regions of the images. The findings presented herein highlight the most advanced state of handheld breast cancer OA imaging performance.

This chapter is partially based on the following publications, whose parts appear verbatim without further notice:

Kukačka, Jan, Stephan Metz, Christoph Dehner, Alexander Muckenhuber, Korbinian Paul-Yuan, Angelos Karlas, Eva Maria Fallenberg, Ernst Rummeny, Dominik Jüstel, and Vasilis Ntziachristos. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.

Vonk, Jasper, Jan Kukačka, Pieter J. Steinkamp, Jaron G. de Wit, Floris J. Voskuil, Wouter T.R. Hooghiemstra, Max Bader, Dominik Jüstel, Vasilis Ntziachristos, Gooitzen M. van Dam and Max J. H. Witjes. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.

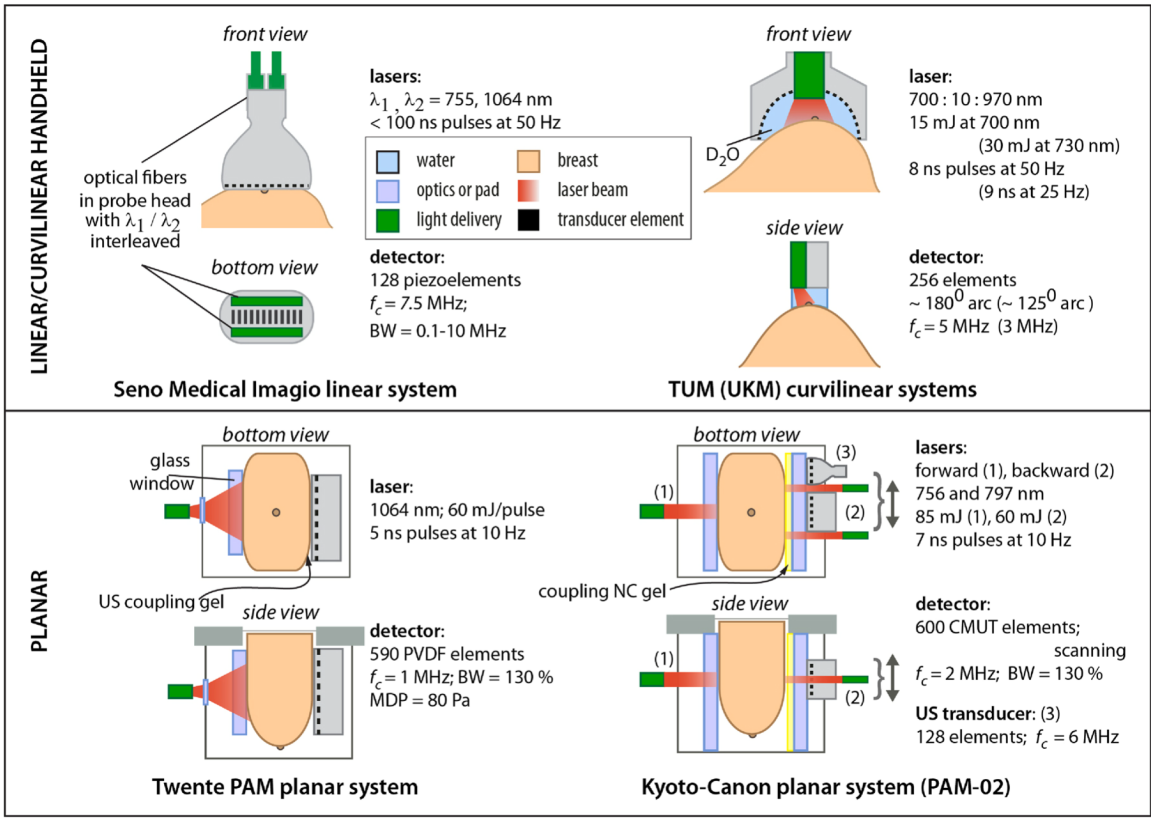

Figure 7.1: Schematic drawings of photoacoustic breast imaging instruments with linear, curvilinear and planar detection apertures. With selection of details of their lasers and ultrasound detectors. **Abbreviations:** BW, bandwidth; CMUT, capacitive micromachined ultrasonic transducer; MDP, minimum detectable pressure; NC, nanocomposite; PAM, photoacoustic mammoscope; PVDF, polyvinylidene fluoride; TUM, Technical University of Munich; UKM, University Hospital Münster; US, ultrasound. Figure reproduced from *Manohar, Srirang, and Maura Dantuma. "Current and future trends in photoacoustic breast imaging." Photoacoustics 16 (2019): 100134.*

7.1.1 Related work

Imaging of breast cancer has been attempted from the early days of OAT. In 1999, Oraevsky *et al*. demonstrated the feasibility of imaging endogenous optical contrast deep in breast tissue on phantoms and an *ex vivo* mastectomy specimen with a system called *LOIS* [31]. In 2001, the same group demonstrated a handheld *LOIS* system for imaging breast tissue *in vivo* [141, 220]. However, these early OAT prototypes employed detectors with an overall low number of ultrasound elements (12 and 32, resp.) and produced images using a simple back-projection inversion technique [325]. Correspondingly, the initial image quality left a large space for improvements.

Since then, the quality of *in vivo* breast cancer imaging has been steadily increasing through various implementations of OAT. In 2005, the Twente photoacoustic mammoscope (*PAM*; Fig. 7.1) has been presented [45]. Although utilizing the same 1064 nm illumination wavelength as the *LOIS* systems, *PAM* was pioneering an application-specific design of bed-based OAT scanners. Bed-based systems are integrated into a modified hospital bed where the patient lies prone with the scanned breast pendant through an aperture opening. The breast is immobilized which suppresses motion artifacts and enables 3D scanning and the use of large apertures, leading to image quality improvements and minimized intra-operator variability [104]. *PAM* utilizes a planar detector array with 588 elements. Results of the initial *PAM* clinical study on ten patients showed, similar to the first *LOIS in vivo* study, increased OA signal in breast lesions [50].

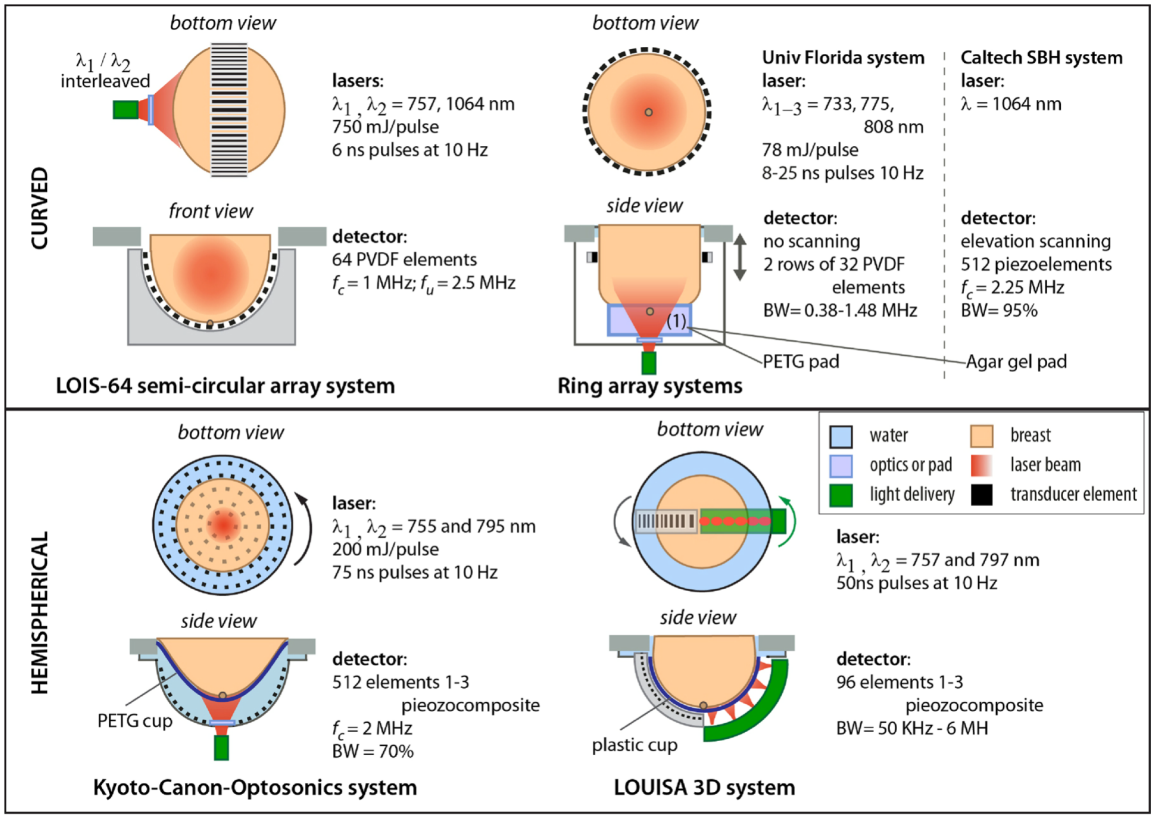

Figure 7.2: Schematic drawings of photoacoustic breast imaging instruments with curved, ring and hemispherical detection apertures. With selection of details of their lasers and ultrasound detectors. **Abbreviations:** BW, bandwidth; PETG, polyethylene terephthalate; PVDF, polyvinylidene fluoride; SBH, single breath-hold. Figure reproduced from *Manohar, Srirang, and Maura Dantuma. "Current and future trends in photoacoustic breast imaging." Photoacoustics 16 (2019): 100134.*

Later—in 2014—the bed-based design with a planar detector array was utilized also by the Kyoto University–Canon consortium and their *PAM-01* [326, 327] and *PAM-02* [59] scanner designs (Fig. 7.1). Those systems were equipped with a tunable Q-switched Nd:YAG lasers, allowing illumination at multiple wavelengths (756 nm and 797 nm). From the two wavelengths, a semi-quantitative estimate of oxygen saturation (sO_2) called S-factor could be computed. A clinical study using *PAM-01* on 42 patients has shown a micro-vessel-related signals at the tumor site in 74.4% cases [58]. Furthermore, reduced sO_2 levels relative to subcutaneous vasculature and contralateral breast was observed in the imaged tumors. With an improved resolution of ~1 mm, *PAM-02* images could resolve an important pattern of individual blood vessels centripetally arranged around the imaged tumor. A later design by the same group, *PAM-03* [47], was based on a scanner design with a hemispherical arrangement of detectors forming a cup around the imaged breast, developed by Optosonics Inc. (USA) in 2013 [57]. *PAM-03* (Fig. 7.2) employs a similar wavelength selection to its predecessors but yields a superior resolution of 0.37 mm axially and 0.57 mm laterally. A clinical study of *PAM-03* published in 2017 included 25 breast cancer patients and revealed the typical centripetal vessel arrangement around the tumor in 61% of invasive disease, compared to only 35% cases of ductal carcinoma in situ (DCIS) lesions. Moreover, additional blood vessel signals were observed in one tumor imaged post a chemotherapy compared to a pre-treatment image, suggesting the ability of *PAM-03* to observe the therapy-related increase in tumor blood flow before any apparent changes to the tumor size.

Yet another design utilized in bed-based devices is a circular or semi-circular detector arrangement. The first scanner of this type, *LOIS-64*, was introduced in 2009 by the Oraevsky group [328]. A semi-circular array of 64 transducers resides in a plane perpendicular to the chest wall (Fig. 7.2).

Using a 757 nm Q-switched Alexandrite laser for illumination, *LOIS-64* has shown the tumor as a bulk of OA contrast in 18 out of 20 malignant cases imaged in the initial clinical study. The authors have demonstrated that OAT can detect tumors also in dense breasts which are inaccessible to XRM. In 2012, a scanner with a circular detector array parallel to the chest wall was demonstrated by a group from University of Florida [329, 330] (Fig. 7.2). Using three wavelengths (733, 775, and 808 nm), the authors unmixed the oxy- and deoxyhemoglobin concentrations to obtain functional OAT images. Clinical study on ten participants has shown that the functional OAT images revealed patterns consistent with MRI imaging. In 2018, another scanner with a circular detector construction has been presented by a Caltech group, called single-breath-hold photoacoustic tomography of the breast (*SBH-PACT*) [48]. Featuring 512 detectors (2.25 MHz central frequency) with an elevational scanning capability (Fig. 7.2), the scanner records 3D images of the breast in high 255 μm in-plane resolution within ~15 seconds. The produced images are obtained with a single wavelength illumination (1064 nm) and show in great detail the vascular networks inside the breast. An initial clinical study on one healthy volunteer and seven breast cancer patients has revealed increased vessel density at the tumor sites in eight of the nine malignant lesions, indicative of present angiogenesis. This observation was, again, unaffected by the breast density, supporting the complementary role of OAT to XRM.

Despite the range of designs presented to date, bed-based scanners remain limited by the cost and difficulty of interfacing a complex 3D scanning geometry to the human breast. Moreover, their integration into the existing breast examination workflow may be challenging since dedicated scanners require separate examinations. Therefore, despite the attention given to bed-based scanners, handheld systems can offer ubiquitous OA examination, especially since they can be seamlessly integrated with handheld ultrasonography, which is routinely employed in clinical breast examinations, adjunct to x-ray mammography [221-223]. US and OA utilize different contrast mechanisms, and thus capture complementary morphologic and functional features of a tumor and the surrounding tissue that could enhance the performance of the examination, while fitting seamlessly into today's clinical workflow.

Handheld OPUS imaging has been demonstrated by two prominent systems. The *Imagio*® scanner (Seno Medical Instruments Inc.; Texas, USA) uses a linear transducer array with 128 piezo elements and a pair of lasers delivering light at two wavelengths (755 nm and 1064 nm), providing co-registered OA and US images (Fig. 7.1). *Imagio*® achieves axial and lateral resolutions of 420 μm and 730 μm, respectively, relying on the filtered back-projection reconstruction algorithm [108]. This scanner was employed in two large, multi-center studies, for adjustment of grading of suspicious breast lesions via predefined semi-quantitative OA features [55, 56] and identification of cancer subtypes [331]. Based on those studies, *Imagio*® recently became the first OAT scanner with an FDA approval for use in breast cancer patients [95].

The *Acuity Echo*® multi-spectral optoacoustic-ultrasound scanner (iThera Medical GmbH; Germany), employed in the following study, features a curvilinear transducer array with 256 piezo elements and a 145° angular coverage (Fig. 7.1; TUM system). The curvilinear design affords superior imaging quality compared to linear arrays [66]. Moreover, using a fast-tunable laser, the *Acuity* scanner enables the collection of 28 images at different wavelengths within 1.1 seconds, offering high spectral definition in the 680−980 nm wavelength range while minimizing motion artifacts. Images reconstructed by filtered back-projection are delivered to the operator in real time, but simple model-based reconstructions computed off-line have also been used to achieve better image quality with axial and lateral resolution of 320 μm and 510 μm, respectively [176]. Variants of the *Acuity* scanner have been successfully utilized in clinical studies of melanoma metastatic status [74], Crohn's disease [70], brown fat metabolism [69, 332], normal vasculature [333] and vascular malformations [64], thyroid disease [82], systemic sclerosis [334], and Duchenne muscular dystrophy [335]. In two pilot breast cancer studies, multi-spectral optoacoustic

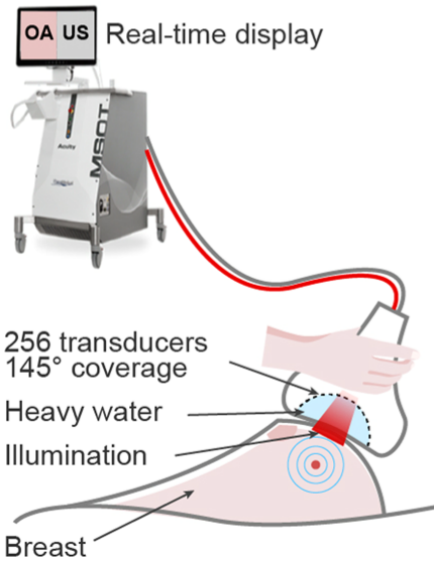

Figure 7.3: Visualization of the Acuity Echo scanner and the handheld probe during the data acquisition. US transducers are arranged in a 145° arc. The arc cavity is filled with heavy water. Laser light is delivered through an optical fiber bundle and a diffuser. Ultrasound (US) and optoacoustic (OA) images are displayed on the scanner screen in real time. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

tomography (MSOT) revealed increased vascularization in the periphery of tumors and a concomitant reduction in the tumor core, as well as heterogeneous total blood volume and irregular deoxy-hemoglobin (Hb) and oxy-hemoglobin (HbO_2) signal patterns in the tumor area [46, 49].

For a more in-depth overview of the field of the OAT breast cancer imaging, the reader is encouraged to refer to existing review articles: [104, 263].

7.1.2 Methods

OPUS setup. *In vivo* measurements were acquired with an *Acuity Echo*® (iThera Medical GmbH, Munich, Germany) hybrid handheld multispectral optoacoustic and ultrasound scanner (Fig. 7.3). Tissue illumination was performed in short light pulses (duration ~8 ns) that were produced by a tunable laser and delivered to the probe via an optical fiber bundle. A diffuser was used to produce a rectangular illumination spot on the skin (ca. 0.5×3 cm). The peak pulse energy was ≈16 mJ, which is within the permissible energy exposure limits set by the American National Standards Institute [113]. Fluctuations in the laser power were compensated by an inbuilt amplitude correction mechanism. Ultrasound signals, produced upon absorption of light energy in the tissue through the optoacoustic effect, were received by 256 piezoelectric elements (4 MHz central frequency) arranged to a curved linear array with 145° coverage and 6 cm distance between the endpoints. The cavity inside the arc was filled with heavy water (D_2O) to ensure acoustic coupling while minimizing absorption of near-infra-red light. We recorded images at 28 separate wavelengths between 700 and 970 nm at 10 nm intervals using a single pulse-per-image acquisition with framerate of 25 Hz; acquisition of a full multispectral OA frame consisting of 28 single-wavelength images took 1.1 s. Pulse-echo ultrasound images were acquired synchronously with the OA images during the pauses between individual laser pulses. The synthetic transmit aperture method with spatial compounding of sub-apertures was used [336]. The US images were acquired at the repetition rate of 6.25 Hz.

Both the US and the OA images were displayed in real-time on the screen of the scanner. These images were reconstructed using a delay-and-sum algorithm, which can run in real-time but

Table 7.1: Overview of the analyzed masses (n=22) by their type and cancer grade distribution.

Lesion type	Grade: G1/G2/G3/Gx	Total
Benign		6
Fibroadenoma		4
Fibrosis		2
Malignant		16
No special type (NST)	2 / 7 / 2 / 1	12
Lobular	1 / 1 / 0 / 0	2
Ductal-lobular	0 / 1 / 0 / 0	1
Inflammatory	0 / 1 / 0 / 0	1

produces images of inferior quality compared to off-line model-based reconstructions. To enable high-quality off-line reconstructions, the raw optoacoustic signals were stored.

Image processing. High-quality OA images were obtained via the 2G-OPUS pipeline, consisting of band-pass filtering, model-based reconstruction, frame averaging, and dual-band visualization with local contrast normalization and sigmoid normalization. Hybrid OPUS images were obtained in a similar fashion to dual-band visualizations, using OA (LCO_{100}) and US images as two channels mapped to RGB color space using the following colormap:

$$\boldsymbol{C}_{\mathrm{OPUS}} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{7.1.1}$$

Study protocol. The aim of this clinical study was characterization of the features extracted by hybrid 2G-OPUS imaging from breast tumor tissue using the 2G data processing pipeline. We

Table 7.2: Overview of the cases described in detail herein (Figures 7.4–7.6). Size denotes the diameter of the hypoechoic tumor core in the selected image, depth refers to the distance between the tumor center of mass and the skin surface. Estrogen receptor and progesterone receptor statuses are reported in percent, except Case 3 which was rated by a different histopathologist and is reported as immunoreactive score. All patients were female. **Abbreviations:** ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor 2; ILC, invasive lobular carcinoma.

						Immunohistochemistry		
Case	**Age**	**Lesion type**	**Size (cm)**	**Depth (cm)**	**Grade**	**ER**	**PR**	**HER2**
Case 1	75	ILC	1.5	1.3	G1	100	100	Neg.
Case 2	61	No special type	3.1	2.2	Gx	70	10	Neg.
Case 3	51	No special type	3.2	1.6	G2	9/12	6/12	Pos.
Case 4	76	No special type	0.8	0.8	G3	3	0	Pos.
Case 5	27	Fibroadenoma	1.7	0.8	N/A	N/A	N/A	N/A
Case 6	50	No special type	1.1	2.3	G2	10	0	Neg.
Case 7	49	No special type	0.8	0.8	G2	100	100	Neg.
Case 8	62	Inflammatory	1.1	2.1	G2	10	0	Neg.

employed scans from 22 female patients (age range 21–79) with breast tumors that were clearly visible in the US and the OA images and were accompanied with complete information for each scan. The scans were performed by a senior radiologist experienced in breast ultrasonography who received additional training for 2G-OPUS imaging. The study was approved by the local ethics committee of the Technical University of Munich (Nr. 27/18 S) and all participants gave written informed consent upon recruitment. The types of scanned tumors are summarized in Table 7.1.

Patients were scanned in the supine position in a quiet room with normal temperature (≈23°C). Standard ultrasound gel was used to ensure acoustic coupling between the probe and the tissue. Selection of the optimal field-of-view was performed manually by the operator as in a routine ultrasound examination, relying on the simultaneous co-registered ultrasound imaging provided by the hybrid 2G-OPUS setup for anatomical navigation. Each tumor was scanned for several seconds to record multiple frames. Nevertheless, the patients were not required to hold their breath during the examination period. The total examination length did not exceed 15 minutes, including patient preparation.

Irrespective of the study participation, all patients underwent surgical removal of the tumor as part of the planned treatment. Histology images shown herein were obtained from the excised, paraffin-embedded tissue and stained with hematoxylin-eosin for standard histopathological analysis. Evaluation of the images for our study was performed retrospectively by a senior pathologist.

7.1.3 Results

Despite the state-of-the-art hardware specifications of the *Acuity Echo*® scanner, which features a curvilinear detector and fast wavelength switching ability, the image quality of handheld scanners is challenged by the limited view angle, which reduces the resolution and contrast achieved and introduces imaging artifacts [337, 338]. In Chapter 4, we have presented the 2G-OPUS image processing pipeline, aiming to achieve the next level of image quality and accuracy in handheld OA imaging. To demonstrate the benefits of the improved image quality and evaluate the clinical potential of the technology, we have conducted a clinical study involving n=22 patients. We inspected the obtained OA images of breast lesions, grouped the observations that correspond to malignancy, and linked the findings to histological analysis and the clinical description of the lesions imaged. Figures 7.4–7.6 exemplify our findings on eight representative cases (see Table 7.2). Fig. 7.7 contains clinical images of the eight cases obtained with standard radiological modalities (XRM, US, MRI) to provide additional details about the displayed masses.

