Department of Physics and Astronomy

University of Heidelberg

Master Thesis

in Physics

submitted by

Dominik Jens Elias Waibel

born in Tübingen

March 2018

Photoacoustic image reconstruction to solve the acoustic inverse problem with deep learning

This master thesis has been carried out by Dominik Jens Elias Waibel

at the

German Cancer Research Center

under the supervision of

Prof. Dr. Peter Bachert

and

Prof. Dr. Lena Maier-Hein

Photoacoustic image reconstruction to solve the acoustic inverse problem with deep learning

Photoacoustic (PA) imaging is a promising and emerging technique for detection and characterization of diseases, such as superficial tumors. It combines the strengths of optical imaging and ultrasound: high optical contrast as well as high imaging depth and spatial resolution. Handheld PA devices, which are popular due to their versatility and ease of use, only provide a limited field of view caused by their sensor geometry. This restrain in hardware as well as the optical and acoustical attenuation in tissue limit the ability to reconstruct the initial pressure distribution, which is necessary for quantification of the underlying tissue. Today only approximations can be reconstructed with state-of-the-art methods. In this master thesis a novel approach for reconstructing the initial pressure distribution in tissue from limited view PA data is presented. It is based on machine learning, using a fully convolutional deep neural network with a U-Net-like architecture. The data necessary to train this network and validate the approach was generated *in silico*. The Initial pressure distribution of single and multiple blood vessels and the propagation of ultrasound was simulated in tissue. The results show a promising path to solve the acoustic inverse problem and offer a quantitative and qualitative improvement over state-of-the-art techniques. The method developed in this thesis can be a further step on the road towards clinical quantitative photoacoustic imaging.

Rekonstruktion photoakustischer Bilder zur Lösung des akustisch inversen Problems mit maschinellen Lernverfahren

Die photoakustische (PA) Bildgebung ist eine neuartige und vielversprechende Technologie für die Detektion und Charakterisierung von Krankheiten, wie beispielsweise oberflächennahen Tumoren. Sie kombiniert die Vorteile von optischer Bildgebung und Ultraschall: hohen optischen Kontrast mit großer Bildtiefe und hoher räumliche Auflösung. Herkömmliche Ultraschall-Sonden, die wegen ihrer Vielseitigkeit und Benutzerfreundlichkeit populär sind, können aufgrund ihrer Sensorgeometrie nur Bilder mit einem limitierten Blickwinkel aufnehmen. Wegen dieser Beschränkung und der optischen und akustischen Abschwächung im Gewebe ist aktuell keine genaue Rekonstruktion der initialen Druckverteilung möglich. In dieser Masterarbeit wird ein Lösungsansatz für die Rekonstruktion der initialen Druckverteilung im Gewebe aus PA Bilddaten mit limitiertem Blickwinkel präsentiert. Dieser basiert auf einem maschinellem Lernverfahren mit einer U-Net Architektur. Um das Netzwerk zu trainieren und um unseren Ansatz zu validieren wurden *in silico* Bilddaten von Blutgefäßen in Gewebe generiert. Die Ergebnisse dieser Arbeit zeigen eine qualitative und quantitative Verbesserung gegenüber aktuell eingesetzter Technik. Damit wurde in dieser Masterarbeit ein weiterer Schritt gemacht, um die quantitative Photoakustik klinisch anzuwenden.

Contents

List of abbreviations used in this master thesis

AR-PAM Acoustic Resolution Photoacoustic Microscopy

CT Computed Tomography

FFT Fast Fourier Transformation

IQR Interquartile Range

LReLu Leaky Rectified Linear Unit

MITK Medical Imaging Interaction Toolkit

MRI Magnetic Resonance Imaging

OR-PAM Optic Resolution Photoacoustic Microscopy

PA Photoacoustic

PAE Photoacoustic Effect

PA-FC Photoacoustic Flow Cytometry

PAI Photoacoustic Imaging

PAM Photoacoustic Microscopy

PAT Photoacoustic Tomography

PET Positron Emission Tomography

PS Pseudo Spectral (Method)

qPAT quantitative PAT

ReLu Rectified Linear Unit

US Ultrasound

Acknowledgments

I am greatly thankful for Professor Bachert's supervision of my master thesis from a physics' point of view, on top of his already impressive workload. He inspired me to specialize in medical physics and accompanied my studies and exams in this field, for which my enthusiasm is ever growing.