Case 1 (Fig. 7.4a–d; used also for the demonstration of image improvement in Chapters 4 and 5) is an invasive lobular carcinoma. Dual-band visualization (Fig. 7.4a) showed high vascular density surrounding the tumor core, with elongated blood vessels (arrows), up to 2 mm in diameter, starting more than 1 cm away from the lesion and centripetally arranged around the tumor mass. Moreover, vascular contrast in a patchy distribution is seen within the tumor mass. Histological analysis identified a lesion with a well-defined central core and spiculated infiltration typical of invasive lobular carcinoma (Fig. 7.4c). Vasodilation in the tumor neighborhood was histologically confirmed (Fig. 7.4d). The appearance of the OA image is markedly different than that of the ultrasound, showing involvement of a larger part of the breast tissue compared to the ultrasound-based characterization of the tumor extent (Fig. 7.4b; grayscale).

Case 2 (Fig. 7.4e–h) is a hormone-positive mamma carcinoma of no special type (NST). Histology identified a necrotic core and a highly perfused rim region with high cancer cell density (Fig. 7.4g, h) and confirmed the OA appearance showing high vascular density around the tumor rim (Fig. 7.4e, f); one such area is highlighted with an arrow on Fig. 7.4e. Case 3 (Fig. 7.4i, j) is also shown to better illustrate the carcinoma pattern seen in Case 2 and depicts a HER2-positive NST tumor. MSOT reveals a dense vascular bed surrounding the tumor mass (Fig. 7.4i, arrow). Fig. 7.4k

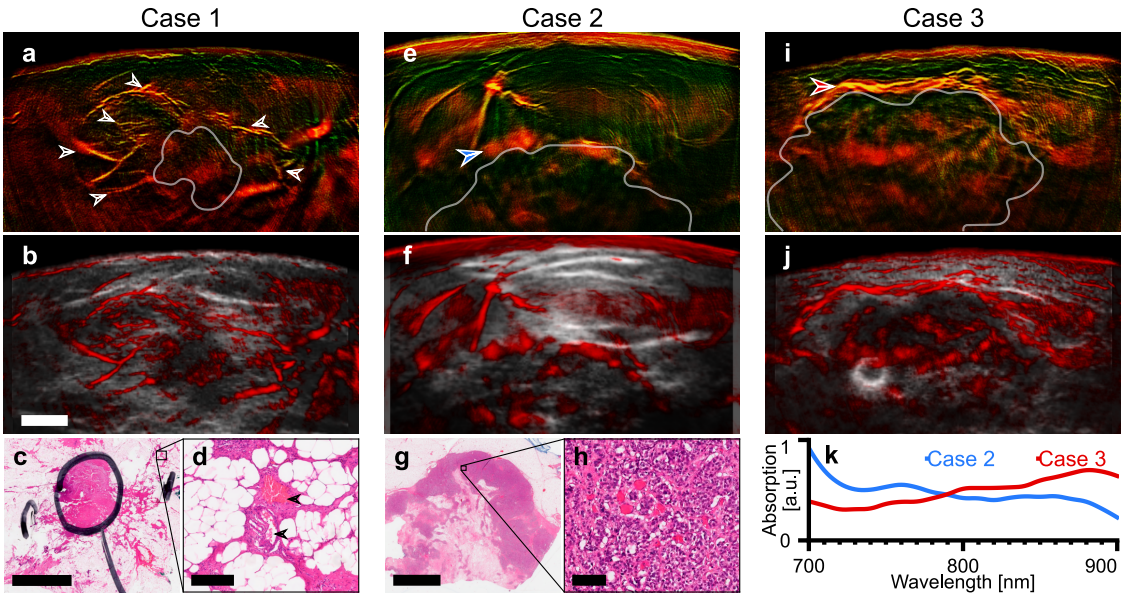

Figure 7.4: MSOT visualizes vascularization and perfusion in the tumor area at high resolution. a–d, Case 1, an invasive lobular carcinoma. Dual-band visualization (a) and OPUS visualization (b) of the median of images in the 850–890 nm range shows dense vascular network (arrows) around the tumor core (white contour, segmented manually in the ultrasound image). Histopathology of the excised tumor (c; H&E stain) revealed that this pattern is likely caused by vasodilation induced by cancer cell infiltration outside of the main core (black pen circle). Magnification cut-out (d) shows blood vessels (arrows) outside the tumor core surrounded by cancer cells. Whereas grayscale ultrasound underestimates the true tumor extent, 2G-OPUS provides additional information about the true cancer spread into the surrounding tissue. **e–h**, Case 2, a hormone-positive mamma carcinoma NST. Dual-band visualization (e) and OPUS visualization (f) of the median of images in the 700–740 nm range shows patches of markedly increased signal in the upper rim of the tumor (e.g., arrow). Histology slice of the excised tumor (g; H&E stain) shows a necrotic center of the mass (result of an earlier therapy) and a region of densely concentrated tumor cells at the lesion upper rim. The high-resolution cut-out (h) shows that the rim region is highly perfused by a dense network of tiny capillaries. The individual capillaries are too small for 2G-OPUS to resolve; instead, the increased perfusion is exhibited as a patch of stronger signal. **i,j**, Case 3, a HER2-positive NST mamma carcinoma. Dual-band visualization (i) and OPUS visualization (j) of the median of images at 850–890 nm reveals marked increase in vascularization in the upper tumor rim. Contrary to the Case 2 shown in (e), the vessels in this tumor rim are larger and can be clearly recognized in the 2G-OPUS image. **k**, Comparison of mean absorption spectra of the enhanced rims in Cases 2 and 3 (denoted by arrows in e and i). The spectrum of rim of Case 2 has a notable signature of deoxyhemoglobin, which indicates reduced function of the capillary network. This observation correlates with the patient's lack of response to neoadjuvant chemotherapy (cT2→ypT3). On the other hand, the spectrum of rim of Case 3 exhibits higher oxygen saturation and suggests better vascular function in supplying the tumor microenvironment. This correlates with this patient's positive response to neoadjuvant chemotherapy (cT2→ypT1c). Scalebars represent 5 mm (b, c, g), 200 µm (d), and 100 µm (h). Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

compares the absorption spectra of the rims in Cases 2 and 3. Case 2 has an absorption spectrum dominated by deoxyhemoglobin, whereas Case 3 shows better oxygenation. This observation suggests that the vasculature in the rim of Case 3 is more functional than in Case 2. Following the MSOT scans, both patients underwent neoadjuvant chemotherapy. The Case 2 patient did not exhibit any response (increase from cT2 to ypT3), whereby Case 3 had a positive response (decrease from cT2 to ypT1c). This finding can be explained by previous observations showing that tumor oxygenation correlates with neoadjuvant chemotherapy outcome [339, 340].

Case 4 (Fig. 7.5a–d) shows a HER2-positive mamma carcinoma NST with irregularly shaped peritumoral structures seen on the OA images (Fig. 7.5a, b, arrows 1–3). These peritumoral structures were found to contain increased hemoglobin signal (see Fig. 7.5e) compared to small

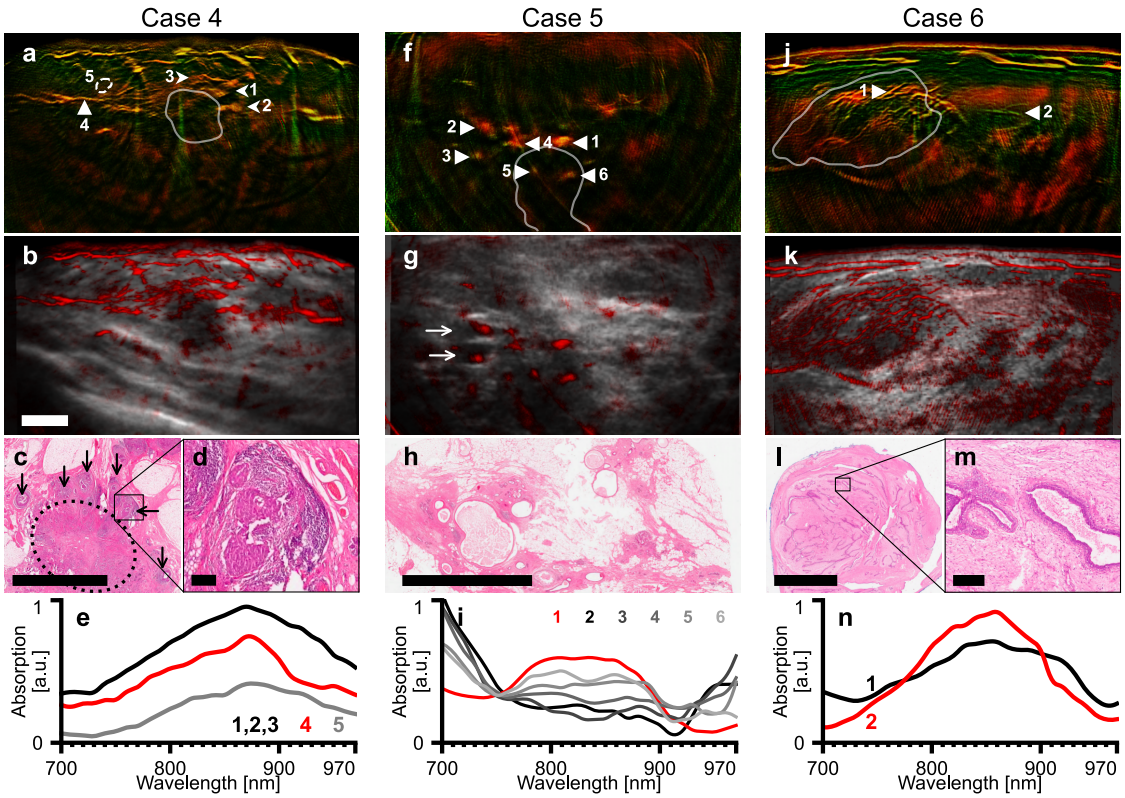

Figure 7.5: 2G-OPUS visualizes signs of inflammation and benign features. a–d, Case 4, a HER2-positive mamma carcinoma NST. Dual-band (a) and hybrid OPUS (b) visualizations of median of images in the 850–890 nm range show numerous structures with irregular shape in the vicinity of the tumor core (arrows 1–3). Histology slice (H&E stain) of the excised tumor (c) shows extensive pre-cancerous proliferation in the ducts (thin arrows) around the invasive tumor (dotted ellipse). The high-resolution cut-out (d) shows accompanying chronic periductal inflammation causing edema and increased blood perfusion. **e,** comparison of mean absorption spectra of the inflammation foci (arrows 1–3), a small blood vessel (arrow 4), and background tissue (circle 5). **f–h**, Case 5, an ER-positive mamma carcinoma NST. Dual-band (f) and hybrid OPUS (g) visualizations of median of images at 700–730 nm show numerous patches of increased signal around the tumor site (arrows in f). Patches 2 and 3 are colocalized with hypoechoic regions on the US (arrows in g). Histology slice (H&E stain) of the excision at the tumor site (h) shows numerous cystically dilated ducts filled with proteinaceous fluid and foam cells. **i**, comparison of absorption spectra of the six patches marked by arrows in f. Whereas the spectrum of patch 1 resembles hemoglobin and corresponds to a cross-sectionally imaged blood vessel, patches 2–6 have a markedly different absorption spectrum appearing as mixture of deoxyhemoglobin and water. **j–m,** Case 6, a fibroadenoma. Dual-band (j) and hybrid OPUS (k) visualizations of median of images in the 700–740 nm range show numerous tubular branching structures inside the tumor core (white contour). Histology slice (H&E stain) of the excised tumor (l) shows a typical pattern for fibroadenomas: dense stromal proliferation with a network of compressed ducts. These ducts match the observed pattern in the optoacoustic images in both size and density. The high-resolution cut-out (m) shows a detail of a compressed duct filled with proteinaceous fluid. **n,** comparison of absorption spectra of the ducts (1) and a blood vessel in similar depth (2; marked by white arrows in j). The duct spectrum is similar to the blood vessel spectrum, indicating bleeding, occurring possibly in relation to a past compression of the mass. No signs of bleeding are present in the histological sample which was excised 7 months after the scanning was performed. Tumor cores in (a), (f), and (j), denoted by white contours, were segmented in co-registered ultrasound images. Scalebars represent 5 mm (b, c, h, l) and 200 µm (d, m). Panels a, b, f, g, j, k have the same resolution. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

blood vessels (Fig. 7.5a, arrow 4) and to background tissue (Fig. 7.5a, circle 5). To perform this comparison between tumor and background hemoglobin signals, we selected lesions that were approximately the same depth to minimize possible depth-related attenuation effects on the OA signal. Decrease of the OA signals after 870 nm (observed uniformly in all cases) is caused by

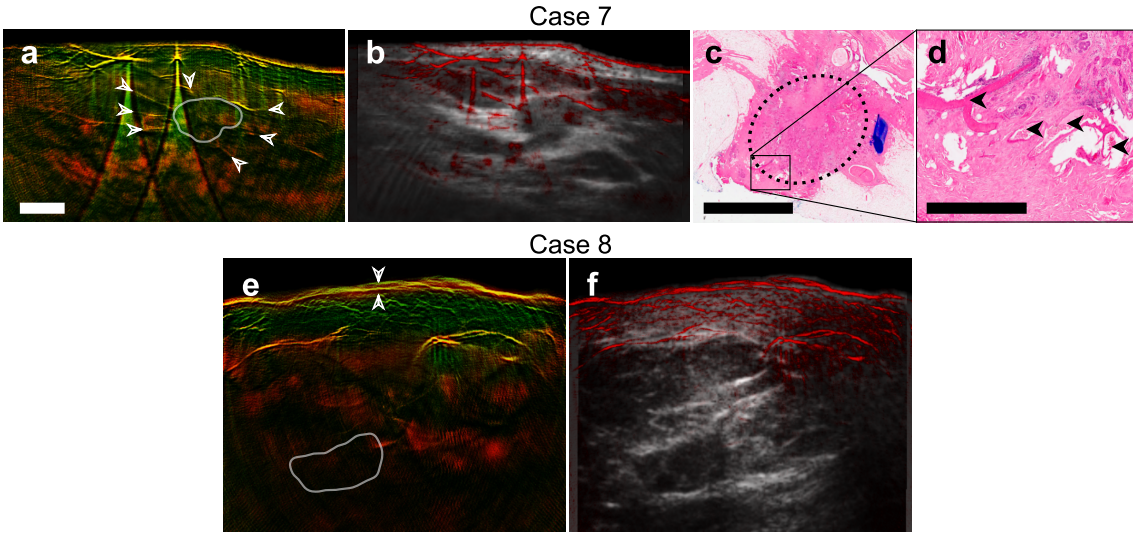

Figure 7.6: Additional cases of optoacoustic features of breast cancer. a–d, Case 7, a HER2-negative mamma carcinoma NST. Dual-band visualization (a) of median of images in the 850–890 nm shows blood vessels (arrows in a) arranged centripetally around the hypoechoic core (segmented in co-registered ultrasound image (b), white contour). Histopathology of the excised tumor (c; H&E stain; marked by dotted ellipse) confirmed presence of numerous vessels (arrows in d) adjacent to the tumor with sizes corresponding to OA observations). **e, f**, Case 8, an inflammatory mamma carcinoma. Dual-band visualization of median of images at 850–890 nm shows thickened cutis (between arrows) and conspicuous dilation of subcutaneous vasculature. This is related to the inflamed appearance of the skin accompanying the tumor proliferation. White contour delineates the tumor core, segmented in a co-registered ultrasound image (f). The scalebar represents 1 mm in (d) and 5 mm in (a) and (c). Panels (a), (b), (e), (f) have the same resolution. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

spectral coloring due to the absorption of lipids and water dominating the longer wavelengths. Histology results (Fig. 7.5c) revealed the presence of numerous pre-cancerous ductal proliferations (black arrows) around the invasive tumor (dotted ellipse), accompanied by chronic periductal inflammation causing edema and increased blood perfusion, as also seen at magnification (Fig. 7.5d). The presence of the inflammation foci explains our observations of the peritumoral structures described above, as their sizes and spatial arrangement match, and increased hemoglobin concentration has been previously observed in chronic inflammation of the gut [70].

Case 5 (Fig. 7.5f–i) shows an ER-positive mamma carcinoma with notable cystic dilation of the ducts around the tumor site visible in the histology slice (Fig. 7.5h) as well as the conventional US (Fig. 7.7). OA images of this tumor (Fig. 7.5f, g) show numerous patches of increased signal (arrows 1–6), as opposed to previous observations from tumors. The patch 1 corresponds to a cross-section of a larger blood vessel, well visible in the contrast enhanced magnetic resonance image (Fig. 7.7), and its spectrum is dominated by oxyhemoglobin (Fig. 7.5i). The mean absorption spectra of patches 2–6 are different than patch 1 or those taken from peritumoral lesions and attain a spectral profile representative of a mixture of deoxyhemoglobin and water, similar to absorption profiles of breast cysts previously reported using optical imaging [341-343]. Some of the patches are colocalized with anechoic capsules on the US (Fig. 7.5g, arrows), supporting our hypothesis that they correspond to cysts or cystically dilated ducts. Compared to the optical imaging studies, the high resolution and spectral contrast afforded by our OA image formation pipeline enables more accurate signal interpretation and the differentiation of cysts from blood vessels.

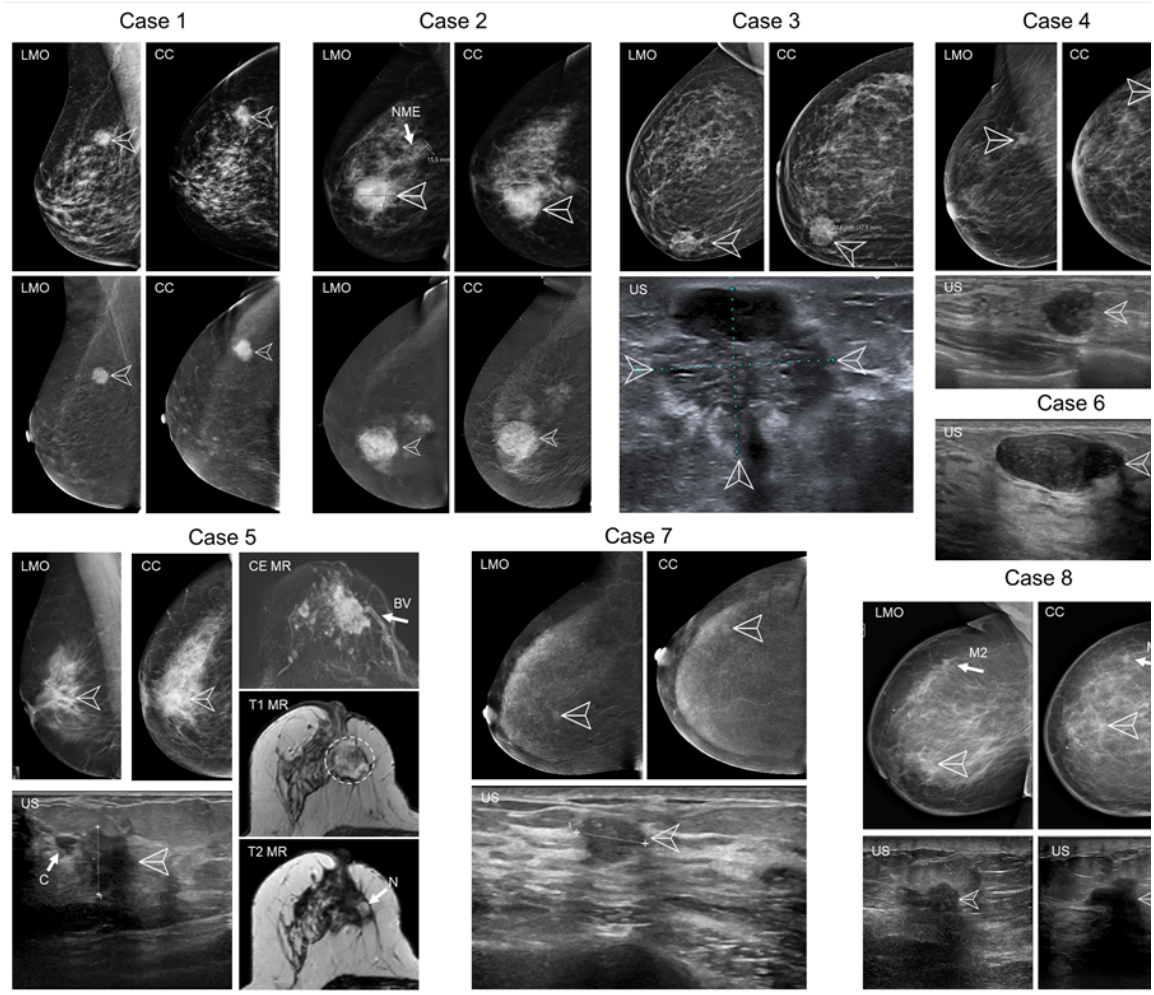

Figure 7.7: Conventional clinical images of the eight cases presented in the results. Lesions are marked by arrowheads. **Case 1:** Contrast-enhanced mammography of the right breast showing the enhancing mass with indistinct borders in the upper outer quadrant in lateromedial oblique (LMO) and craniocaudal (CC) views. **Case 2:** Contrast-enhanced mammography (LMO and CC) of the right breast showing a mainly circumscribed, partially indistinct tumor and adjacent non-mass enhancement (arrow NME) without clear findings in the Low energy (upper) image. **Case 3:** Mammography (LMO and CC) and a conventional ultrasound (US) image of the tumor containing well circumscribed hypoechoic parts (either cystic or containing more aggressively growing cells) and some spiculated, less hypoechoic infiltrating parts. This example reflects tumor heterogeneity. **Case 4:** Mammographs (LMO and CC) and a conventional ultrasound image of the microlobulated round tumor. **Case 5:** Mammographs (LMO and CC), a conventional ultrasound, contrast-enhanced, T1- and unenhanced T2-weighted magnetic resonance (MR) images of the partially necrotic tumor. A large feeding blood vessel subtracted CE-MR, arrow BV) is visible near the lesion in a position matching the observation in the OA image (Fig. 7.5F, arrow 1). Necrosis is visible in the tumor core (T2 image, arrow N) usually reflecting a very fast tumor growth. The ultrasound shows a cyst or a dilated duct near the tumor core (arrow C). **Case 6:** An ultrasound image of a fibroadenoma. The lesion has a typical appearance: hypoechoic, oval-shaped, with a uniformly distributed texture and smooth borders. **Case 7:** Contrast-enhanced mammography (LMO and CC) and an ultrasound image of the breast containing the tumor. The 10 mm large lesion is oval to round, with unsharp borders and a disruption of some anatomical structures. **Case 8:** Mammography (LMO and CC) and ultrasound images of the breast with a recurring inflammatory mamma carcinoma with skin thickening and two masses: arrowheads mark the mass imaged in this study, appearing in the as irregular mass with shadowing features. A secondary mass is marked by arrow M2. Figure reproduced from *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343.*

In Case 6, we also showcase MSOT features revealed from a benign lesion—a fibroadenoma (Fig. 7.5j–n). The OA images show numerous tubular structures within the tumor core (Figs. 7.5j, k). However, there is no dense vascular network around the rim as seen in malignant tumors (Fig. 7.4). Histology slices of this tumor show that the observed structures correspond to numerous compressed ducts filled with proteinaceous fluid that form a linear branching network, a pattern often seen in fibroadenomas (Fig. 7.5l, m). Fig. 7.5n shows a comparison of the mean absorption spectrum in one of the ducts (arrow 1) and a small blood vessel at similar depth (arrow 2). The similar appearance of both spectra indicates the presence of blood in the ducts, possibly due to compression-induced bleeding.

Case 7 (Fig. 7.6a–d) is a hormone-positive mamma carcinoma NST. The MSOT image of the tumor (Fig. 7.6a) as well as hybrid OPUS image (Fig. 7.6b) exhibit small vessels in a centripetal arrangement around the tumor core (arrows), similar to Case 1. Histopathology (Fig. 7.6c, d) confirms the presence of numerous vessels (arrows) around the tumor (ellipse) with sizes matching our optoacoustic observations.

Case 8 (Fig. 7.6e, f) is an instance of inflammatory breast cancer (IBC). Although IBC is not an actual inflammation, it is typically accompanied by symptoms resembling acute mastitis, including skin thickening and redness. MSOT captures both symptoms: the cutis layer (between arrows) is considerably denser than in other scans (cf. Figs. 7.6a, 7.4e, 7.5j) and conspicuous dilation of subcutaneous vasculature correlates with observed erythema.

Table 7.3 summarizes the frequency of the observed patterns in the OA images across all the 22 analyzed lesions. Due to the exploratory nature of our analysis and the small sample size, we do not draw direct conclusions regarding the diagnostic value of these patterns.

7.1.4 Discussion

In this study, we utilized our new image processing platform and set a new mark in the image quality of handheld optoacoustic breast cancer imaging. The unprecedented image quality, achieved by 2G-OPUS, allowed the characterization of OA features in malignant and benign breast tumors. We examined 22 breast lesions and linked observed patterns to available clinical data and micrographs of the excised tissues. We could identify patterns of malignancy, such as enhanced rim and vascular density and functional parameters, in particular oxygenation/hypoxia, and showcased representative examples seen in eight cases.

We demonstrated that using improved image processing tools—reconstruction models with TIR correction, averaging of motion-corrected frames, and color-coded visualization of two frequency bands—we could achieve optoacoustic imaging quality never before demonstrated with a handheld scanner (*Acuity Echo*®). Our approach showed the ability to resolve blood vessels with diameters as small as 200 µm at depths up to 2 cm *in vivo* (see Case 1—Fig. 7.4 and Case 7—Fig. 7.6).

Table 7.3: Summary of presence of observed optoacoustic patterns in analyzed lesions grouped by the lesion malignancy status. Our study focuses on better understanding of the patterns visible thanks to high spatial resolution and spectral contrast of 2G-OPUS. We do not draw direct conclusions from this study regarding the diagnostic value of these patterns.

Lesion type	Malignant (n=16)	Benign (n=6)
Rim enhancement	5	0
Blood vessels Ø > 1 mm	5	2
Blood vessels Ø < 1 mm	13	5
Cysts, ducts	2	3

Comparable imaging quality has been so far reported only with dedicated bed-based scanners *SBH-PACT* [48] and *PAM-03* [47], which nevertheless yielded lower axial resolution (255 μm and 370 μm, respectively). Unlike stationary imaging systems, handheld optoacoustic imaging can be seamlessly integrated in routine breast ultrasound [221-223], improving the information obtained in the imaging session.

Malignant tumors generally exhibited patterns of small blood vessels arranged centripetally around the tumor core, a finding that has been previously described as one of the most reliable OA imaging features to indicate malignancy [47, 54, 56, 58]. Compared to previous studies however, the superior image quality achieved herein facilitated detailed visualization of this feature, which could increase the confidence in diagnosis of suspicious borderline lesions, thus reducing the number of unnecessary biopsies. In a smaller number of cases, the findings also demonstrated the presence of dilated vasculature surrounding the tumor core. In Case 1, examination of the tumor histopathology revealed that the vasodilation was caused by infiltration of the cancer cells into the surrounding tissue. Whereas standalone US often underestimates the size of invasive lobular carcinoma [344, 345], our results show, in line with earlier *ex vivo* experiments [346], that optoacoustics can help with assessment of tumor margins *in vivo*. 2G-OPUS could also characterize functional parameters associated with the tumor rim microenvironment. In Cases 2 and 3 we observed different oxygenation levels in the rim, reflecting the functional condition of these vessels, and corresponding differences in NAT response. Compromised vascular functionality exhibited through hypoxia reduces the efficiency of therapeutic drug delivery [347, 348] and has been shown to correlate with poor NAT outcomes [339, 340]. This observation indicates a potential for 2G-OPUS to offer detailed functional images serving as a predictor for NAT outcomes.

We further observed features associated with benign lesions, such as cysts, and chronic inflammation around ducts containing pre-cancerous proliferations. Although the relationship between inflammation and invasive progression of DCIS is not fully understood [349], some studies show that the presence of chronic inflammation relates to increased risk of recurrent invasive disease [350, 351]. MSOT has been used to monitor chronic inflammation of the gut [70, 352]; however, this is the first time it has been used to visualize chronic inflammation in the human breast. Although the inflammation could not be distinguished from malignancy based on functional features such as oxygen saturation, a combination of 2G-OPUS with traditional modalities could facilitate identification of such risky DCIS cases early and could support better treatment decisions.

Remaining limitations of handheld 2G-OPUS include the imaging depth limited to ~2 cm due to light scattering, which can be mostly remedied in breast imaging by positioning the handheld probe in a favorable location and applying mild pressure to bring the tumor into the FOV. Moreover, speeding up the image processing pipeline to provide high-quality OA images to the operator in real-time would increase the efficacy of handheld OA breast cancer examinations. Specialized hardware implementations or deep-learning-based acceleration are promising directions towards that goal. Furthermore, the motion artifacts corrupting the spectral information could be further remedied by recording less wavelengths to shorten the acquisition (e.g., omitting wavelengths beyond 930 nm). However, recording many wavelengths facilitates more reliable spectral unmixing, and finding the optimal trade-off between the imaging speed and the spectral information remains an open problem. Finally, while this study focused on characterization of tumor features, further research including healthy subjects is needed to fully understand their reproducibility, and diagnostic and predictive value. Overall, the combination of improved image processing with handheld, hybrid optoacoustic-ultrasound acquisition, label-free contrast, and fast operation make 2G-OPUS well suited for incorporation into the established breast cancer examination workflow. It expands the capabilities of routinely used conventional grayscale ultrasound with the ability to image additional pathophysiological features. Simultaneous hybrid acquisition of ultrasound along with optoacoustic images allows easy localization of tumors during

scanning. Future improvements to the quality of the built-in ultrasound might eliminate the need for two separate examinations, thus increasing patient comfort. Additionally, the rich spectral information provided by 2G-OPUS imaging allows label-free resolution of endogenous chromophores absorbing in the near-infrared range, primarily oxy- and deoxyhemoglobin, which can be used to estimate oxygen saturation of blood. Since 2G-OPUS can visualize these features within the established clinical routine by expanding the capabilities of common grayscale ultrasound, we see great potential for its translation into standard clinical breast care.

7.2 Cervical lymph node metastases

In majority of human carcinomas, cancer cells spread from the primary tumor initially to regional lymph nodes (LN) and form metastases there [353]. Presence of LN metastases is one of the most important prognostic factors and forms the basis of cancer staging [354]. For example, the nodal involvement in oral cancer patients is linked to a decreased survival rate by as much as 50% [355, 356]. Staging of LN involvement in oral cancer patients is performed using standard radiological techniques such as CT, MRI, or US. Although imaging is more reliable at metastasis detection than palpation, these methods still fail to detect micrometastases [357] and 20–30% of patients with clinically negative neck (cN0) harbor occult nodal metastases. When the probability of occult metastases is >20%, elective neck dissection (END) has been advocated despite the cN0 stage, assuming that early detection and treatment of occult metastases leads to a better oncological outcome [356, 358, 359]. However, this approach results in a considerable overtreatment of patients without any metastases, and the END procedure is associated with frequent shoulder dysfunction morbidity, impairing the quality of life [360, 361].

As an alternative to non-invasive clinical imaging, SLN biopsy can be performed. SLN is the first LN draining the primary tumor site and thus the first LN to harbor a metastasis. Pathological evaluation of the biopsied sentinel LN is thus a good predictor for the metastatic status of the remaining cervical LNs [362]. Localization of the SLN involves a peritumoral injection of a radioactive tracer followed by a gamma camera examination. Whereas the SLN biopsy has a high sensitivity of 87–92% and a negative predictive value of 94–96% [356], it is a time-consuming and invasive procedure with considerable morbidity and exposing both the patient and the caretaking personnel to radiation [363]. As such, there remains a need for improved non-invasive pre-operative detection of LN metastases.

Current imaging methods for cervical LN metastasis detection rely mainly on the LN size, the presence of central necrosis, the appearance of extranodal extension, or the contrast enhancement [364]. However, these morphological changes are preceded by biochemical changes triggered by the metastasizing cells. By targeting the early biochemical changes, molecular imaging approaches may have a diagnostic advantage and higher sensitivity to detect early-stage LN metastases. OAT is well-suited to the task at hand, as it resolves endogenous and exogenous optical contrast, provides images at sufficiently high resolution, and can achieve depth penetration of several centimeters required for imaging cervical LNs.

Indeed, imaging LNs with OAT has been described in multiple works. With the aid of optical contrast agents, optoacoustic detection of SLNs [90, 365] and micrometastases [366] has been achieved in the pre-clinical setting. Moreover, optoacoustic scans of a mouse model of oral cancer have revealed significantly reduced background-subtracted oxygen saturation in metastatic LNs [367]. Guggenheim *et al.* have obtained high-resolution volumetric images of healthy excised human LNs showing the intranodal vasculature and the lipid layer surrounding the LNs [368]. Using the clinically approved optical agent indocyanine green, Stoffels *et al.* were able to localize

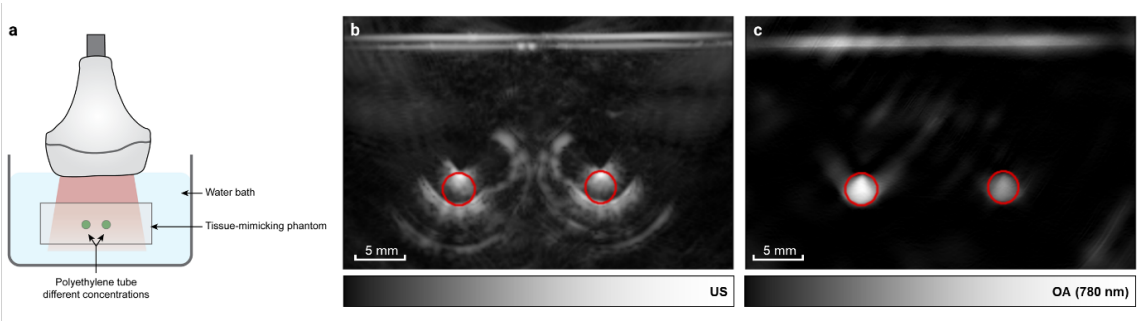

Figure 7.8: Phantom experiment. a, Schematic overview of the phantom experiment. **b,** Ultrasound image of the phantom with two embedded tubes (marked by red circles). Acoustic impedance difference of the tubes and the surrounding agar causes reverberations producing visible artifacts. **c,** Optoacoustic image (λ=780 nm) of the two tubes (marked by red circles) with highest concentrations of cetuximab-CW800. Abbreviations: US, ultrasound; OA, optoacoustics. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

SLNs *in vivo* with MSOT [74, 75]. However, optoacoustic evaluation of the nodal metastatic status in patients *in vivo* remains an open problem.

A possible avenue towards detection for nodal metastasis detection with OAT is the use of tumor-specific optical contrast agents. Multiple studies have shown the use of Epidermal Growth Factor Receptor (EGFR)-targeted fluorescent tracers (e.g., panitumumab-800CW and cetuximab-800CW) to detect malignant LNs *ex vivo* in oral cancer patients using fluorescence molecular imaging [369, 370]. Recently, Nishio *et al.* demonstrated the visualization of panitumumab-800CW in excised LNs using MSOT, differentiating the benign LNs from malignant ones [371].

Expanding upon the previous achievements, we are the first to demonstrate the use of 2G-OPUS for *in vivo* imaging of cervical LNs in oral cancer patients. We characterize the optoacoustic properties of the tumor-specific fluorescent tracer cetuximab-800CW in a tissue-mimicking phantom and determine an estimated minimal detectable concentration *in vitro*. Next, we image the LNs *in vivo* before and after cetuximab-800CW administration in patients and identify limitations that prevent a successful detection of the tracer. Furthermore, we analyze the features of the imaged LNs provided by resolving the intrinsic chromophores, HbO_2 and Hb. Although the small sample size prevents a large-scale evaluation, we observe increased variance of Hb distribution in malignant LNs and demonstrate that clinically significant features of LNs can be observed with 2G-OPUS, providing the rationale and need for a larger clinical trial to substantiate these findings.

7.2.1 Methods

Production of cetuximab-800CW. Cetuximab-800CW was manufactured in the University Medical Center Groningen according to good manufacturing-practice guidelines, as previously described [372]: Cetuximab (Erbitux®) and IRDye800CW NHS Ester (LI-COR Biosciences, Lincoln, NE, USA) were conjugated under regulated conditions with a dye/antibody ratio of 1:2. Cetuximab-800CW was formulated in a sodium phosphate solution at a concentration of 1 mg/ml.

Characterization of cetuximab-800CW. Prior to the clinical study, the optoacoustic spectra of cetuximab-800CW and IRDye800CW were determined with a MS-OPUS system. A tissue-mimicking phantom was fabricated using 300 mL deionized water, 2% agarose and 6% intralipid to mimic the optical properties of biological tissue. At 1 cm depth, polyethylene tubes with a diameter of 3 mm were inserted. These were filled with cetuximab-800CW 1 mg/mL and IRDye800CW with an optical density of 2 as a reference. The phantom was placed in a water bath

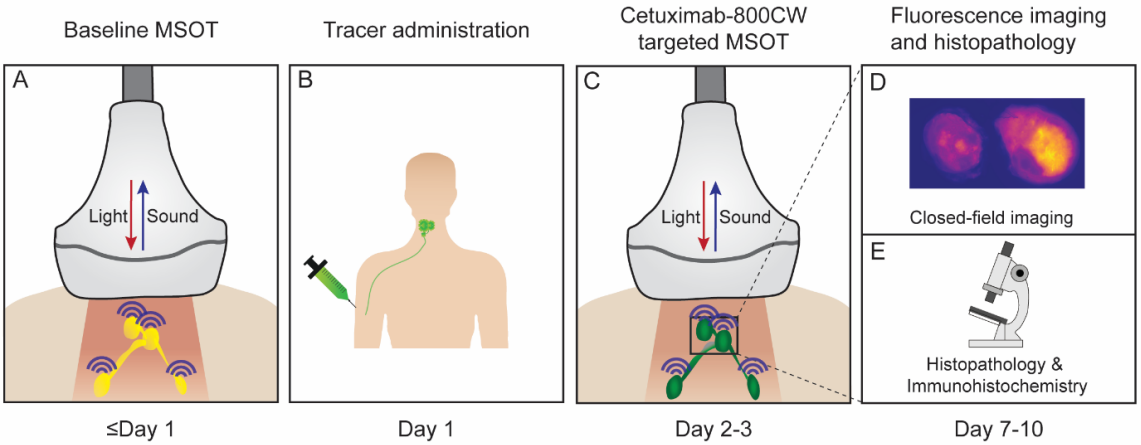

Figure 7.9: Study workflow. After baseline multispectral optoacoustic tomography (MSOT), all patients were intravenously administered with 75 mg cetuximab followed by 15 mg cetuximab-800CW. Two days later, MSOT was performed again. After surgical removal of the lymph nodes, single lymph nodes were imaged during pathology processing of the nodal specimen and correlated with final histopathology. Abbreviations: MSOT, multispectral optoacoustic tomography. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

to ensure optimal coupling between the transducer and the phantom and imaging was performed at 660–900 nm with 5 nm step size. The tubes were manually segmented in the reconstructed optoacoustic images and the spectra of cetuximab-800CW and IRDye800CW were determined as the mean absorption in the segmented regions. Fig. 7.8 shows an overview of the experiment and example US and MSOT images.

To verify the linear relationship between the cetuximab-800CW concentration and the optoacoustic signal strength, and to estimate the minimal detectable concentration, a two-fold dilution series down to 1.6 µM was imaged in the phantom. Six wavelengths were used here (700, 730, 760, 780, 800 and 850 nm) to match the in vivo imaging procedure. Non-negative least squares linear spectral unmixing using spectra of Hb, HbO2, and cetuximab-800CW was applied to compute the cetuximab-800CW concentrations in the optoacoustic images. Again, the mean of the unmixing coefficients in the ROI was computed to estimate the concentration measured by MSOT.

Clinical study. A proof-of-concept clinical study was performed with the objective of visualization of tumor-specific fluorescent tracer cetuximab-800CW using MSOT. The study was performed at the University Medical Center Groningen, Groningen, the Netherlands. Approval was obtained at the Institutional Review Board of the hospital and the Central Committee on Research Involving Human Subjects. The study was performed in compliance with the Dutch Act on Medical Research involving Medical Subjects and the Declaration of Helsinki (adapted version 2013, Fortaleza, Brazil). The trial was registered at clinicaltrials.gov (NCT03757507).

Participants. Patients ≥18 years eligible for inclusion in this study had histology-confirmed oral cancer and were already included in a concurrent clinical trial on fluorescence-guided surgery for margin assessment (ICON-study, NCT03134846). Subjects that underwent a surgical procedure of the neck with a concurrent primary tumor surgery were included, as histopathology was the reference standard of *in vivo* MSOT. All participants provided a written informed consent prior to any study-related procedure.

Exclusion criteria. Patients with any of the following conditions were excluded from participation in the ICON study [373]: concurrent uncontrolled medical conditions, inadequately controlled hypertension, receiving an investigational drug <30 days prior to scheduled tracer administration,

an event of myocardial infarction, cerebrovascular accident, uncontrolled cardiac heart failure, significant liver disease or unstable angina <6 months prior to enrolment. Other exclusion criteria were a life expectancy <26 weeks, Karnofsky performance status <70%, pregnancy, history of infusion reactions to cetuximab or other monoclonal antibody therapies, QTc prolongation (>440 in males and >450 in females) and magnesium, potassium, or calcium deviations of CTCAE grade II and higher.

Study protocol. Fig. 7.9 shows an overview of the study workflow. Included patients were scanned with MSOT to obtain baseline images of the neck LNs. Then, two days prior to surgery, patients were administered with 15 mg cetuximab-800CW preceded by 75 mg unlabeled cetuximab to prevent rapid plasma clearance and to occupy off-target receptors, previously determined as the optimal dosing strategy for primary tumor imaging [373]. The second MSOT scanning was performed before the surgery to obtain the post-administration images of the LNs. Following the surgery, the excised LNs were processed and analyzed by a head and neck pathologist. Clinical and pathological TNM (tumor–node–metastasis) staging was performed according to the 8th edition of the American Joint Committee on Cancer criteria.

OPUS setup. MSOT imaging was performed with a prototype of *Acuity Echo*® (iThera Medical GmbH, Munich, Germany) hybrid handheld multispectral optoacoustic and ultrasound scanner (Fig. 7.3). The scanner comprises of a fast-tunable Nd:YAG laser (25 Hz pulse repetition rate, 4-7 ns pulse duration) and a 256-element 125° arc-shaped ultrasound transducer array (3.4 MHz central frequency). The maximum output energy of this system is in accordance with the American National Standards Institute safety limit for laser exposure [113]. The cavity inside the arc was filled with a gel pad to ensure acoustic coupling while minimizing absorption of near-infra-red light. MSOT images were acquired at six wavelengths (700, 730, 760, 780, 800 and 850 nm) selected to reflect the characteristics of the absorption spectra of Hb, HbO2, and IRDye800CW. Using a single pulse-per-image acquisition with framerate of 25 Hz, acquisition of a full multispectral OA frame consisting of six single-wavelength images took 0.24 s. Pulse-echo ultrasound images were acquired synchronously with the OA images during the pauses between individual laser pulses. The synthetic transmit aperture method with spatial compounding of sub-apertures was used [336]. The US images were acquired at the repetition rate of 6.25 Hz.

The imaging was performed in a dedicated MSOT-imaging room following all safety regulations for safe use of class IV lasers (e.g., laser interlock system, safety goggles). All patients were imaged in a supine position with the neck in hyperextension and turned away from the imaging area. The neck was explored for LNs using the inbuilt ultrasound guidance of the OPUS system. Of all identified LNs, a recording was obtained of ~10 seconds. The anatomical location of each identified LN within the various levels of the neck was mapped. Patients were asked for any symptoms present during or after imaging, and their skin was visually inspected.

Image processing. High-quality OA images were obtained via the 2G-OPUS pipeline, consisting of band-pass filtering (Butterworth filter, 0.5–12 MHz), model-based reconstruction, and frame averaging. The LNs were manually segmented on the ultrasound images generated by the scanner along with MSOT to specify ROIs for further analysis.

A depth-gain correction procedure (Fig. 7.10) was applied to avoid depth-effects biasing our quantitative analysis. Given an original MSOT image (Fig. 7.10a) $I_{\mathrm{orig}}\colon \Omega \subset \mathbb{R}^2 \times \Lambda \to \mathbb{R}$, where Ω denotes the set of image pixels and Λ the set of acquired wavelengths, a depth correction factor c is computed, as a function of depth D (Fig. 7.10b) and wavelength λ, to be the median (Fig. 7.10c) of all pixels in the same depth which are not inside the LN ROI Ω_{LN}:

$$c(D,\ \lambda) = \mathrm{median}\big(I_{\mathrm{orig}}(\boldsymbol{x},\ \lambda) \mid \boldsymbol{x} \in \Omega \setminus \Omega_{\mathrm{LN}} \wedge d(\boldsymbol{x}) = D\big). \tag{7.2.1}$$

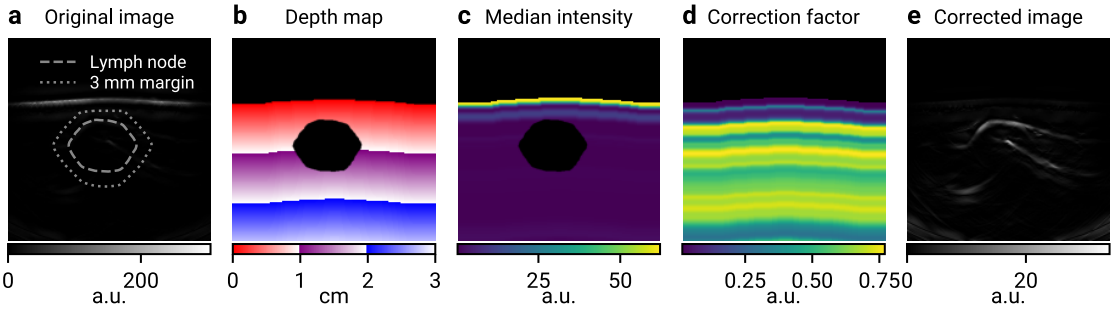

Figure 7.10: Depth correction procedure demonstrated on the image of malignant lymph node (LN) M1. **a,** Uncorrected optoacoustic image (λ = 700 nm) demonstrates the signal intensity decaying with depth. Only the top layer of the skin is visible, all other structures have too low intensity to be seen. The location of the LN and its 3 mm margin are denoted by dashed and dotted lines, respectively. **b,** Depth map. Pixels above the skin surface and inside the LN region are excluded. **c,** Map of median intensities. Each pixel shows the median intensity over all pixels in the image that are equally deep. Pixels in the LN region are excluded from the computation. As in (a), intensity decaying quickly with depth can be seen. **d,** Map of reciprocals to the median intensities shown in (c), which are used as multiplicative correction factors. **e,** Depth-corrected image, obtained by multiplying images (a) and (d), shows that the approach balances the signal intensities over the whole field of view. **Abbreviations:** a.u., arbitrary unit. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

Here, $d(\boldsymbol{x})$ represents the depth of a pixel $\boldsymbol{x}$. Then, the corrected image I_{corr} (Fig. 7.10e) is defined as

$$I_{\mathrm{corr}}(\boldsymbol{x}, \lambda) = \frac{I_{\mathrm{orig}}(\boldsymbol{x},\ \lambda)}{c(d(\boldsymbol{x}),\ \lambda) + \varepsilon}\ , \tag{7.2.2}$$

where ε is a correction term to ensure numerical stability in the regions of the image with near-zero values. In our analysis, ε was set to 1. Map of correction factors $(c + \varepsilon)^{-1}$ is shown in Fig. 7.10d. Furthermore, to limit the artifacts of discretization of the depth map, the function $c(\cdot,\ \lambda)$ was smoothened with a Gaussian kernel with σ = 0.3 mm.

Image analysis. The concentrations of chromophores (Hb, HbO_2, lipids, and cetuximab-800CW) were estimated using non-negative linear spectral unmixing. Specifically, we obtained the vector of concentrations $\boldsymbol{c} \geq 0$ for each pixel $\boldsymbol{x}$ with spectrum $\boldsymbol{s} \equiv [I_{\mathrm{corr}}(\boldsymbol{x}, \lambda_1), \ldots, I_{\mathrm{corr}}(\boldsymbol{x}, \lambda_{|\Lambda|})]^T$ by solving the non-negative least squares problem $\mathrm{argmin}_{\boldsymbol{c}} \|\boldsymbol{A}\boldsymbol{c} - \boldsymbol{s}\|_{\boldsymbol{2}}$, where $\boldsymbol{A}$ is a matrix with the absorption spectra of the four chromophores as columns (see section 2.7).

The distributions of Hb and HbO_2 concentrations in each LN were characterized using their variance relative to 3 mm margins around the LNs (Fig. 7.10a) as $\mathrm{Var}(\mathrm{Hb}_{\mathrm{LN}})\ /\ \mathrm{Var}(\mathrm{Hb}_{\mathrm{Margin}})$, where $\mathrm{Hb}_{\mathrm{LN}}$ and $\mathrm{Hb}_{\mathrm{Margin}}$ are the sets of Hb unmixing coefficients for all pixels in the respective ROIs, and a corresponding formula is applied to HbO_2 as well. This approach has a twofold benefit for the robustness of our analysis. First, the use of relative variance eliminates any subject-dependent linear effects (i.e., constant offset and multiplicative bias) on the results that might be caused e.g., by varying melanin content between the subjects [374]. Second, normalizing by variance of neighborhood, whose size is proportional to the size of the LN, minimizes the possibility that an increased value is observed solely due to increased size of a LN.

Statistical analysis. Statistical analyses were conducted using SciPy (version 1.8.0) [208]. Due to the limited sample size of this study, all data was considered non-normally distributed. The Mann-

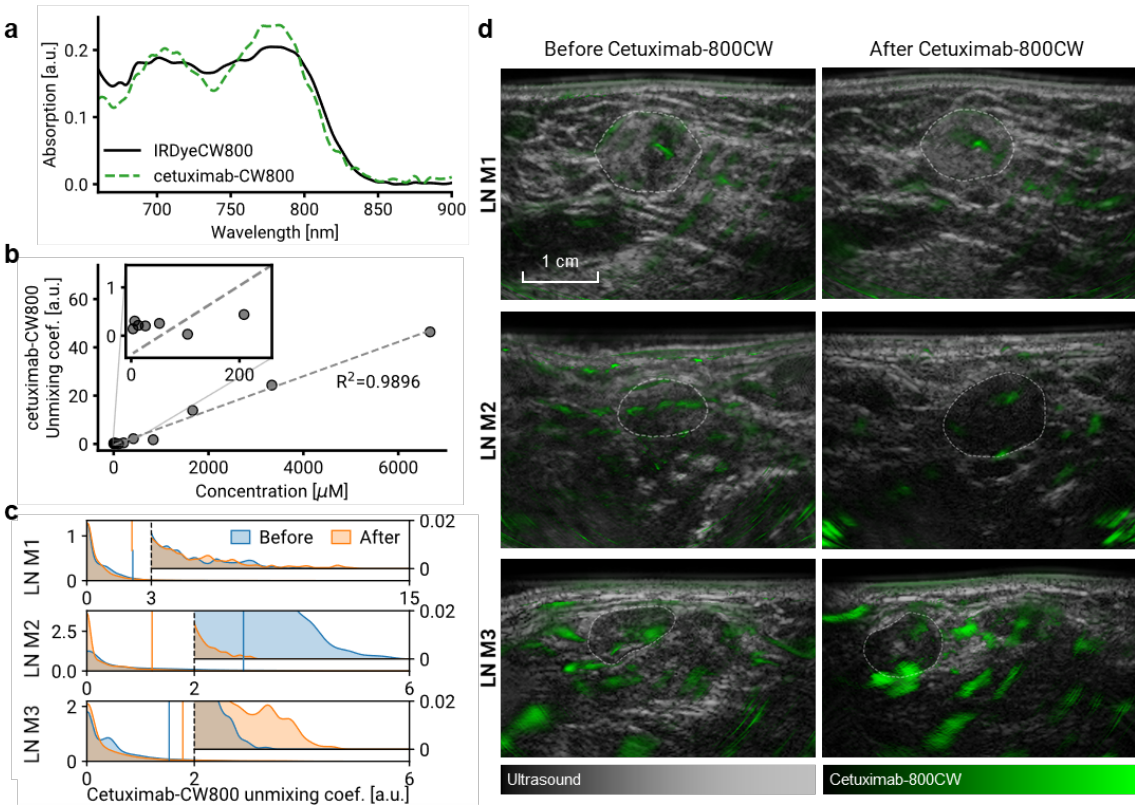

Figure 7.11: Optoacoustic characterization of cetuximab-800CW and its detection in malignant lymph nodes *in vivo*. a, Absorption spectra of IRDye800CW and cetuximab-800CW solutions in a tissue-mimicking phantom, acquired with the MSOT Acuity Echo. **b,** Unmixing coefficient of cetuximab-800CW in optoacoustic phantom images as a function of its concentration, showing a linear relationship. **c,** Distribution of unmixing coefficients of cetuximab-800CW in three malignant lymph nodes *in vivo* before and after injection of the contrast agent. Insets show the right tails of the distributions where a potential localized increase could be observed. Vertical lines denote 95th percentiles. **d,** Hybrid visualizations of cetuximab-800CW unmixing coefficients (green) overlapping greyscale ultrasound, providing both morphological and molecular information on the tissue of interest. Cetuximab-800CW signal cannot reliably be detected with the current setup since it is not specifically present in the malignant LNs. The signal visualized throughout the image is most likely due to errors of linear unmixing, and presumably originates from hemoglobin contrast. In addition, administration of cetuximab-800CW does not result in an apparent increase in optoacoustic signal within the malignant LNs. LN locations are denoted by dashed lines. **Abbreviations:** LN M, malignant LN. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

Whitney U test was used to analyze distributions of HbO_2 and Hb. A p-value <0.05 was considered statistically significant.

7.2.2 Results

Optoacoustic characterization of cetuximab-800CW in a phantom

First, we characterized the spectrum of cetuximab-800CW in MSOT and we verified a linear relationship between the contrast agent concentration and the unmixing coefficients computed from MSOT images. Figure 7.11a displays the spectra of cetuximab-800CW and unconjugated IRDye800CW as recorded by MSOT *in vitro* using a tissue-mimicking phantom. The spectrum of cetuximab-800CW has peaks at 780 nm and 700 nm, the former mirroring a peak of IRDye800CW and the latter related to the forming of H-aggregates [371, 375]. Figure 7.11b shows that the MSOT unmixing coefficients of cetuximab-800CW increase linearly with its concentration in the phantom

(R^2=0.9896) and can be reliably distinguished from the background signals at concentrations above 400 µM. At concentrations below 400 µM, the linear relationship does not hold (R^2<0).

Clinical study participants

Seven patients participated in this clinical study. Table 7.4 summarizes patient demographics and clinical characteristics. All patients received the study drugs and completed the imaging protocol. No adverse events or complaints were reported related to the MSOT procedure. Four patients presented with a clinically negative neck, and three patients with a clinically positive neck. Four patients underwent an END, of which one was extended to a modified radical neck dissection as a malignant LN was intraoperatively identified through frozen section biopsy. Three patients underwent a modified radical neck dissection, of which one also received an END on the contralateral side. The study was ended prematurely as the primary endpoint could be assessed earlier.

Table 7.4: Patient characteristics. Summary of patient and LN characteristics. **Abbreviations:** cTN, clinical tumor and nodal stage; pTN, pathological tumor and nodal stage; N/A, not applicable.

Patient	Sex	Ethnicity	Age	Tumor location	cTN	pTN	# metastases
1	Female	White	65	Tongue	cT2N0	pT2N1	1
2	Female	White	78	Mandible	cT4N0	pT4N0	N/A
3	Female	White	63	Maxilla	cT3/4N0	pT2N0	N/A
4	Female	White	66	Mandible	cT3-4aN1	pT4aN2a	1
5	Male	White	29	Tongue	cT1N0	pT1N0	N/A
6	Female	White	65	Tongue	cT4aN2b	pT4aN3b	7
7	Female	White	47	Mandibula	cT4aN0	pT4aN0	N/A

Table 7.5: Overview of lymph nodes analyzed with MSOT.
Abbreviations: LN, lymph node; N/A, not applicable; US, ultrasound.

LN	Patient	Status	Depth (mm)	Diameter (US; mm)	Diameter (histology; mm)
M1	4	Malignant	4.1	12.9	13.8 x 8.4
M2	6	Malignant	3.7	11.3	10.4 x 9.5
M3	6	Malignant	2.9	8.1	11.4 x 9.3
B1	1*	Benign	12.0	18.7	N/A
B2	1*	Benign	5.7	8.3	N/A
B3	7	Benign	10.4	7.7	N/A
B4	7	Benign	5.0	6.0	N/A
B5	1*	Benign	11.6	20.2	N/A
B6	2	Benign	7.1	16.6	N/A

***** This patient presented with one solitary metastasis in the contralateral side of the neck.

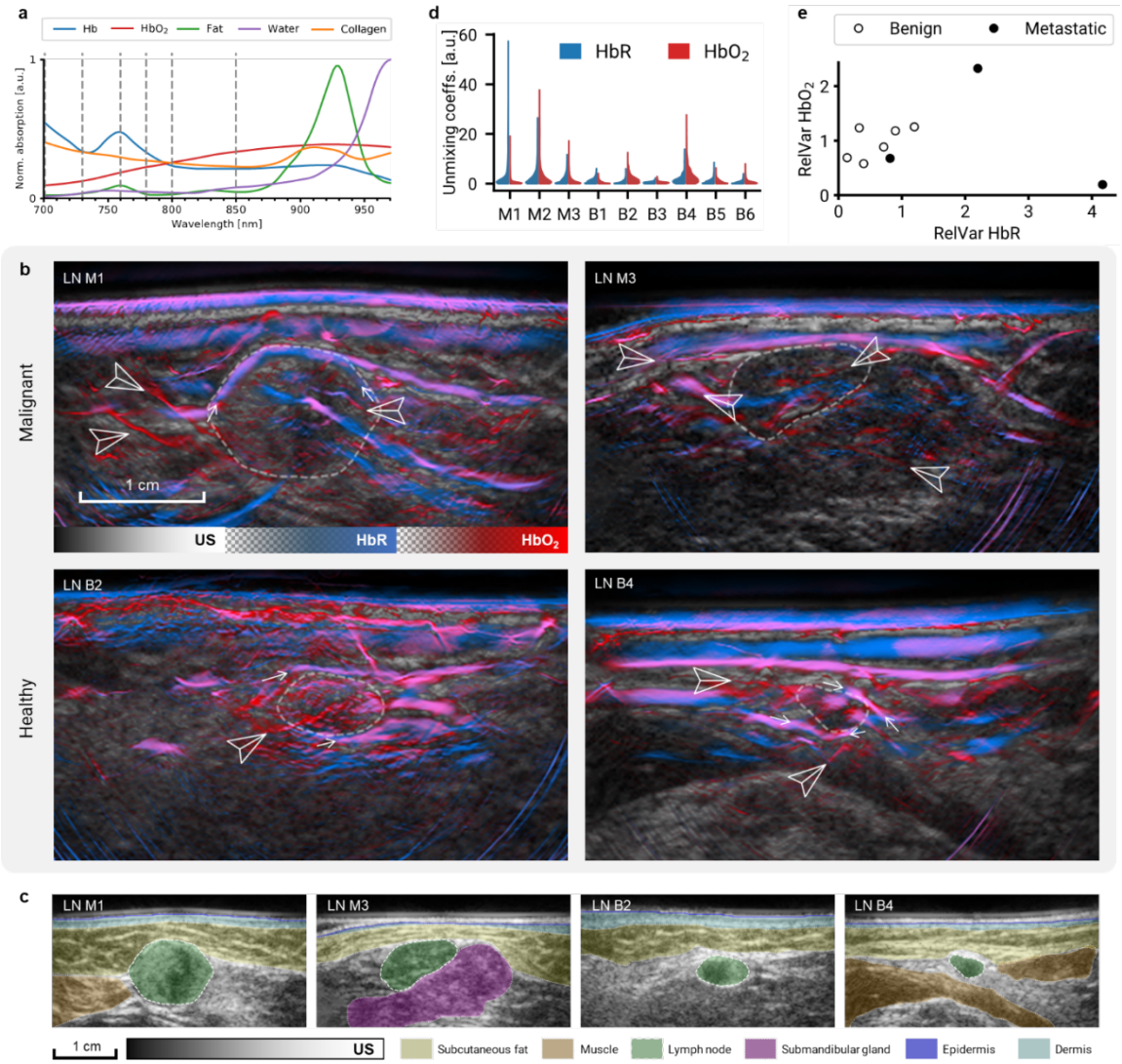

Figure 7.12: Intrinsic contrast in lymph nodes. a, Absorption spectra of main endogenous chromophores in the illumination range of the MSOT Acuity Echo scanner. Dashed lines denote the wavelengths used for the multispectral image acquisition (700, 730, 760, 780, 800, 850 nm). **b,** Ultrasound and linear spectral unmixing images of lymph nodes (LNs) M1, M3, B2, and B4. Arrowheads mark peripheral, feeding, and intranodal blood vessels. Small arrows mark signal enhancement around the LNs. Scalebar applies to all four images. Separate HbR and HbO_2 maps are provided in Fig. S4. Strong HbR signal in the epidermis can be attributed to the presence of melanin. **c,** Ultrasound images of the LNs displayed in (b) with localization of surrounding anatomical landmarks. **d,** Distribution of Hb and HbO_2 unmixing coefficients in three malignant LNs (M1-3) and six benign LNs (B1-6). **e,** Variance of chromophore concentrations in malignant and benign LNs relative to their 3 mm margins. Malignant LNs exhibit large increase of relative deoxy-hemoglobin variance. **Abbreviations:** LN M, malignant lymph node; LN B, benign LN; HbO2, oxyhemoglobin; HbR, deoxyhemoglobin. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

In vivo MSOT of lymph nodes in patients with oral cancer

Due to the nature of the surgical procedure, exact tracking of all imaged LNs between *in vivo* imaging and histopathology was not possible. Only a subset of LNs where the metastatic status could be assigned with certainty was included in the *in vivo* image analysis (Table 7.5). Specifically, we included only LNs of patients showing no malignancies at all at the histopathology, classified as benign (n=11), and malignant LNs that were either palpable or close to anatomical landmarks and thus could be tracked until histopathology (n=3). Five benign LNs that allowed correlation with histopathology were excluded from the analysis due to image quality

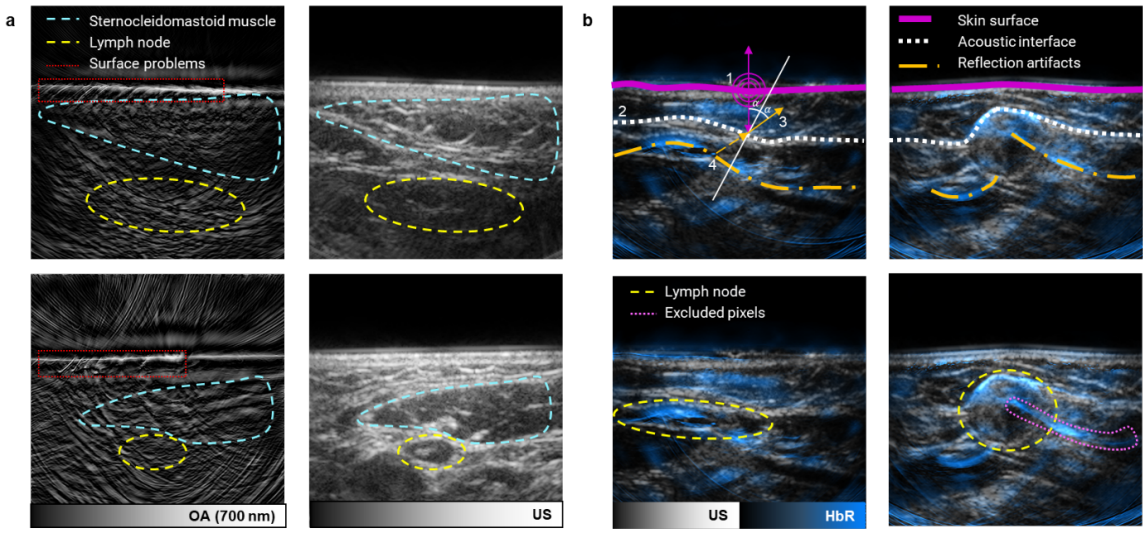

Figure 7.13: Image quality issues and reflection artifacts. a, Two examples of surface contact issues (red dotted line) and occlusion of the lymph nodes (yellow dashed line) by the sternocleidomastoid muscle (aqua dashed line). Both problems contribute to bad image quality and too weak signal in the lymph node region, as seen in the left optoacoustic images (λ=700 nm). Ultrasound images are shown on the right. **b,** Two examples of the reflection artifacts and a scheme of the physical mechanism behind the phenomenon. An acoustic pressure pulse is emitted from an optical absorber in all directions (1). Part of the wave travels directly to the detectors, but part propagates into the tissue. Acoustic interfaces inside the tissue (2) act as reflectors. The wave gets reflected from the interface and travels outwards to the transducers (3). The reconstruction algorithm is oblivious to the reflections affecting the incoming waves and assumes the signal came along a straight line from an absorber at a distance given by the travel time and the speed of sound (4). A reflection artifact is thus observed behind the reflecting interface at a distance proportional to the distance between the interface and the true pressure wave origin. The most observed reflection is of the skin where the light fluence is highest, and thus, the optoacoustic signal is strongest. The skin reflection mainly affects the images of deoxyhemoglobin. This happens because the main absorber in the skin is melanin, whose absorption spectrum coincides with deoxyhemoglobin in the MSOT wavelength range. **Abbreviations:** LN+, malignant lymph node; LN-, benign lymph node. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

issues: in one case the LN was obscured by skin reflection artifacts and four cases had surface contact issues causing strong artifacts while also having the view on the LN obscured by the sternocleidomastoid muscle. One malignant case (M1) was also partially affected by the skin reflection artifact but could still be used for the analysis after excluding the affected pixels. Fig. 7.13 demonstrates these problems in detail.

Analyzing the *in vivo* images, we observed that cetuximab-800CW could not be reliably detected in our setup. Fig. 7.11c shows the distributions of cetuximab-800CW unmixing coefficients in three malignant LNs pre- and post-injection. Comparing the tails of the cetuximab-800CW unmixing coefficient distributions, where a localized increase of coefficients would be apparent, we observed that the LN M3 exhibited an increase, LN M2 showed a decrease and LN M1 remained unchanged. We obtained the same result also when comparing the 95th percentiles or considering the means of the upper top 10% of the coefficient values, as reported in the earlier *ex vivo* study [371]. Fig. 7.11d shows visualizations of the cetuximab-800CW signal as a green overlay on the grayscale ultrasound images. The post-injection scan of the LN M3 clearly shows patches of increased signal, albeit a discrepancy exists between imaging angles pre- and post-injection, preventing us from conclusively proving the possibility to detect cetuximab-800CW *in vivo*.

On the other hand, using MSOT we could visualize intrinsic chromophores and related features in lymph nodes *in vivo*. Figure 7.12a shows the absorption spectra of the main tissue chromophores

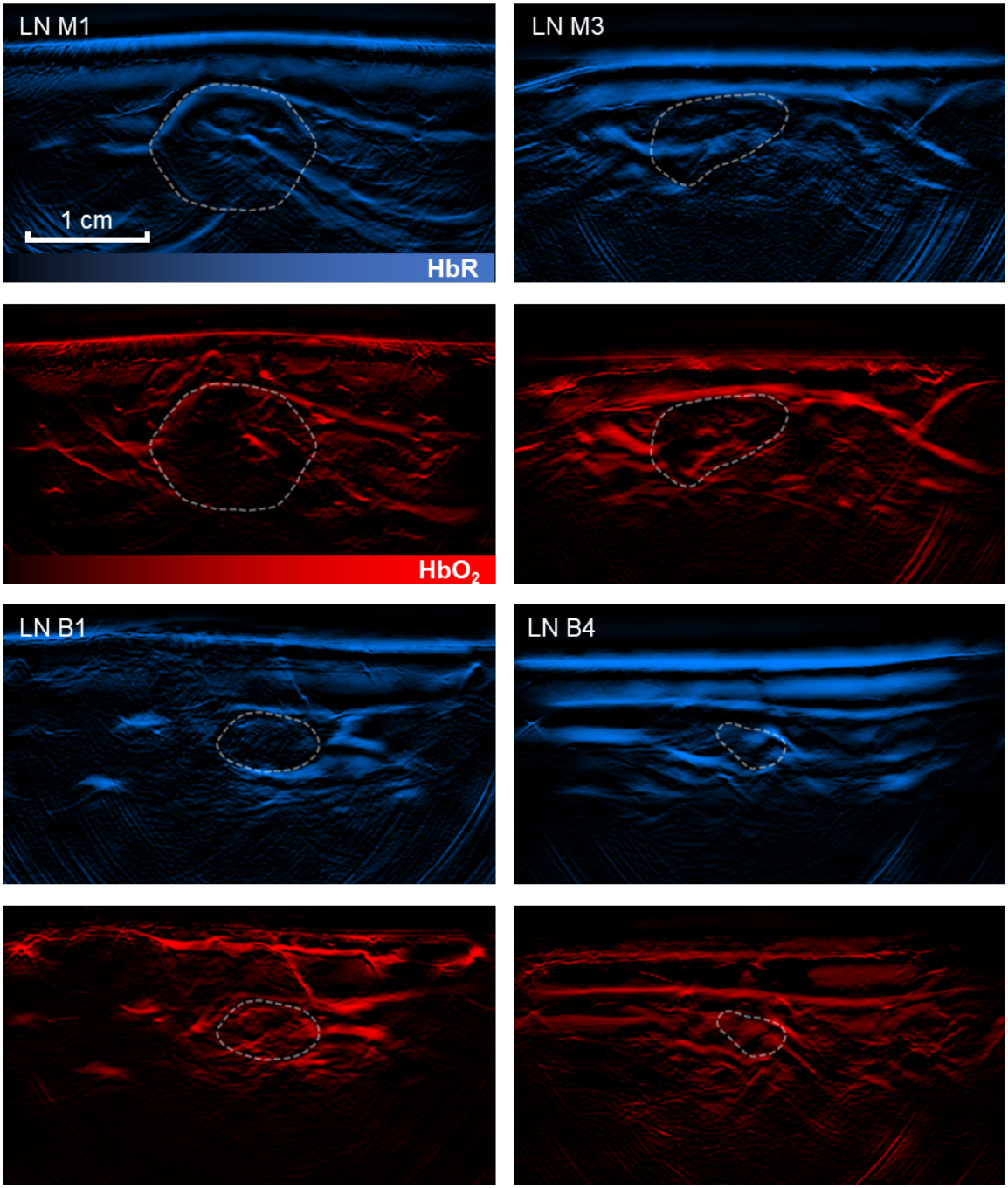

Figure 7.14: Intrinsic chromophore maps in malignant and benign lymph nodes. Standalone visualizations of deoxyhemoglobin (HbR) and oxyhemoglobin (HbO_2) maps in malignant lymph nodes M1 and M3 and benign lymph nodes B1 and B4. Figure reproduced from *Vonk, Kukačka et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study." Photoacoustics 26 (2022): 100362.*

in the wavelength range of MSOT. Dashed lines denote the wavelengths acquired during this study, allowing distinguishing Hb and HbO_2. Figure 7.12b shows images of malignant LNs (M1 and M3) and benign LNs (B1 and B4) with distinct features, such as vasculature (peripheral, feeding, and internodal; marked by arrowheads) and enhancement of Hb and HbO_2 signals in the LN borders (small arrows), although we could not establish apparent differences between malignant and benign LNs. Separate Hb and HbO_2 images are provided in Fig. 7.14. Figure 7.12c shows US images with

localization of adjacent anatomical structures to provide a better spatial context to the displayed MSOT images.

Quantitatively, malignant LNs exhibited a larger variance of Hb coefficients than the benign ones. The distributions of the unmixing coefficients of Hb and HbO_2 throughout the LN ROIs are shown in Fig. 7.12c. Figure 7.12d shows the variances of Hb and HbO_2 distributions of malignant and benign LNs relative to their margins. Malignant LNs exhibit significantly higher variance of Hb than benign ones ($p=0.047$, $n=9$). No significant difference in HbO_2-variance was observed between malignant and benign LNs ($p=0.349$, $n=9$).

***Ex vivo* analysis of surgical specimens**
To verify the tumor-specific contrast of cetuximab-800CW, the fluorescence images of all formalin-fixed LNs that were surgically excised during this study were analyzed ($n=149$). Significantly increased mean and maximum fluorescence intensities were observed in the malignant cases compared to the benign ones, confirming the increased concentration of cetuximab-800CW.

7.2.3 Discussion

This study demonstrated MSOT for *in vivo* imaging of LNs in oral cancer patients. We characterized the optoacoustic properties of cetuximab-800CW using a tissue-mimicking phantom. Furthermore, we provided the first clinical results of *in vivo* EGFR-targeted molecular imaging with MSOT. Although we validated the suitability of cetuximab-800CW as a tumor-specific contrast agent for MSOT through our phantom and *ex vivo* experiments, we identified a number of obstacles preventing its reliable detection *in vivo*. Next, we compiled a list of recommendations for future studies. As a secondary goal, we assessed the qualitative and quantitative optoacoustic features of LNs in oral cancer patients. MSOT was able to detect altered tissue metabolism in malignant LNs, showcased by an increase in deoxyhemoglobin variance.

Our phantom experiment demonstrated the ability of MSOT to detect the EGFR-targeted fluorescent probe, cetuximab-800CW, at concentrations above 400 µM. Furthermore, the analysis of the excised specimens verified the tumor-specific intake of the tracer in LNs, confirming observations from previous fluorescence molecular imaging studies [369, 370, 376]. Recently, Nishio *et al.* [371] reported that a IRDye800CW-labeled antibody could be used for *ex vivo* optoacoustic detection of LN metastases. We could not extend these results by observing the accumulation of cetuximab-800CW in MSOT scans of LNs *in vivo*. We observed an increase in cetuximab-800CW signal only in one out of three examined cases, but here also big discrepancy between the pre- and post-injection imaging angles was observed. On the contrary, the case with the best alignment between the pre- and post-injection images did not show any change. Multiple reasons could explain the differences between our results and the previously reported *ex vivo* study. First, the light scattering and absorption in the overlying tissue present during *in vivo* imaging cause spectral coloring and signal decay due to reduced light fluence, complicating the detection of optical contrast in the LNs compared to *ex vivo* imaging. Secondly, the maximum permissible light fluence exposure for in-human use [113] is well below the laser power used by Nishio *et al.*, further limiting the MSOT signal strength *in vivo*. Overall, we surmise that the selected dose of cetuximab-800CW lies below the detection limit of MSOT *in vivo* at depth, under the selected image acquisition parameters.

A secondary aim of this study was to explore MSOT of intrinsic contrast, particularly HbO_2 and HbR. The *in vivo* imaging of Hb and HbO_2 allowed the visualization of multiple endogenous features within and around LNs, such as lymphoid vasculature which seemed more extensively present in malignant LNs [377, 378]. Furthermore, we observed a significant increase in Hb variance in malignant LNs. The increased Hb heterogeneity in malignant LNs may be explained

by the fact that malignant LNs contain both healthy tissue, that typically exhibits lower oxygen metabolism, and tumor tissue that is characterized by regions of hypoxia due to decreased oxygen supply (from dysfunctional microvasculature) and increased oxygen demand (from the hypermetabolic state of tumor cells) [367, 379, 380]. The non-invasive detection of malignant LNs with MSOT can help to better stratify oral cancer patients for surgical treatment of the neck and avoid overtreatment resulting from the limited sensitivity of current imaging methods. Moreover, the use of intrinsic contrast disposes the need for exogenous contrast agents, allowing for seamless implementation into standard of care.

This first clinical proof-of-concept study revealed several limitations of the chosen methodology. First, the difficulties in linking *in vivo* imaging to the final histopathology aggravated the already quite low sample size, prohibiting significant results. Unlike in studies assessing the performance of established techniques (e.g., ultrasound), the novelty of MSOT LN imaging and the explorative nature of this study require meticulous correlation with histopathology before certain image characteristics can be attributed to the presence of tumor. Second, the mismatch of imaging positions between the pre- and post-injection scans complicates the comparison of chromophore quantities. Third, in this study we acquired images at six wavelengths matching the absorption spectrum characteristics of IRDye800CW. There are multiple other chromophores in the tissue and the spectral coloring increases the range of observable spectra even further, in which case six wavelengths do not facilitate reliable spectral unmixing. Finally, besides the challenges in the study procedure itself, the issue of detecting the contrast agent *in vivo* remains. We studied IRDye800CW as we could determine its potential noninvasively by including patients already administered with cetuximab-800CW as part of a fluorescence-guided surgery trial. However, better contrast agents could be considered to improve the tumor-specific signal generation for MSOT, such as gold nanoparticles [381]. The main criteria for optoacoustic contrast agents include strong and sharply peaked optical absorption in the near-infrared window, high optoacoustic efficiency, and optimal biocompatibility [37, 382]. Also, when evaluating novel optoacoustic contrast agents, it is helpful to realize that the accumulation of contrast agents involves complex pharmacokinetic processes, including both active binding of the target of interest (e.g., receptor) as well as nonspecific accumulation due to variety in lymph and blood physiological processes [383, 384]. To better quantify target expression one could use a paired-imaging approach, where a non-targeted tracer with a different absorption spectrum is administered simultaneously and used to correct for nonspecific accumulation [385, 386].

Based on the limitations that we identified during our study, we compiled the following recommendations for future studies: 1) ensure node-by-node comparison by establishing specific study designs [387] or including only patients of which the LNs identified with preoperative imaging can tracked until final histopathology (e.g. preoperative lymphoscintigraphy and intraoperative Geiger meter-detection in the sentinel node procedure) [388, 389]; 2) when multiple sentinel LNs are identified, the endpoints should comprise the number of LNs identified, the metastatic state of the sentinel LN specimen and the number of malignant LNs; 3) a baseline MSOT could be performed to study pre- and post-injection scans; if a good match between pre- and post-injection scans cannot be established, multiple angles should be scanned and their results averaged; 4) additional wavelengths should be included in the image acquisition to increase the accuracy of spectral unmixing (i.e., at 25 wavelengths, the image acquisition takes 1s and motion is thus not a problem, but the redundancy in the spectral dimension makes spectral unmixing more robust to noise and light fluence effects); 5) blind spectral unmixing (e.g., non-negative matrix factorization) may be preferred to simple linear spectral unmixing when recording sufficient number wavelengths, as it can better adapt to observed spectral variations; 6) MSOT and, if the quantum yield of the of the contrast agent allows, fluorescence imaging of the excised and grossed LNs can be performed to validate *in vivo* imaging results. Here, MSOT of the excised LNs would require scanning in a water bath to improve acoustic coupling; 7) all sentinel lymph nodes should be

examined through histopathology to determine their metastatic status. Finally, when suspicious features are determined based on MSOT, subsequent studies can correlate the number of identified LNs with final histopathology as in studies with ultrasound.

In conclusion, we identified possible reasons which prevent achieving the primary goal of detecting cetuximab-800 CW *in vivo*. Furthermore, we demonstrated that in vivo MSOT can observe clinically important features in LNs, specifically using hemoglobin parameters. In particular, increased variance of the deoxyhemoglobin distribution inside the LNs could serve as an optoacoustic marker of LN metastases. Additionally, we suggested several approaches to improve the efficiency of follow-up MSOT studies on LN imaging, such as a confined study population that allows correlation with histopathology, using more wavelengths during the image acquisition and the use of advanced spectral unmixing algorithms. Following our suggestions, future studies may better evaluate the clinical benefit of optoacoustic contrast agents as well as the intrinsic LN features observed with MSOT, facilitating improved preoperative detection of LN metastases.

CHAPTER 8

Conclusion

With the aim to advance clinical optoacoustic tomography, this work has introduced improvements across the whole lifecycle of an optoacoustic image: from the recorded electrical signals to the clinical interpretation of the image contents. Altogether, this work presents the second generation of optoacoustic tomography methods and defines an improved standard for future clinical applications.

First, we have proposed a formal framework called Motion score for quantification of motion in hybrid optoacoustic-ultrasound recordings. Motion is an inevitable element of handheld imaging, corrupting the spectral dimension of optoacoustic images. Selecting frames which are the least affected by motion is the first step towards achieving an optimal image quality. We have validated that Motion score can select stationary frames with a human-like accuracy and can be used for a reproducible, automated frame selection. Furthermore, we have shown that the frames selected by Motion score have indeed better spectral quality than other frames, ensuring higher precision of image analysis.

Second, we have presented the current state-of-the-art optoacoustic image reconstruction procedure and proposed additional post-processing steps. The combination of these methods leads to a new level of image quality produced by a handheld optoacoustic tomography system. Images are reconstructed using a mathematical model of the imaging system which accounts for heterogeneities in the speed-of-sound and the related refraction on the probe membrane, as well as for the spatial and electrical impulse response of the ultrasound detectors. Explicit modeling of these effects ensures spatial fidelity of the reconstructed images as well as an improved spatial resolution of 200 μm. The subsequent post-processing is performed by alignment and aggregation of the optoacoustic images across the time and the wavelength dimensions. This step contributes towards artifact reduction as well as improving the contrast-to-noise ratio. Together, this image processing pipeline allows resolving minute features deep in the tissue to a level never before attained in handheld optoacoustic tomography.

Third, we have addressed the topic of visualization of optoacoustic images, which is essential for optimal qualitative evaluation of the features present in the images. We have proposed solutions to two issues inherent to optoacoustic tomography visualization. First, we introduced the dual-band visualization that equalizes the clarity of visual features on different scales. This equalization is achieved by combining a full-band image with an image obtained from high-pass filtered signals. This way, the small features (corresponding to the high-frequency portion of the signal) are emphasized. A special colormap used for rendering of the two images avoids noise overamplification. Second, we have introduced a pair of non-linear adaptive filters to adjust an extremely skewed pixel-value distribution caused by the light fluence decay in depth. Implemented within a user-friendly GUI, local contrast normalization and sigmoid normalization allow optimal contrast enhancement for viewing the optoacoustic image features.

Next, we have considered the possibility of extracting morphological information from optoacoustic images by means of automated image segmentation. Achieving an accurate segmentation of the tissue in optoacoustic tomography images would be an important milestone towards quantitative optoacoustics. First, quantitative biomarkers could be extracted directly from the morphological information provided by the segmentation, such as vascular network metrics. Secondly, a tissue model based on the segmentation could be incorporated into an optical inversion scheme, enabling more precise chromophore concentration estimates. Although deep learning methods offer high accuracy, their use in optoacoustic tomography is limited by the availability of annotated data. To bypass this issue, we have explored the self-supervised learning paradigm for training models on unlabeled data. Our experiments with fundus images have validated advantages of the approach which could be transferred to the optoacoustic domain.

Finally, we have applied the advanced tools presented herein to two oncological clinical studies. Our study on breast cancer patients has identified various features of breast tumors that can be resolved thanks to the improved image quality. To interpret the meaning of our observations, we have correlated them to histology images and clinical information on eight representative cases. This study serves as a guide for future studies on breast imaging in terms of recommended data processing as well as target features to validate on a larger cohort. Our second study was concerned with detection of lymph node metastases in oral cancer patients using a tumor-specific optical contrast agent. Despite positive results reported by earlier *ex vivo* studies, we could not detect the tracer in lymph nodes *in vivo*. We have outlined possible reasons and suggested improvements for future studies. Furthermore, we have proposed an alternative marker of malignancy based on the deoxyhemoglobin variance within the lymph node.

8.1 Outlook

Future clinical translation of optoacoustic tomography would benefit from advances in the following directions:

Image processing
The image processing pipeline proposed herein is limited by the offline computation requirement, preventing its utilization for real-time guidance of the operator during image acquisition. Advances in model-based image formation accelerated by deep learning have the potential to overcome this limitation and will surely play an important role in the future of clinical optoacoustic tomography.

Furthermore, the proposed image processing pipeline would benefit from a stricter standardization and an open-source implementation that could be utilized by other groups working with *Acuity Echo* MSOT scanners. Although some parts of the pipeline have already been made available, using them still requires considerable expertise.

Optoacoustic image processing would also benefit from the development of neural representations of the images. Our experiments have hinted at the possibility of learning image feature detectors solely from unlabeled data. Such models could be utilized for regularization in the image reconstruction step to promote certain patterns or suppress artifacts. Furthermore, they could be finetuned to perform tissue segmentation, which would in turn serve as a prior for an improved image reconstruction.

Image visualization
The presented image visualization GUI produces acceptable images under the default parameter settings, yet manual parameter tweaking is required for achieving the optimal contrast in each

image. This procedure may be time-consuming and not reproducible. The introduction of suitable image quality metrics would enable automatization of this process.

Image analysis
Blind unmixing methods are a promising direction for analysis of multispectral optoacoustic images. Although the development of accurate optical inversion methods would eliminate the need for blind unmixing, it is a useful tool for clinical image analysis in the meantime.

Furthermore, the development of tissue segmentation models would enable automated extraction of quantitative morphological features. Whereas the curation of large, standardized datasets and open benchmarks would be needed, utilization of unsupervised neural representations of optoacoustic images could alleviate the lack of annotated data.

Clinical validation
Finally, the clinical translation of optoacoustic tomography requires clinical validation of the features and biomarkers that can be extracted from the images. Clinical studies should be planned carefully with the involvement of experts across the whole image lifecycle—radiologists, oncologists, engineers, and data scientists—to ensure all data are collected in the form that is required and the study goals can be achieved. Additionally, the interdisciplinary exchange should also take place during the clinical study to provide feedback between the involved parties, facilitate learning, and identify problems quickly. Furthermore, larger studies would require multi-center collaborations and standardization of the optoacoustic procedures. Experiments on the reproducibility of the measurements should be performed to identify potential issues related to larger studies.

◆ ◆ ◆

Bibliography

1. Erdmann, F., et al., *Krebs in Deutschland für 2017/2018.* 2021.
2. Statistisches Bundesamt (Destatis), *Todesursachenstatistik 2020: Zahl der Todesfälle um 4,9 % gestiegen*. 2021: Wiesbaden.
3. Kooperationsgemeinschaft Mammographie, *Jahresbericht Evaluation 2018. Deutsches Mammographie-Screening-Programm*. 2020: Berlin.
4. Maxwell, J.R., et al., *Imaging-guided core needle biopsy of the breast: study of psychological outcomes.* The breast journal, 2000. **6**(1): p. 53-61.
5. Marmot, M.G., et al., *The benefits and harms of breast cancer screening: an independent review.* British journal of cancer, 2013. **108**(11): p. 2205-2240.
6. Narod, S.A., et al., *Breast cancer mortality after a diagnosis of ductal carcinoma in situ.* JAMA oncology, 2015. **1**(7): p. 888-896.
7. Sikov, W.M., et al., *General principles of neoadjuvant management of breast cancer*. 2020, UpToDate.
8. Koga, K.H., et al., *Monitoring the response to neoadjuvant chemotherapy in breast cancer*, in *Neoadjuvant Chemotherapy—Increasing Relevance in Cancer Management*, M.M. Markman, Editor. 2013, IntechOpen. p. 25-48.
9. Romeo, V., et al., *Assessment and Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer: A Comparison of Imaging Modalities and Future Perspectives.* Cancers (Basel), 2021. **13**(14).
10. Le, M.T., et al., *Is the false-positive rate in mammography in North America too high?* The British journal of radiology, 2016. **89**(1065): p. 20160045.
11. Mario, J., et al., *Benign breast lesions that mimic cancer: Determining radiologic-pathologic concordance.* Appl Radiol, 2015. **44**(9): p. 28-32.
12. Kerlikowske, K., et al., *Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy.* JAMA internal medicine, 2013. **173**(9): p. 807-816.
13. Lehman, C.D., et al., *Accuracy and value of breast ultrasound for primary imaging evaluation of symptomatic women 30-39 years of age.* American journal of roentgenology, 2012. **199**(5): p. 1169-1177.
14. Lee, J.M., et al., *Performance of screening ultrasonography as an adjunct to screening mammography in women across the spectrum of breast cancer risk.* JAMA internal medicine, 2019. **179**(5): p. 658-667.
15. Stavros, A.T., et al., *Solid breast nodules: use of sonography to distinguish between benign and malignant lesions.* Radiology, 1995. **196**(1): p. 123-134.
16. Rahbar, G., et al., *Benign versus malignant solid breast masses: US differentiation.* Radiology, 1999. **213**(3): p. 889-894.
17. Calas, M., et al., *Intraobserver interpretation of breast ultrasonography following the BI-RADS classification.* European journal of radiology, 2010. **74**(3): p. 525-528.
18. Geisel, J., M. Raghu, and R. Hooley. *The role of ultrasound in breast cancer screening: the case for and against ultrasound*. in *Seminars in Ultrasound, CT and MRI*. 2018. Elsevier.
19. Lin, X., et al., *The diagnostic performance of automated versus handheld breast ultrasound and mammography in symptomatic outpatient women: a multicenter, cross-sectional study in China.* European Radiology, 2021. **31**(2): p. 947-957.
20. Moss, H.A., et al., *How reliable is modern breast imaging in differentiating benign from malignant breast lesions in the symptomatic population?* Clinical radiology, 1999. **54**(10): p. 676-682.
21. Gonzaga, M.A., *How accurate is ultrasound in evaluating palpable breast masses?* Pan African Medical Journal, 2010. **7**(1).

22. Zanello, P.A., et al., *Breast ultrasound diagnostic performance and outcomes for mass lesions using Breast Imaging Reporting and Data System category 0 mammogram.* Clinics, 2011. **66**: p. 443-448.
23. Newell, M., et al., *ACR practice parameter for the performance of contrast enhanced magnetic resonance imaging (MRI) of the breast.* American College of Radiology, Reston, VA, 2018.
24. Knopp, M., et al., *Pathophysiologic basis of contrast enhancement in breast tumors.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 1999. **10**(3): p. 260-266.
25. Thompson, J.L. and G.P. Wright, *The role of breast MRI in newly diagnosed breast cancer: an evidence-based review.* The American Journal of Surgery, 2021. **221**(3): p. 525-528.
26. Kumar, A., et al., *Doppler ultrasound scoring to predict chemotherapeutic response in advanced breast cancer.* World Journal of Surgical Oncology, 2007. **5**(1): p. 1-6.
27. Tromberg, B.J., et al., *Predicting responses to neoadjuvant chemotherapy in breast cancer: ACRIN 6691 trial of diffuse optical spectroscopic imaging.* Cancer research, 2016. **76**(20): p. 5933-5944.
28. Manohar, S. and D. Razansky, *Photoacoustics: a historical review.* Advances in optics and photonics, 2016. **8**(4): p. 586-617.
29. Nagy, J.A., et al. *Heterogeneity of the tumor vasculature*. in *Seminars in thrombosis and hemostasis*. 2010. © Thieme Medical Publishers.
30. Hanahan, D. and J. Folkman, *Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis.* cell, 1996. **86**(3): p. 353-364.
31. Oraevsky, A.A., et al. *Laser optoacoustic imaging of the breast: detection of cancer angiogenesis*. in *Optical Tomography and Spectroscopy of Tissue III*. 1999. International Society for Optics and Photonics.
32. Tromberg, B.J., et al., *Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy.* Neoplasia, 2000. **2**(1-2): p. 26-40.
33. Walsh, J.C., et al., *The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities.* Antioxidants & redox signaling, 2014. **21**(10): p. 1516-1554.
34. Zhu, Q., et al., *Assessment of functional differences in malignant and benign breast lesions and improvement of diagnostic accuracy by using US-guided diffuse optical tomography in conjunction with conventional US.* Radiology, 2016. **280**(2): p. 387-397.
35. Ntziachristos, V. and B. Chance, *Breast imaging technology: Probing physiology and molecular function using optical imaging-applications to breast cancer.* Breast Cancer Research, 2000. **3**(1): p. 41.
36. Liapis, E., et al., *Resolution of Spatial and Temporal Heterogeneity in Bevacizumab-Treated Breast Tumors by Eigenspectra Multispectral Optoacoustic TomographyeMSOT-Resolved Oxygenation in Bevacizumab-Treated Tumors.* Cancer Research, 2020. **80**(23): p. 5291-5304.
37. Weber, J., P.C. Beard, and S.E. Bohndiek, *Contrast agents for molecular photoacoustic imaging.* Nature methods, 2016. **13**(8): p. 639-650.
38. Hudson, S.V., et al., *Targeted noninvasive imaging of EGFR-expressing orthotopic pancreatic cancer using multispectral optoacoustic tomography.* Cancer research, 2014. **74**(21): p. 6271-6279.
39. Mallidi, S., et al., *Molecular specific optoacoustic imaging with plasmonic nanoparticles.* Optics Express, 2007. **15**(11): p. 6583-6588.
40. Weissleder, R., et al., *In vivo imaging of tumors with protease-activated near-infrared fluorescent probes.* Nature biotechnology, 1999. **17**(4): p. 375-378.
41. Huang, X., S. Lee, and X. Chen, *Design of "smart" probes for optical imaging of apoptosis.* American journal of nuclear medicine and molecular imaging, 2011. **1**(1): p. 3.
42. Yang, Q., et al., *In vivo photoacoustic imaging of chemotherapy-induced apoptosis in squamous cell carcinoma using a near-infrared caspase-9 probe.* Journal of Biomedical Optics, 2011. **16**(11): p. 116026.

43. Oraevsky, A.A., et al. *Laser-based optoacoustic imaging in biological tissues*. in *Laser-Tissue Interaction V; and Ultraviolet Radiation Hazards*. 1994. International Society for Optics and Photonics.
44. Kruger, R.A. and P. Liu, *Photoacoustic ultrasound: Pulse production and detection in 0.5% Liposyn.* Medical physics, 1994. **21**(7): p. 1179-1184.
45. Manohar, S., et al., *The Twente Photoacoustic Mammoscope: system overview and performance.* Physics in Medicine & Biology, 2005. **50**(11): p. 2543.
46. Diot, G., et al., *Multispectral optoacoustic tomography (MSOT) of human breast cancer.* Clinical Cancer Research, 2017. **23**(22): p. 6912-6922.
47. Toi, M., et al., *Visualization of tumor-related blood vessels in human breast by photoacoustic imaging system with a hemispherical detector array.* Scientific reports, 2017. **7**: p. 41970.
48. Lin, L., et al., *Single-breath-hold photoacoustic computed tomography of the breast.* Nature communications, 2018. **9**(1): p. 1-9.
49. Becker, A., et al., *Multispectral optoacoustic tomography of the human breast: characterisation of healthy tissue and malignant lesions using a hybrid ultrasound-optoacoustic approach.* European radiology, 2018. **28**(2): p. 602-609.
50. Manohar, S., et al., *Initial results of in vivo non-invasive cancer imaging in the human breast using near-infrared photoacoustics.* Opt Express, 2007. **15**(19): p. 12277-85.
51. Heijblom, M., et al., *The state of the art in breast imaging using the Twente Photoacoustic Mammoscope: results from 31 measurements on malignancies.* Eur Radiol, 2016. **26**(11): p. 3874-3887.
52. Heijblom, M., et al., *Visualizing breast cancer using the Twente photoacoustic mammoscope: what do we learn from twelve new patient measurements?* Opt Express, 2012. **20**(11): p. 11582-97.
53. Heijblom, M., W. Steenbergen, and S. Manohar, *Clinical photoacoustic breast imaging: the Twente experience.* IEEE Pulse, 2015. **6**(3): p. 42-6.
54. Neuschler, E.I., et al., *A pivotal study of optoacoustic imaging to diagnose benign and malignant breast masses: a new evaluation tool for radiologists.* Radiology, 2018. **287**(2): p. 398-412.
55. Neuschler, E.I., et al., *Downgrading and upgrading gray-scale ultrasound bi-rads categories of benign and malignant masses with optoacoustics: a pilot study.* American Journal of Roentgenology, 2018. **211**(3): p. 689-700.
56. Menezes, G.L., et al., *Downgrading of breast masses suspicious for cancer by using optoacoustic breast imaging.* Radiology, 2018. **288**(2): p. 355-365.
57. Kruger, R.A., et al., *Dedicated 3D photoacoustic breast imaging.* Med Phys, 2013. **40**(11): p. 113301.
58. Fakhrejahani, E., et al., *Clinical Report on the First Prototype of a Photoacoustic Tomography System with Dual Illumination for Breast Cancer Imaging.* PLoS One, 2015. **10**(10): p. e0139113.
59. Asao, Y., et al., *Photoacoustic mammography capable of simultaneously acquiring photoacoustic and ultrasound images.* Journal of biomedical optics, 2016. **21**(11): p. 116009.
60. Matsumoto, Y., et al., *Visualising peripheral arterioles and venules through high-resolution and large-area photoacoustic imaging.* Scientific reports, 2018. **8**(1): p. 1-11.
61. Matsumoto, Y., et al., *Label-free photoacoustic imaging of human palmar vessels: a structural morphological analysis.* Scientific reports, 2018. **8**(1): p. 1-8.
62. Nagae, K., et al., *Real-time 3D photoacoustic visualization system with a wide field of view for imaging human limbs.* F1000Research, 2018. **7**.
63. Karlas, A., et al., *Multispectral optoacoustic tomography of lipid and hemoglobin contrast in human carotid atherosclerosis.* Photoacoustics, 2021. **23**: p. 100283.
64. Masthoff, M., et al., *Use of multispectral optoacoustic tomography to diagnose vascular malformations.* JAMA dermatology, 2018. **154**(12): p. 1457-1462.

65. Karlas, A., et al., *Multispectral optoacoustic tomography of muscle perfusion and oxygenation under arterial and venous occlusion: A human pilot study.* Journal of Biophotonics, 2020: p. e201960169.
66. Dima, A. and V. Ntziachristos, *Non-invasive carotid imaging using optoacoustic tomography.* Optics express, 2012. **20**(22): p. 25044-25057.
67. van den Berg, P.J., et al., *Feasibility of photoacoustic/ultrasound imaging of synovitis in finger joints using a point-of-care system.* Photoacoustics, 2017. **8**: p. 8-14.
68. Jo, J., et al., *Photoacoustic tomography for human musculoskeletal imaging and inflammatory arthritis detection.* Photoacoustics, 2018. **12**: p. 82-89.
69. Reber, J., et al., *Non-invasive measurement of brown fat metabolism based on optoacoustic imaging of hemoglobin gradients.* Cell metabolism, 2018. **27**(3): p. 689-701. e4.
70. Knieling, F., et al., *Multispectral Optoacoustic Tomography for Assessment of Crohn's Disease Activity.* The New England journal of medicine, 2017. **376**(13): p. 1292.
71. Jose, J., et al., *Initial results of imaging melanoma metastasis in resected human lymph nodes using photoacoustic computed tomography.* Journal of biomedical optics, 2011. **16**(9): p. 096021-096021-5.
72. Grootendorst, D., et al., *First experiences of photoacoustic imaging for detection of melanoma metastases in resected human lymph nodes.* Lasers in surgery and medicine, 2012. **44**(7): p. 541-549.
73. Garcia-Uribe, A., et al., *Dual-modality photoacoustic and ultrasound imaging system for noninvasive sentinel lymph node detection in patients with breast cancer.* Scientific reports, 2015. **5**(1): p. 1-8.
74. Stoffels, I., et al., *Metastatic status of sentinel lymph nodes in melanoma determined noninvasively with multispectral optoacoustic imaging.* Science translational medicine, 2015. **7**(317): p. 317ra199-317ra199.
75. Stoffels, I., et al., *Assessment of nonradioactive multispectral optoacoustic tomographic imaging with conventional lymphoscintigraphic imaging for sentinel lymph node biopsy in melanoma.* JAMA network open, 2019. **2**(8): p. e199020-e199020.
76. Kang, J., et al., *Ex vivo estimation of photoacoustic imaging for detecting thyroid microcalcifications.* PLoS One, 2014. **9**(11): p. e113358.
77. Dogra, V.S., et al., *Preliminary results of ex vivo multispectral photoacoustic imaging in the management of thyroid cancer.* American Journal of Roentgenology, 2014. **202**(6): p. W552-W558.
78. Yang, M., et al., *Photoacoustic/ultrasound dual imaging of human thyroid cancers: an initial clinical study.* Biomedical optics express, 2017. **8**(7): p. 3449-3457.
79. Levi, J., et al., *Design, synthesis, and imaging of an activatable photoacoustic probe.* Journal of the American Chemical Society, 2010. **132**(32): p. 11264-11269.
80. Dima, A. and V. Ntziachristos, *In-vivo handheld optoacoustic tomography of the human thyroid.* Photoacoustics, 2016. **4**(2): p. 65-69.
81. Kroenke, M., et al., *Multispectral optoacoustic tomography: a novel label-free imaging technique for the assessment of hyperthyroid diseases*. 2019, Soc Nuclear Med.
82. Roll, W., et al., *Multispectral Optoacoustic Tomography of Benign and Malignant Thyroid Disorders: A Pilot Study.* Journal of Nuclear Medicine, 2019. **60**(10): p. 1461-1466.
83. Kim, J., et al., *Multiparametric Photoacoustic Analysis of Human Thyroid Cancers in Vivophotoacoustic Analysis of Human Thyroid Cancers.* Cancer Research, 2021. **81**(18): p. 4849-4860.
84. Kothapalli, S.-R., et al., *Simultaneous transrectal ultrasound and photoacoustic human prostate imaging.* Science translational medicine, 2019. **11**(507): p. eaav2169.
85. Nedosekin, D.A., et al., *In vivo ultra-fast photoacoustic flow cytometry of circulating human melanoma cells using near-infrared high-pulse rate lasers.* Cytometry Part A, 2011. **79**(10): p. 825-833.
86. Zhang, E.Z., et al., *Multimodal photoacoustic and optical coherence tomography scanner using an all optical detection scheme for 3D morphological skin imaging.* Biomedical optics express, 2011. **2**(8): p. 2202-2215.

87. Attia, A.B.E., et al., *Noninvasive real-time characterization of non-melanoma skin cancers with handheld optoacoustic probes.* Photoacoustics, 2017. **7**: p. 20-26.
88. Breathnach, A., et al., *Preoperative measurement of cutaneous melanoma and nevi thickness with photoacoustic imaging.* Journal of Medical Imaging, 2018. **5**(1): p. 015004.
89. Hai, P., et al., *Label-free high-throughput photoacoustic tomography of suspected circulating melanoma tumor cells in patients in vivo.* Journal of biomedical optics, 2020. **25**(3): p. 036002.
90. Liu, S., et al., *In vivo photoacoustic sentinel lymph node imaging using clinically-approved carbon nanoparticles.* IEEE Transactions on Biomedical Engineering, 2019. **67**(7): p. 2033-2042.
91. Kim, C., et al., *Handheld array-based photoacoustic probe for guiding needle biopsy of sentinel lymph nodes.* Journal of biomedical optics, 2010. **15**(4): p. 046010.
92. Kang, J., et al., *Real-time sentinel lymph node biopsy guidance using combined ultrasound, photoacoustic, fluorescence imaging: in vivo proof-of-principle and validation with nodal obstruction.* Scientific reports, 2017. **7**(1): p. 1-9.
93. Tummers, W.S., et al., *Intraoperative pancreatic cancer detection using tumor-specific multimodality molecular imaging.* Annals of surgical oncology, 2018. **25**(7): p. 1880-1888.
94. Horiguchi, A., et al., *A pilot study of photoacoustic imaging system for improved real-time visualization of neurovascular bundle during radical prostatectomy.* The Prostate, 2016. **76**(3): p. 307-315.
95. U.S. Food and Drug Administration, *Premarket Approval: Imagio Breast Imaging System*. 2021, FDA.
96. Pleitez, M.A., et al., *Label-free metabolic imaging by mid-infrared optoacoustic microscopy in living cells.* Nature biotechnology, 2020. **38**(3): p. 293-296.
97. Seeger, M., et al., *Pushing the boundaries of optoacoustic microscopy by total impulse response characterization.* Nature communications, 2020. **11**(1): p. 1-13.
98. Omar, M., J. Aguirre, and V. Ntziachristos, *Optoacoustic mesoscopy for biomedicine.* Nature biomedical engineering, 2019. **3**(5): p. 354-370.
99. Omar, M., et al., *Pushing the optical imaging limits of cancer with multi-frequency-band raster-scan optoacoustic mesoscopy (RSOM).* Neoplasia, 2015. **17**(2): p. 208-214.
100. Guo, H., et al., *Photoacoustic endoscopy: A progress review.* Journal of Biophotonics, 2020. **13**(12): p. e202000217.
101. Joseph, J., et al., *Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects.* Journal of Nuclear Medicine, 2017. **58**(5): p. 807-814.
102. Xia, J. and L.V. Wang, *Small-animal whole-body photoacoustic tomography: a review.* IEEE Transactions on Biomedical Engineering, 2013. **61**(5): p. 1380-1389.
103. Razansky, D., A. Buehler, and V. Ntziachristos, *Volumetric real-time multispectral optoacoustic tomography of biomarkers.* Nature Protocols, 2011. **6**(8): p. 1121-1129.
104. Manohar, S. and M. Dantuma, *Current and future trends in photoacoustic breast imaging.* Photoacoustics, 2019.
105. Zhang, G., et al., *Developing a photoacoustic whole-breast imaging system based on the synthetic matrix array.* Frontiers in Physics, 2020. **8**: p. 600589.
106. Niederhauser, J.J., et al., *Combined ultrasound and optoacoustic system for real-time high-contrast vascular imaging in vivo.* IEEE transactions on medical imaging, 2005. **24**(4): p. 436-440.
107. Buehler, A., et al., *Real-time handheld multispectral optoacoustic imaging.* Optics letters, 2013. **38**(9): p. 1404-1406.
108. Oraevsky, A., et al., *Clinical optoacoustic imaging combined with ultrasound for coregistered functional and anatomical mapping of breast tumors.* Photoacoustics, 2018. **12**: p. 30-45.
109. Diebold, G., T. Sun, and M. Khan, *Photoacoustic monopole radiation in one, two, and three dimensions.* Physical review letters, 1991. **67**(24): p. 3384.
110. Hysi, E., E. Strohm, and M. Kolios, *Probing different biological length scales using photoacoustics: from 1 to 1000 MHz.* Handbook of Photonics for Biomedical Engineering, 2014. **10**: p. 978-94.

111. Ntziachristos, V., *Going deeper than microscopy: the optical imaging frontier in biology.* Nature methods, 2010. **7**(8): p. 603.
112. Shnaiderman, R., et al., *A submicrometre silicon-on-insulator resonator for ultrasound detection.* Nature, 2020. **585**(7825): p. 372-378.
113. Institute, A.N.S., *American National Standard for Safe Use of Lasers*. 2000, Laser Institute of America New York.
114. Wang, L.V., *Tutorial on photoacoustic microscopy and computed tomography.* IEEE Journal of Selected Topics in Quantum Electronics, 2008. **14**(1): p. 171-179.
115. Croft, W.J., *Under the microscope: a brief history of microscopy*. Vol. 5. 2006: World Scientific.
116. Henderson, F., *Door-mats and penumbras: Hooke's contributions to the English Language*, in *Robert Hooke's London*. 2014.
117. Davis, I.M., *"Round, red globules floating in a crystalline fluid"–Antoni van Leeuwenhoek's observations of red blood cells and hemocytes.* Micron, 2022. **157**: p. 103249.
118. Brunton, J., *A new otoscope or speculum auris.* The Lancet, 1865. **86**(2205): p. 617-618.
119. Hofmann, F., *Beitrag zur Untersuchung des äusseren Gehörganges.* Casper's Wochenschrift für die gesammte Heilkunde, Berlin, 1841: p. 10-14.
120. Hartmann, A., *The diseases of the ear and their treatment*. 1887, New York: Putnam.
121. von Helmholtz, H., *Beschreibung eines Augen-Spiegels zur Untersuchung der Netzhaut im lebenden Auge*. 1851, Berlin: A. Förstner'sche Verlagsbuchhandlung.
122. Röntgen, W.C., *Über eine neue Art von Strahlen*. Aus den Sitzungsberichten der Würzburger Physik.-medic. Gesellschaft. 1895, Würzburg: Stahel'sche K. Hof- und Universitätsbuch- und Kunsthandlung.
123. Frankel, R.I., *Centennial of Röntgen's discovery of x-rays.* Western journal of medicine, 1996. **164**(6): p. 497-501.
124. Bell, A.G., *Upon the production and reproduction of sound by light.* Journal of the Society of Telegraph Engineers, 1880. **9**(34): p. 404-426.
125. Wang, L.V. and H.-i. Wu, *Biomedical optics: principles and imaging*. 2012: John Wiley & Sons.
126. Preece, W.H., *On the conversion of radiant energy into sonorous vibrations.* Proceedings of the Royal Society of London, 1881. **31**(206-211): p. 506-520.
127. Rosencwaig, A. and A. Gersho, *Photoacoustic effect with solids: a theoretical treatment.* Science, 1975. **190**(4214): p. 556-557.
128. Veingerov, M., *New method of gas analysis based on tyndall-roentgen optoacoustic effect.* Doklady Akademii Nauk SSSR, 1938. **19**(687): p. 8.
129. Pfund, A., *Atmospheric contamination.* Science, 1939. **90**(2336): p. 326-327.
130. Luft, K., *Über eine neue Methode der registrierenden Gasanalyse mit Hilfe der Absorption ultraroter Strahlen ohne spektrale Zerlegung.* Z. tech. Phys, 1943. **24**: p. 97-104.
131. Veingerov, M., *An optical-acoustic method of gas analysis.* Nature, 1946. **158**: p. 28.
132. Bell, A.G., *The Spectrophone.* Bull. Philos. Soc., 1881. **4**: p. 42.
133. Curie, J. and P. Curie, *Développement par compression de l'électricité polaire dans les cristaux hémièdres à faces inclinées.* Bulletin de minéralogie, 1880. **3**(4): p. 90-93.
134. Amar, L., *Detection d'omes elastiques (ultrasonores) sur l'os occipital, induites par impulsions laser dans l'oeil d'un lapin.* CR Acad. Sc. Paris, 1964. **259**: p. 3653-3655.
135. Amar, L., et al., *On the detection of laser induced ultrasonic waves in the human eye and the elaboration of a theory on the fundamental mechanism of vision.* Zeitschrift für angewandte Mathematik und Physik ZAMP, 1965. **16**: p. 182-183.
136. Cleary, S.F. and P.E. Hamrick, *Laser-induced acoustic transients in the mammalian eye.* The Journal of the Acoustical Society of America, 1969. **46**(4B): p. 1037-1044.
137. Rosencwaig, A., *Photoacoustic spectroscopy of biological materials.* Science, 1973. **181**(4100): p. 657-658.
138. Campbell, S.D., S.S. Yee, and M.A. Afromowitz, *Applications of photoacoustic spectroscopy to problems in dermatology research.* IEEE Transactions on Biomedical Engineering, 1979(4): p. 220-227.

139. Bowen, T. *Radiation-Induced Thermoacoustic Soft Tissue Imaging*. in *Ultrasonics Symposium* 1981. IEEE.
140. Chen, Q., et al., *Photo-acoustic probe for intra-arterial imaging and therapy.* Electronics Letters, 1993. **18**(29): p. 1632-1633.
141. Khamapirad, T., et al. *Diagnostic imaging of breast cancer with LOIS: clinical feasibility*. in *Photons Plus Ultrasound: Imaging and Sensing 2005: The Sixth Conference on Biomedical Thermoacoustics, Optoacoustics, and Acousto-optics*. 2005. International Society for Optics and Photonics.
142. Wang, X., D.L. Chamberland, and D.A. Jamadar, *Noninvasive photoacoustic tomography of human peripheral joints toward diagnosis of inflammatory arthritis.* Optics letters, 2007. **32**(20): p. 3002-3004.
143. Sun, Y., E. Sobel, and H. Jiang, *Quantitative three-dimensional photoacoustic tomography of the finger joints: an in vivo study.* Journal of biomedical optics, 2009. **14**(6): p. 064002-064002-5.
144. Xi, L. and H. Jiang, *High resolution three-dimensional photoacoustic imaging of human finger joints in vivo.* Applied Physics Letters, 2015. **107**(6): p. 063701.
145. van Es, P., et al., *Initial results of finger imaging using photoacoustic computed tomography.* Journal of biomedical optics, 2014. **19**(6): p. 060501-060501.
146. Yuan, J., et al., *Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization.* Journal of biomedical optics, 2013. **18**(8): p. 086001-086001.
147. Xiao, J., et al., *Quantitative two-dimensional photoacoustic tomography of osteoarthritis in the finger joints.* Optics express, 2010. **18**(14): p. 14359-14365.
148. Jo, J., et al., *A functional study of human inflammatory arthritis using photoacoustic imaging.* Scientific reports, 2017. **7**(1): p. 15026.
149. Akers, W.J., et al., *Multimodal sentinel lymph node mapping with single-photon emission computed tomography (SPECT)/computed tomography (CT) and photoacoustic tomography.* Translational Research, 2012. **159**(3): p. 175-181.
150. Su, J., et al., *Photoacoustic imaging of clinical metal needles in tissue.* Journal of biomedical optics, 2010. **15**(2): p. 021309-021309-6.
151. Wang, H., et al., *Three-dimensional interventional photoacoustic imaging for biopsy needle guidance with a linear array transducer.* Journal of biophotonics, 2019. **12**(12): p. e201900212.
152. Lediju Bell, M.A. and J. Shubert, *Photoacoustic-based visual servoing of a needle tip.* Scientific reports, 2018. **8**(1): p. 15519.
153. Bell, M.A.L., et al., *Localization of transcranial targets for photoacoustic-guided endonasal surgeries.* Photoacoustics, 2015. **3**(2): p. 78-87.
154. Shubert, J. and M.A. Lediju Bell, *Photoacoustic imaging of a human vertebra: implications for guiding spinal fusion surgeries.* Physics in Medicine & Biology, 2018. **63**(14): p. 144001.
155. Horiguchi, A., et al., *Pilot study of prostate cancer angiogenesis imaging using a photoacoustic imaging system.* Urology, 2017. **108**: p. 212-219.
156. Morse, P.M. and K.U. Ingard, *Theoretical acoustics*. 1986: Princeton university press.
157. Wang, L.V., *Photoacoustic imaging and spectroscopy*. Optical science and engineering. 2009, Boca Raton: CRC. xx, 499 p.
158. Queirós, D., *Image Reconstruction in Optoacoustics*. 2015, Technische Universität München.
159. Wang, K. and M.A. Anastasio, *Photoacoustic and thermoacoustic tomography: image formation principles*, in *Handbook of Mathematical Methods in Imaging*. 2015.
160. Kuchment, P. and L. Kunyansky, *Mathematics of photoacoustic and thermoacoustic tomography*, in *Handbook of mathematical methods in imaging*. 2011, Springer. p. 817-865.
161. Anastasio, M.A., et al., *Half-time image reconstruction in thermoacoustic tomography.* IEEE Trans Med Imaging, 2005. **24**(2): p. 199-210.

162. Rosenthal, A., V. Ntziachristos, and D. Razansky, *Acoustic inversion in optoacoustic tomography: A review.* Current Medical Imaging, 2013. **9**(4): p. 318-336.
163. Cox, B.T. and P.C. Beard, *Fast calculation of pulsed photoacoustic fields in fluids using k-space methods.* The Journal of the Acoustical Society of America, 2005. **117**(6): p. 3616-3627.
164. Paltauf, G., et al., *Iterative reconstruction algorithm for optoacoustic imaging.* The Journal of the Acoustical Society of America, 2002. **112**(4): p. 1536-1544.
165. O'Donnell, M., E. Jaynes, and J. Miller, *General relationships between ultrasonic attenuation and dispersion.* The Journal of the Acoustical Society of America, 1978. **63**(6): p. 1935-1937.
166. Deán-Ben, X.L., D. Razansky, and V. Ntziachristos, *The effects of acoustic attenuation in optoacoustic signals.* Physics in Medicine & Biology, 2011. **56**(18): p. 6129.
167. Burgholzer, P., et al. *Image reconstruction in photoacoustic tomography using integrating detectors accounting for frequency-dependent attenuation*. in *Photons Plus Ultrasound: Imaging and Sensing 2010*. 2010. SPIE.
168. Treeby, B.E. and B.T. Cox, *Modeling power law absorption and dispersion for acoustic propagation using the fractional Laplacian.* The Journal of the Acoustical Society of America, 2010. **127**(5): p. 2741-2748.
169. Haltmeier, M., R. Kowar, and L.V. Nguyen, *Iterative methods for photoacoustic tomography in attenuating acoustic media.* Inverse Problems, 2017. **33**(11): p. 115009.
170. Cox, B.T., S.R. Arridge, and P.C. Beard, *Estimating chromophore distributions from multiwavelength photoacoustic images.* JOSA A, 2009. **26**(2): p. 443-455.
171. Hoelen, C.G., et al. *Photoacoustic blood cell detection and imaging of blood vessels in phantom tissue*. in *Optical and Imaging Techniques for Biomonitoring III*. 1998. SPIE.
172. Xu, M. and L.V. Wang, *Universal back-projection algorithm for photoacoustic computed tomography.* Physical Review E, 2005. **71**(1): p. 016706.
173. Burgholzer, P., et al., *Exact and approximative imaging methods for photoacoustic tomography using an arbitrary detection surface.* Physical Review E, 2007. **75**(4): p. 046706.
174. Wang, K., et al., *Accelerating image reconstruction in three-dimensional optoacoustic tomography on graphics processing units.* Medical physics, 2013. **40**(2): p. 023301.
175. Rosenthal, A., V. Ntziachristos, and D. Razansky, *Model-based optoacoustic inversion with arbitrary-shape detectors.* Medical physics, 2011. **38**(7): p. 4285-4295.
176. Chowdhury, K.B., et al., *A Synthetic Total Impulse Response Characterization Method for Correction of Hand-held Optoacoustic Images.* IEEE Transactions on Medical Imaging, 2020.
177. Dean-Ben, X.L., et al., *Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography.* IEEE transactions on medical imaging, 2012. **31**(10): p. 1922-1928.
178. Provost, J. and F. Lesage, *The application of compressed sensing for photo-acoustic tomography.* IEEE transactions on medical imaging, 2008. **28**(4): p. 585-594.
179. Hansen, P.C., *Analysis of discrete ill-posed problems by means of the L-curve.* SIAM review, 1992. **34**(4): p. 561-580.
180. Han, Y., et al., *Sparsity-based acoustic inversion in cross-sectional multiscale optoacoustic imaging.* Medical physics, 2015. **42**(9): p. 5444-5452.
181. Wang, K., et al., *Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography.* Physics in Medicine & Biology, 2012. **57**(17): p. 5399.
182. Wright, S.J., R.D. Nowak, and M.A. Figueiredo, *Sparse reconstruction by separable approximation.* IEEE Transactions on signal processing, 2009. **57**(7): p. 2479-2493.
183. Kutyniok, G., W.-Q. Lim, and R. Reisenhofer, *Shearlab 3D: Faithful digital shearlet transforms based on compactly supported shearlets.* ACM Transactions on Mathematical Software (TOMS), 2016. **42**(1): p. 1-42.
184. Ding, L., et al., *Efficient non-negative constrained model-based inversion in optoacoustic tomography.* Physics in Medicine & Biology, 2015. **60**(17): p. 6733.

185. Lawson, C.L. and R.J. Hanson, *Solving least squares problems*. Prentice-Hall series in automatic computation. 1974, Englewood Cliffs, N.J.,: Prentice-Hall. xii, 340 p.
186. Wang, K., et al., *An imaging model incorporating ultrasonic transducer properties for three-dimensional optoacoustic tomography.* IEEE transactions on medical imaging, 2010. **30**(2): p. 203-214.
187. Bu, S., et al., *Model-based reconstruction integrated with fluence compensation for photoacoustic tomography.* IEEE transactions on biomedical engineering, 2012. **59**(5): p. 1354-1363.
188. Pinton, G.F., et al., *A heterogeneous nonlinear attenuating full-wave model of ultrasound.* IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 2009. **56**(3): p. 474-488.
189. Cox, B., et al., *Quantitative spectroscopic photoacoustic imaging: a review.* Journal of biomedical optics, 2012. **17**(6): p. 061202-061202.
190. Laufer, J., et al., *Quantitative spatially resolved measurement of tissue chromophore concentrations using photoacoustic spectroscopy: application to the measurement of blood oxygenation and haemoglobin concentration.* Physics in Medicine & Biology, 2006. **52**(1): p. 141.
191. Keshava, N., *A survey of spectral unmixing algorithms.* Lincoln laboratory journal, 2003. **14**(1): p. 55-78.
192. Tzoumas, S., et al., *Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues.* Nature communications, 2016. **7**: p. 12121.
193. Olefir, I., et al., *A Bayesian approach to eigenspectra optoacoustic tomography.* IEEE transactions on medical imaging, 2018. **37**(9): p. 2070-2079.
194. Olefir, I., et al., *Deep learning-based spectral unmixing for optoacoustic imaging of tissue oxygen saturation.* IEEE transactions on medical imaging, 2020. **39**(11): p. 3643-3654.
195. Pearson, K., *LIII. On lines and planes of closest fit to systems of points in space.* The London, Edinburgh, and Dublin philosophical magazine and journal of science, 1901. **2**(11): p. 559-572.
196. Hyvärinen, A. and E. Oja, *A fast fixed-point algorithm for independent component analysis.* Neural computation, 1997. **9**(7): p. 1483-1492.
197. Glatz, J., et al., *Blind source unmixing in multi-spectral optoacoustic tomography.* Optics express, 2011. **19**(4): p. 3175-3184.
198. Arabul, M., et al., *Unmixing multi-spectral photoacoustic sources in human carotid plaques using non-negative independent component analysis.* Photoacoustics, 2019. **15**: p. 100140.
199. An, L. and B.T. Cox, *Estimating relative chromophore concentrations from multiwavelength photoacoustic images using independent component analysis.* Journal of biomedical optics, 2018. **23**(7): p. 076007-076007.
200. Jüstel, D., et al., *Spotlight on nerves: Portable multispectral optoacoustic imaging of peripheral nerve vascularization and morphology.* arXiv preprint arXiv:2207.13978, 2022.
201. Grasso, V., R. Willumeit-Römer, and J. Jose, *Superpixel spectral unmixing framework for the volumetric assessment of tissue chromophores: A photoacoustic data-driven approach.* Photoacoustics, 2022. **26**: p. 100367.
202. Ron, A., et al., *Self-gated respiratory motion rejection for optoacoustic tomography.* Applied Sciences, 2019. **9**(13): p. 2737.
203. Taruttis, A., et al., *Motion clustering for deblurring multispectral optoacoustic tomography images of the mouse heart.* Journal of biomedical optics, 2012. **17**(1): p. 016009.
204. Özbek, A., X.L. Deán-Ben, and D. Razansky, *Optoacoustic imaging at kilohertz volumetric frame rates.* Optica, 2018. **5**(7): p. 857-863.
205. Erlöv, T., et al., *Regional motion correction for in vivo photoacoustic imaging in humans using interleaved ultrasound images.* Biomedical Optics Express, 2021. **12**(6): p. 3312-3322.

206. Jeng, G.-S., et al., *Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction.* Nature communications, 2021. **12**(1): p. 1-12.
207. Kukačka, J., et al., *Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients.* Photoacoustics, 2022. **26**: p. 100343.
208. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nature methods, 2020. **17**(3): p. 261-272.
209. Ilg, E., et al. *Flownet 2.0: Evolution of optical flow estimation with deep networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
210. Wang, Z., et al., *Image quality assessment: from error visibility to structural similarity.* IEEE transactions on image processing, 2004. **13**(4): p. 600-612.
211. Van der Walt, S., et al., *scikit-image: image processing in Python.* PeerJ, 2014. **2**: p. e453.
212. Fasoula, N.-A., et al., *Multicompartmental non-invasive sensing of postprandial lipemia in humans with multispectral optoacoustic tomography.* Molecular Metabolism, 2021. **47**: p. 101184.
213. Li, C. and L.V. Wang, *Photoacoustic tomography and sensing in biomedicine.* Physics in Medicine & Biology, 2009. **54**(19): p. R59.
214. Kratkiewicz, K., et al., *Technical considerations in the Verasonics research ultrasound platform for developing a photoacoustic imaging system.* Biomedical Optics Express, 2021. **12**(2): p. 1050-1084.
215. Manwar, R., et al., *Photoacoustic signal enhancement: towards utilization of low energy laser diodes in real-time photoacoustic imaging.* Sensors, 2018. **18**(10): p. 3498.
216. Manwar, R., M. Zafar, and Q. Xu, *Signal and image processing in biomedical photoacoustic imaging: a review.* Optics, 2020. **2**(1): p. 1-24.
217. Xavierselvan, M., M.K.A. Singh, and S. Mallidi, *In vivo tumor vascular imaging with light emitting diode-based photoacoustic imaging system.* Sensors, 2020. **20**(16): p. 4503.
218. Avants, B.B., et al., *A reproducible evaluation of ANTs similarity metric performance in brain image registration.* Neuroimage, 2011. **54**(3): p. 2033-2044.
219. McCormick, M.M., et al., *ITK: enabling reproducible research and open science.* Frontiers in neuroinformatics, 2014. **8**: p. 13.
220. Oraevsky, A.A., et al. *Laser optoacoustic imaging of breast cancer in vivo*. in *Biomedical Optoacoustics II*. 2001. International Society for Optics and Photonics.
221. Teh, W. and A. Wilson, *The role of ultrasound in breast cancer screening. A consensus statement by the European Group for Breast Cancer Screening.* European journal of cancer, 1998. **34**(4): p. 449-450.
222. Lee, C.H., et al., *Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer.* Journal of the American college of radiology, 2010. **7**(1): p. 18-27.
223. Hooley, R.J., L. Andrejeva, and L.M. Scoutt, *Breast cancer screening and problem solving using mammography, ultrasound, and magnetic resonance imaging.* Ultrasound quarterly, 2011. **27**(1): p. 23-47.
224. Dehner, C., et al., *DeepMB: Deep neural network for real-time model-based optoacoustic image reconstruction with adjustable speed of sound.* arXiv preprint arXiv:2206.14485, 2022.
225. Goss, S., L. Frizzell, and F. Dunn, *Ultrasonic absorption and attenuation in mammalian tissues.* Ultrasound in medicine & biology, 1979. **5**(2): p. 181-186.
226. Aguirre, J., et al., *Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy.* Nature Biomedical Engineering, 2017. **1**(5): p. 1-8.
227. Moore, M.J., et al., *Photoacoustic F-mode imaging for scale specific contrast in biological systems.* Communications Physics, 2019. **2**(1): p. 1-10.
228. Longo, A., D. Jüstel, and V. Ntziachristos, *Disentangling the frequency content in optoacoustics.* IEEE Transactions on Medical Imaging, 2022.

229. Nguyen, H.N.Y. and W. Steenbergen, *Reducing artifacts in photoacoustic imaging by using multi-wavelength excitation and transducer displacement.* Biomedical optics express, 2019. **10**(7): p. 3124-3138.
230. Paltauf, G. and R. Nuster, *Artifact removal in photoacoustic section imaging by combining an integrating cylindrical detector with model-based reconstruction.* Journal of Biomedical Optics, 2014. **19**(2): p. 026014.
231. Zalev, J. and M.C. Kolios, *Opto-acoustic image reconstruction and motion tracking using convex optimization.* IEEE Transactions on Computational Imaging, 2021. **7**: p. 1161-1175.
232. Kimpe, T. and T. Tuytschaever, *Increasing the number of gray shades in medical display systems—how much is enough?* Journal of digital imaging, 2007. **20**(4): p. 422-432.
233. Barten, P.G. *Physical model for the contrast sensitivity of the human eye*. in *Human Vision, Visual Processing, and Digital Display III*. 1992. SPIE.
234. Hall, E.L., et al., *A survey of preprocessing and feature extraction techniques for radiographic images.* IEEE Transactions on Computers, 1971. **100**(9): p. 1032-1044.
235. Hummel, R., *Image enhancement by histogram transformation.* Unknown, 1975.
236. Ketcham, D.J., R.W. Lowe, and J.W. Weber, *Image enhancement techniques for cockpit displays*. 1974, Hughes Aircraft Co Culver City Ca Display Systems Lab.
237. Pizer, S.M., *Intensity mappings for the display of medical images.* Functional Mapping of Organ Systems and Other Computer Topics, 1981: p. 205-217.
238. Merčep, E., et al., *Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals.* Light: Science & Applications, 2019. **8**(1): p. 1-12.
239. Park, S., et al., *Normalization of optical fluence distribution for three-dimensional functional optoacoustic tomography of the breast.* Journal of biomedical optics, 2022. **27**(3): p. 036001.
240. Pattyn, A., et al., *Model-based optical and acoustical compensation for photoacoustic tomography of heterogeneous mediums.* Photoacoustics, 2021. **23**: p. 100275.
241. Cox, B.T., et al., *Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method.* Applied optics, 2006. **45**(8): p. 1866-1875.
242. Yao, L., Y. Sun, and H. Jiang, *Quantitative photoacoustic tomography based on the radiative transfer equation.* Optics letters, 2009. **34**(12): p. 1765-1767.
243. Yuan, Z., Q. Wang, and H. Jiang, *Reconstruction of optical absorption coefficient maps of heterogeneous media by photoacoustic tomography coupled with diffusion equation based regularized Newton method.* Optics express, 2007. **15**(26): p. 18076-18081.
244. Rosenthal, A., D. Razansky, and V. Ntziachristos, *Quantitative optoacoustic signal extraction using sparse signal representation.* IEEE transactions on medical imaging, 2009. **28**(12): p. 1997-2006.
245. Nagy, J., et al., *Why are tumour blood vessels abnormal and why is it important to know?* British journal of cancer, 2009. **100**(6): p. 865-869.
246. Schellenberg, M., et al., *Semantic segmentation of multispectral photoacoustic images using deep learning.* Photoacoustics, 2022. **26**: p. 100341.
247. Jose, J., et al., *Speed-of-sound compensated photoacoustic tomography for accurate imaging.* Medical physics, 2012. **39**(12): p. 7262-7271.
248. Deán-Ben, X.L., V. Ntziachristos, and D. Razansky, *Effects of small variations of speed of sound in optoacoustic tomographic imaging.* Medical physics, 2014. **41**(7): p. 073301.
249. Brochu, F.M., et al., *Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography.* IEEE transactions on medical imaging, 2016. **36**(1): p. 322-331.
250. Liang, Z., et al., *Automatic 3-D segmentation and volumetric light fluence correction for photoacoustic tomography based on optimal 3-D graph search.* Medical Image Analysis, 2022. **75**: p. 102275.
251. Laufer, J., et al., *Quantitative determination of chromophore concentrations from 2D photoacoustic images using a nonlinear model-based inversion scheme.* Applied optics, 2010. **49**(8): p. 1219-1233.

252. Perekatova, V., et al. *Quantitative Characterization of Age-Related Changes in Peripheral Vessels of a Human Palm Using Raster-Scan Optoacoustic Angiography*. in *Photonics*. 2022. MDPI.
253. Estrada, H., et al., *Intravital optoacoustic and ultrasound bio-microscopy reveal radiation-inhibited skull angiogenesis.* Bone, 2020. **133**: p. 115251.
254. Rebling, J., et al., *Long-term imaging of wound angiogenesis with large scale optoacoustic microscopy.* Advanced Science, 2021. **8**(13): p. 2004226.
255. Sun, M., et al., *Full three-dimensional segmentation and quantification of tumor vessels for photoacoustic images.* Photoacoustics, 2020. **20**: p. 100212.
256. Lafci, B., et al., *Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images.* IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 2020. **68**(3): p. 688-696.
257. Raumonen, P. and T. Tarvainen, *Segmentation of vessel structures from photoacoustic images with reliability assessment.* Biomedical optics express, 2018. **9**(7): p. 2887-2904.
258. Ly, C.D., et al., *Full-view in vivo skin and blood vessels profile segmentation in photoacoustic imaging based on deep learning.* Photoacoustics, 2022. **25**: p. 100310.
259. Chlis, N.-K., et al., *A sparse deep learning approach for automatic segmentation of human vasculature in multispectral optoacoustic tomography.* Photoacoustics, 2020. **20**: p. 100203.
260. Rajpurkar, P., et al., *AI in health and medicine.* Nature Medicine, 2022. **28**(1): p. 31-38.
261. Willemink, M.J., et al., *Preparing medical imaging data for machine learning.* Radiology, 2020. **295**(1): p. 4-15.
262. Jungo, A., et al. *On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
263. Lin, L. and L.V. Wang, *The emerging role of photoacoustic imaging in clinical oncology.* Nature Reviews Clinical Oncology, 2022. **19**(6): p. 365-384.
264. Sangha, G.S., et al., *Assessment of photoacoustic tomography contrast for breast tissue imaging using 3D correlative virtual histology.* Scientific reports, 2022. **12**(1): p. 1-13.
265. Kang, J., et al., *Photoacoustic imaging of breast microcalcifications: a validation study with 3-dimensional ex vivo data and spectrophotometric measurement.* Journal of biophotonics, 2015. **8**(1-2): p. 71-80.
266. Heijblom, M., et al., *Photoacoustic image patterns of breast carcinoma and comparisons with Magnetic Resonance Imaging and vascular stained histopathology.* Sci Rep, 2015. **5**: p. 11778.
267. Menezes, G.L., et al., *Optoacoustic imaging of the breast: correlation with histopathology and histopathologic biomarkers.* European radiology, 2019. **29**(12): p. 6728-6740.
268. Longo, A., et al., *Assessment of hessian-based Frangi vesselness filter in optoacoustic imaging.* Photoacoustics, 2020. **20**: p. 100200.
269. Vogt, W.C., et al., *Biologically relevant photoacoustic imaging phantoms with tunable optical and acoustic properties.* Journal of biomedical optics, 2016. **21**(10): p. 101405.
270. Singh, M.K.A. and W. Steenbergen, *Photoacoustic-guided focused ultrasound (PAFUSion) for identifying reflection artifacts in photoacoustic imaging.* Photoacoustics, 2015. **3**(4): p. 123-131.
271. Nguyen, H.N.Y., A. Hussain, and W. Steenbergen, *Reflection artifact identification in photoacoustic imaging using multi-wavelength excitation.* Biomedical optics express, 2018. **9**(10): p. 4613-4630.
272. Frikel, J.r. and E.T. Quinto, *Artifacts in incomplete data tomography with applications to photoacoustic tomography and sonar.* SIAM Journal on Applied Mathematics, 2015. **75**(2): p. 703-725.
273. Gröhl, J., et al., *Deep learning for biomedical photoacoustic imaging: A review.* Photoacoustics, 2021. **22**: p. 100241.
274. Boink, Y.E., S. Manohar, and C. Brune, *A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation.* IEEE transactions on medical imaging, 2019. **39**(1): p. 129-139.

275. Zhang, J., et al. *Pathology study for blood vessel of ocular fundus images by photoacoustic tomography*. in *2018 IEEE International Ultrasonics Symposium (IUS)*. 2018. IEEE.
276. Cuadros, J. and G. Bresnick, *EyePACS: an adaptable telemedicine system for diabetic retinopathy screening.* Journal of diabetes science and technology, 2009. **3**(3): p. 509-516.
277. Zhou, Z., et al. *Models genesis: Generic autodidactic models for 3d medical image analysis*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019. Springer.
278. Taleb, A., et al., *3d self-supervised methods for medical imaging.* arXiv preprint arXiv:2006.03829, 2020.
279. Holmberg, O.G., et al., *Self-supervised retinal thickness prediction enables deep learning from unlabelled data to boost classification of diabetic retinopathy.* Nature Machine Intelligence, 2020. **2**(11): p. 719-726.
280. Hervella, Á.S., et al. *Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images*. in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020. IEEE.
281. Chen, T., et al., *A simple framework for contrastive learning of visual representations.* arXiv preprint arXiv:2002.05709, 2020.
282. He, K., et al. *Momentum contrast for unsupervised visual representation learning*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
283. Chen, T., et al., *Big self-supervised models are strong semi-supervised learners.* arXiv preprint arXiv:2006.10029, 2020.
284. Galdran, A., et al., *The Little W-Net That Could: State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models.* arXiv preprint arXiv:2009.01907, 2020.
285. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
286. Chen, X., et al., *Improved baselines with momentum contrastive learning.* arXiv preprint arXiv:2003.04297, 2020.
287. Sohn, K. *Improved deep metric learning with multi-class n-pair loss objective*. in *Advances in neural information processing systems*. 2016.
288. Oord, A.v.d., Y. Li, and O. Vinyals, *Representation learning with contrastive predictive coding.* arXiv preprint arXiv:1807.03748, 2018.
289. Staal, J., et al., *Ridge-based vessel segmentation in color images of the retina.* IEEE transactions on medical imaging, 2004. **23**(4): p. 501-509.
290. Odstrcilik, J., et al., *Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database.* IET Image Processing, 2013. **7**(4): p. 373-383.
291. Fraz, M.M., et al., *An ensemble classification-based approach applied to retinal blood vessel segmentation.* IEEE Transactions on Biomedical Engineering, 2012. **59**(9): p. 2538-2548.
292. Hoover, A., V. Kouznetsova, and M. Goldbaum, *Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response.* IEEE Transactions on Medical imaging, 2000. **19**(3): p. 203-210.
293. Orlando, J.I., et al. *Towards a glaucoma risk index based on simulated hemodynamics from fundus images*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
294. Estrada, R., et al., *Retinal artery-vein classification via topology estimation.* IEEE transactions on medical imaging, 2015. **34**(12): p. 2518-2534.
295. Holm, S., et al., *DR HAGIS—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients.* Journal of Medical Imaging, 2017. **4**(1): p. 014503.
296. Porwal, P., et al., *Idrid: Diabetic retinopathy–segmentation and grading challenge.* Medical image analysis, 2020. **59**: p. 101561.

297. Wu, Z., et al. *Unsupervised feature learning via non-parametric instance discrimination*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
298. He, K., et al. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
299. Park, K.-B., S.H. Choi, and J.Y. Lee, *M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks.* IEEE Access, 2020. **8**: p. 146308-146322.
300. Wang, D., et al., *Hard attention net for automatic retinal vessel segmentation.* IEEE Journal of Biomedical and Health Informatics, 2020. **24**(12): p. 3384-3396.
301. Laibacher, T. and A. Anjos, *On the Evaluation and Real-World Usage Scenarios of Deep Vessel Segmentation for Retinography.* arXiv preprint arXiv:1909.03856, 2019.
302. Shin, S.Y., et al., *Deep vessel segmentation by learning graphical connectivity.* Medical image analysis, 2019. **58**: p. 101556.
303. Wang, B., S. Qiu, and H. He. *Dual encoding u-net for retinal vessel segmentation*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019. Springer.
304. Jin, Q., et al., *DUNet: A deformable network for retinal vessel segmentation.* Knowledge-Based Systems, 2019. **178**: p. 149-162.
305. Trucco, E., et al., *Validating retinal fundus image analysis algorithms: issues and a proposal.* Investigative ophthalmology & visual science, 2013. **54**(5): p. 3546-3559.
306. Maier-Hein, L., et al., *Why rankings of biomedical image analysis competitions should be interpreted with care.* Nature communications, 2018. **9**(1): p. 1-13.
307. Reguant, R., S. Brunak, and S. Saha, *Understanding inherent image features in CNN-based assessment of diabetic retinopathy.* Scientific Reports, 2021. **11**(1): p. 1-12.
308. Quellec, G., et al., *Deep image mining for diabetic retinopathy screening.* Medical image analysis, 2017. **39**: p. 178-193.
309. Kolesnikov, A., X. Zhai, and L. Beyer. *Revisiting self-supervised visual representation learning*. in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2019.
310. Zhao, N., et al., *Distilling localization for self-supervised representation learning.* arXiv preprint arXiv:2004.06638, 2020.
311. Wallace, B. and B. Hariharan. *Extending and analyzing self-supervised learning across domains*. in *European Conference on Computer Vision*. 2020. Springer.
312. Zhao, N., et al., *What makes instance discrimination good for transfer learning?* arXiv preprint arXiv:2006.06606, 2020.
313. Newell, A. and J. Deng. *How useful is self-supervised pretraining for visual tasks?* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
314. Raghu, M., et al., *Transfusion: Understanding transfer learning for medical imaging.* arXiv preprint arXiv:1902.07208, 2019.
315. Russo, A., et al., *A novel device to exploit the smartphone camera for fundus photography.* Journal of ophthalmology, 2015. **2015**.
316. Haddock, L.J., D.Y. Kim, and S. Mukai, *Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes.* Journal of ophthalmology, 2013. **2013**.
317. Chandrakanth, P., et al., *Trash to treasure Retcam.* Indian journal of ophthalmology, 2019. **67**(4): p. 541.
318. Hinton, G., O. Vinyals, and J. Dean, *Distilling the knowledge in a neural network.* arXiv preprint arXiv:1503.02531, 2015.
319. Armanious, K., et al., *MedGAN: Medical image translation using GANs.* Computerized medical imaging and graphics, 2020. **79**: p. 101684.

320. Gatys, L.A., A.S. Ecker, and M. Bethge. *Image style transfer using convolutional neural networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
321. Rivenson, Y., et al., *Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning.* Nature biomedical engineering, 2019. **3**(6): p. 466-477.
322. Kang, L., et al., *Deep learning enables ultraviolet photoacoustic microscopy based histological imaging with near real-time virtual staining.* Photoacoustics, 2022. **25**: p. 100308.
323. Martell, M.T., et al. *Deep learning-enabled realistic virtual histology with ultraviolet scattering and photoacoustic remote sensing microscopy*. in *Microscopy Histopathology and Analytics*. 2022. Optica Publishing Group.
324. Klimovskaia, A., et al. *Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data*. in *Medical Imaging with Deep Learning*. 2021.
325. Oraevsky, A.A., et al. *Two-dimensional optoacoustic tomography: transducer array and image reconstruction algorithm*. in *Laser-Tissue Interaction X: Photochemical, Photothermal, and Photomechanical*. 1999. SPIE.
326. Kitai, T., et al., *Photoacoustic mammography: initial clinical results.* Breast Cancer, 2014. **21**(2): p. 146-53.
327. Fukutani, K., et al. *Characterization of photoacoustic tomography system with dual illumination*. in *Photons Plus Ultrasound: Imaging and Sensing 2011*. 2011. International Society for Optics and Photonics.
328. Ermilov, S.A., et al., *Laser optoacoustic imaging system for detection of breast cancer.* J Biomed Opt, 2009. **14**(2): p. 024007.
329. Xi, L., et al., *Design and evaluation of a hybrid photoacoustic tomography and diffuse optical tomography system for breast cancer detection.* Medical physics, 2012. **39**(5): p. 2584-2594.
330. Li, X., et al., *High resolution functional photoacoustic tomography of breast cancer.* Med Phys, 2015. **42**(9): p. 5321-8.
331. Dogan, B.E., et al., *Optoacoustic imaging and gray-scale US features of breast cancers: correlation with molecular subtypes.* Radiology, 2019. **292**(3): p. 564-572.
332. Li, Y., et al., *Secretin-activated brown fat mediates prandial thermogenesis to induce satiation.* Cell, 2018. **175**(6): p. 1561-1574. e12.
333. Yang, H., et al., *Soft ultrasound priors in optoacoustic reconstruction: Improving clinical vascular imaging.* Photoacoustics, 2020: p. 100172.
334. Masthoff, M., et al., *Multispectral optoacoustic tomography of systemic sclerosis.* Journal of biophotonics, 2018. **11**(11): p. e201800155.
335. Regensburger, A.P., et al., *Detection of collagens by multispectral optoacoustic tomography as an imaging biomarker for Duchenne muscular dystrophy.* Nature Medicine, 2019. **25**(12): p. 1905-1915.
336. Mercep, E., et al., *Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays.* IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 2015. **62**(9): p. 1651-1661.
337. Buehler, A., et al., *Model-based optoacoustic inversions with incomplete projection data.* Medical physics, 2011. **38**(3): p. 1694-1704.
338. Xu, Y., et al., *Reconstructions in limited-view thermoacoustic tomography.* Med Phys, 2004. **31**(4): p. 724-33.
339. Generali, D., et al., *Hypoxia-inducible factor-1α expression predicts a poor response to primary chemoendocrine therapy and disease-free survival in primary human breast cancer.* Clinical Cancer Research, 2006. **12**(15): p. 4562-4568.
340. Ueda, S., et al., *Baseline tumor oxygen saturation correlates with a pathologic complete response in breast cancer patients undergoing neoadjuvant chemotherapy.* Cancer research, 2012. **72**(17): p. 4318-4328.
341. Gu, X., et al., *Differentiation of cysts from solid tumors in the breast with diffuse optical tomography.* Academic radiology, 2004. **11**(1): p. 53-60.

342. Taroni, P., et al., *Time-resolved optical mammography between 637 and 985 nm: clinical study on the detection and identification of breast lesions.* Physics in Medicine & Biology, 2005. **50**(11): p. 2469.
343. van de Ven, S., et al., *Diffuse optical tomography of the breast: initial validation in benign cysts.* Molecular imaging and biology, 2009. **11**(2): p. 64-70.
344. Vijayaraghavan, G.R., et al., *Unifocal Invasive Lobular Carcinoma: Tumor Size Concordance Between Preoperative Ultrasound Imaging and Postoperative Pathology.* Clinical breast cancer, 2018. **18**(6): p. e1367-e1372.
345. Pritt, B., et al., *Influence of breast cancer histology on the relationship between ultrasound and pathology tumor size measurements.* Modern pathology, 2004. **17**(8): p. 905-910.
346. Goh, Y., et al., *Multispectral optoacoustic tomography in assessment of breast tumor margins during breast-conserving surgery: a first-in-human case study.* Clinical breast cancer, 2018. **18**(6): p. e1247-e1250.
347. Azzi, S., J.K. Hebda, and J. Gavard, *Vascular permeability and drug delivery in cancers.* Frontiers in oncology, 2013. **3**: p. 211.
348. Fukumura, D. and R.K. Jain, *Tumor microenvironment abnormalities: causes, consequences, and strategies to normalize.* Journal of cellular biochemistry, 2007. **101**(4): p. 937-949.
349. Allen, M.D. and L.J. Jones, *The role of inflammation in progression of breast cancer: Friend or foe?* International journal of oncology, 2015. **47**(3): p. 797-805.
350. Pinder, S., et al., *A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial.* British journal of cancer, 2010. **103**(1): p. 94-100.
351. Van Bockstal, M., et al., *Stromal characteristics are adequate prognosticators for recurrence risk in ductal carcinoma in situ of the breast.* European Journal of Surgical Oncology, 2019. **45**(4): p. 550-559.
352. Hupple, C., et al., *Dynamic imaging of chronic inflammation and tumour therapeutic response with multispectral optoacoustic tomography (MSOT).* Journal of Nuclear Medicine, 2015. **56**(supplement 3): p. 620-620.
353. Sleeman, J., *The lymph node as a bridgehead in the metastatic dissemination of tumors.* Lymphatic metastasis and sentinel lymphonodectomy, 2000: p. 55-81.
354. Beahrs, O.H., M.H. Myers, and American Joint Committee on Cancer., *Manual for staging of cancer*. 2nd ed. 1983, Philadelphia: Lippincott. xvii, 250 p.
355. Ho, A.S., et al., *Metastatic lymph node burden and survival in oral cavity cancer.* Journal of Clinical Oncology, 2017. **35**(31): p. 3601.
356. de Bree, R., et al., *Elective neck dissection in oral squamous cell carcinoma: Past, present and future.* Oral oncology, 2019. **90**: p. 87-93.
357. De Bree, R., et al., *Advances in diagnostic modalities to detect occult lymph node metastases in head and neck squamous cell carcinoma.* Head & neck, 2015. **37**(12): p. 1829-1839.
358. Weiss, M.H., L.B. Harrison, and R.S. Isaacs, *Use of decision analysis in planning a management strategy for the stage N0 neck.* Arch Otolaryngol Head Neck Surg, 1994. **120**(7): p. 699-702.
359. D'Cruz, A.K., et al., *Elective versus therapeutic neck dissection in node-negative oral cancer.* New England Journal of Medicine, 2015. **373**(6): p. 521-529.
360. Schiefke, F., et al., *Function, postoperative morbidity, and quality of life after cervical sentinel node biopsy and after selective neck dissection.* Head Neck, 2009. **31**(4): p. 503-12.
361. Bradley, P.J., et al., *Neck treatment and shoulder morbidity: still a challenge.* Head & neck, 2011. **33**(7): p. 1060-1067.
362. Alkureishi, L.W., et al., *Joint practice guidelines for radionuclide lymphoscintigraphy for sentinel node localization in oral/oropharyngeal squamous cell carcinoma.* European journal of nuclear medicine and molecular imaging, 2009. **36**(11): p. 1915-1936.

363. Goldberg, J.I., et al., *Morbidity of sentinel node biopsy: relationship between number of excised lymph nodes and patient perceptions of lymphedema.* Annals of surgical oncology, 2011. **18**(10): p. 2866-2872.
364. de Bondt, R.B., et al., *Detection of lymph node metastases in head and neck cancer: a meta-analysis comparing US, USgFNAC, CT and MR imaging.* Eur J Radiol, 2007. **64**(2): p. 266-72.
365. Luke, G.P., et al., *Silica-coated gold nanoplates as stable photoacoustic contrast agents for sentinel lymph node imaging.* Nanotechnology, 2013. **24**(45): p. 455101.
366. Luke, G.P., et al., *Sentinel lymph node biopsy revisited: ultrasound-guided photoacoustic detection of micrometastases using molecularly targeted plasmonic nanosensors.* Cancer research, 2014. **74**(19): p. 5397-5408.
367. Luke, G.P. and S.Y. Emelianov, *Label-free detection of lymph node metastases with US-guided functional photoacoustic imaging.* Radiology, 2015. **277**(2): p. 435.
368. Guggenheim, J.A., et al., *Photoacoustic imaging of human lymph nodes with endogenous lipid and hemoglobin contrast.* Journal of biomedical optics, 2015. **20**(5): p. 050504.
369. Rosenthal, E.L., et al., *Sensitivity and specificity of cetuximab-IRDye800CW to identify regional metastatic disease in head and neck cancer.* Clinical Cancer Research, 2017. **23**(16): p. 4744-4752.
370. Nishio, N., et al., *Optical molecular imaging can differentiate metastatic from benign lymph nodes in head and neck cancer.* Nature communications, 2019. **10**(1): p. 1-10.
371. Nishio, N., et al., *Photoacoustic Molecular Imaging for the Identification of Lymph Node Metastasis in Head and Neck Cancer Using an Anti-EGFR Antibody–Dye Conjugate.* Journal of Nuclear Medicine, 2021. **62**(5): p. 648-655.
372. Linssen, M.D., et al., *Roadmap for the development and clinical translation of optical tracers cetuximab-800CW and trastuzumab-800CW.* Journal of Nuclear Medicine, 2019. **60**(3): p. 418-423.
373. Voskuil, F.J., et al., *Fluorescence-guided imaging for resection margin evaluation in head and neck cancer patients using cetuximab-800CW: a quantitative dose-escalation study.* Theranostics, 2020. **10**(9): p. 3994.
374. Mantri, Y. and J.V. Jokerst, *Impact of skin tone on photoacoustic oximetry and tools to minimize bias.* Biomedical optics express, 2022. **13**(2): p. 875-887.
375. Fuenzalida Werner, J.P., et al., *Challenging a preconception: optoacoustic spectrum differs from the optical absorption spectrum of proteins and dyes for molecular imaging.* Analytical Chemistry, 2020. **92**(15): p. 10717-10724.
376. Vonk, J., et al., *Epidermal Growth Factor Receptor–Targeted Fluorescence Molecular Imaging for Postoperative Lymph Node Assessment in Patients with Oral Cancer.* Journal of Nuclear Medicine, 2022. **63**(5): p. 672-678.
377. Karaman, S. and M. Detmar, *Mechanisms of lymphatic metastasis.* The Journal of clinical investigation, 2014. **124**(3): p. 922-928.
378. Zhang, S., et al., *The relationship of lymphatic vessel density, lymphovascular invasion, and lymph node metastasis in breast cancer: a systematic review and meta-analysis.* Oncotarget, 2017. **8**(2): p. 2863.
379. Singleton, D.C., A. Macann, and W.R. Wilson, *Therapeutic targeting of the hypoxic tumour microenvironment.* Nature Reviews Clinical Oncology, 2021. **18**(12): p. 751-772.
380. Bhandari, V., et al., *Molecular landmarks of tumor hypoxia across cancer types.* Nature genetics, 2019. **51**(2): p. 308-318.
381. Dumani, D.S., I.-C. Sun, and S.Y. Emelianov, *Ultrasound-guided immunofunctional photoacoustic imaging for diagnosis of lymph node metastases.* Nanoscale, 2019. **11**(24): p. 11649-11659.
382. Luke, G.P., D. Yeager, and S.Y. Emelianov, *Biomedical applications of photoacoustic imaging with exogenous contrast agents.* Annals of biomedical engineering, 2012. **40**(2): p. 422-437.
383. Tichauer, K.M., et al., *Quantitative in vivo cell-surface receptor imaging in oncology: kinetic modeling and paired-agent principles from nuclear medicine and optical imaging.* Physics in Medicine & Biology, 2015. **60**(14): p. R239.

384. Sampath, L., W. Wang, and E.M. Sevick-Muraca, *Near infrared fluorescent optical imaging for nodal staging.* Journal of biomedical optics, 2008. **13**(4): p. 041312.
385. Pressman, D., E.D. Day, and M. Blau, *The use of paired labeling in the determination of tumor-localizing antibodies.* Cancer research, 1957. **17**(9): p. 845-850.
386. Tichauer, K.M., et al., *Microscopic lymph node tumor burden quantified by macroscopic dual-tracer molecular imaging.* Nature medicine, 2014. **20**(11): p. 1348-1353.
387. Nishio, N., et al., *Diagnosis of cervical lymph node metastases in head and neck cancer with ultrasonic measurement of lymph node volume.* Auris Nasus Larynx, 2019. **46**(6): p. 889-895.
388. Boeve, K., et al., *High sensitivity and negative predictive value of sentinel lymph node biopsy in a retrospective early stage oral cavity cancer cohort in the Northern Netherlands.* Clinical Otolaryngology, 2018. **43**(4): p. 1080-1087.
389. Stoeckli, S.J., et al., *Sentinel lymph node evaluation in squamous cell carcinoma of the head and neck.* Otolaryngology—Head and Neck Surgery, 2001. **125**(3): p. 221-226.

Permission to reuse materials from published works

Figure 2.1a was reproduced from from Welcome Collection (https:// wellcomecollection.org/works/atkewunb), published under Creative Commins CC-BY license (https://creativecommons.org/licenses/by/4.0).

Figure 2.1b was reproduced from Brunton, John. "A new otoscope or speculum auris". *The Retrospect of Practical Medicine and Surgery.* Braithwaite, William and Braithwaite, James (eds.) vol. 53, pp 214–216. Simpkin, Marshall & Company, London, 1866. Digitized copy provided by Google Books in the public domain. Accessed at https://books.google.de/books?id=RxQDAAAAYAAJ.

Figure 2.1c was reproduced from von Helmholtz, H., „Beschreibung eines Augen-Spiegels zur Untersuchung der Netzhaut im lebenden Auge." A. Förstner'sche Verlagsbuchhandlung, Berlin, 1851. Digitized copy provided by Internet Archive in the public domain: https://archive.org/details/beschreibungeine00helm.

Figure 2.1d was reproduced from Eder, Josef Maria and Valenta, Eduard, „Versuche über Photographie mittelst der Röntgen'schen Strahlen". R. Lechner, Wien, 1896. Digitized copy provided by Jeff Behary through the Internet Archive: https://archive.org/details/eder-valenta-photography-rontgen-rays.

Figures 4.1–4.4, 5.1, 5.3, 7.3–7.6 were reproduced from the publication *Kukačka et al. "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients." Photoacoustics 26 (2022): 100343*, published by Elsevier under the Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/).

Figures 7.1 and 7.2 were reproduced from the publication *Manohar, Srirang, and Maura Dantuma. "Current and future trends in photoacoustic breast imaging." Photoacoustics 16 (2019): 100134*, published by Elsevier under the Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/).

Figures 7.8–7.14 were reproduced from the publication *Vonk et al. " Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: A proof-of-concept study" Photoacoustics 26 (2022): 100362*, published by Elsevier under the Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/).