I am deeply grateful for Professor Lena Maier-Hein's remarkable support during my master thesis. As head of the 'Computer assisted Medical Interventions' Department at DKFZ she provided me with opportunity to join her lab and work together with all the brilliant people in her group. It was a pleasure to join her exciting research project.

My special thank goes to Janek Gröhl for all that I have learned from him. His support was always superb and he was endlessly patient with my questions about photoacoustic and software development. And of cause thanks for thoroughly proof-reading my thesis.

I thank Thomas Kirchner for his valuable input about physics and photoacoustics and for sharing his expertise about scientific practice: Thanks also for his help when working with the Medical Imaging Interaction Toolkit and for proof reading my thesis.

Without what I learned from Fabian Isensee about deep learning, this thesis would not have been possible, thanks to him.

Many thanks go to the entire MBI department for all the inspiring conversations, all the support and friendship. They have always been kind and helpful.

Last but not least I thank my family and my girlfriend Johanna for supporting me.
This thesis would not have been possible without you.

Thank you!

1 Introduction

Innovations in life and health sciences are often driven by imaging technologies, which connect microscopic discoveries with macroscopic observation [Wang and Yao, 2016] as the recently awarded Nobel prices show. They were awarded for the invention of the super-resolved fluorescence microscopy in 2014 and for the development of cryoelectron microscopy in 2017, which opened research areas for numerous scientists [Hell and Wichmann, 1994, Nobelprize, 2017]. Imaging technologies are key to investigate biological systems with consistent steps of improvement in resolution, imaging speed, contrast, penetration depth and detection sensitivity as well as reducing hazard for the living being of investigation. Especially in medical oncology, innovations in early detection, characterization and monitoring of cancer are driven by imaging techniques. These innovations result in an improved treatment for patients.

Photoacoustic Imaging

For optimal patient care the trend in medical imaging devices is going towards multimodal systems [Cherry, 2009]. These are able to simultaneously deliver diverse information, for example structural and functional information. A combined ultrasound (US) and photoacoustic imaging (PAI) application could be such a multimodal device [Harrison and Zemp, 2011, Talukdar et al., 2013, Merčep et al., 2015]. Advantages of US and PAI in contrast to modalities of similar capabilities are the absence of ionizing radiation, low costs, real-time imaging and portability. These are important factors for clinical application. While US is a fundamental modality in preclinical and clinical practice, PAI is an emerging technique. It is a modality for imaging high optical absorption contrast over a wide range of spatial scales with the possibility of measuring in real time. By the measurement of differences in optical absorption in tissue components such as hemoglobin, lipids, melanin, collagen and water among others, PAI is able to reveal the tissues anatomical, functional, metabolic and histologic properties [Valluru and Willmann, 2016]. Exogenous contrast agents such as dyes, nanoparticles, carbon nanotubes and others can provide molecular and cellular specificity by binding on certain molecules and thereby enhancing optical contrast in PAI [Valluru and Willmann, 2016]. Like US it is a non-invasive method. The spacial scale of PAI can cover organelles, cells, tissues and organs.

In the last decade, photoacoustic tomography (PAT) has been proven to be capable of multiscale imaging with

a consistent contrast mechanism. Thus, it is well suited to bridge the microscopic and macroscopic domains in the life sciences [Wang, 2008]. It is an imaging modality that acoustically detects optical absorption contrast via the photoacoustic effect (PAE), a physical phenomenon that converts absorbed optical energy into acoustic energy [Wang and Yao, 2016]. Biological tissue is orders of magnitude more transparent to sound than to light, when comparing the scattering mean free path. This enables PAI to achieve greater penetration depth than sole optical imaging with a high spatial resolution [Wang and Yao, 2016]. This is especially useful for cancer detection as cancer hallmarks like angiogenesis and oxygenation can be imaged with multispectral PAI. Although the PA phenomena was first reported by Bell in 1880 [Bell, 1880] a century passed before Bowen proposed to use this phenomena for imaging, using ionizing radiation for excitation in 1983 [Bowen, 1983]. Seven years later laser induced excitation was proved to work [Kruger, 1994]. In recent years the development of PAT has accelerated as exemplary shown by [Wang et al., 2003, Razansky et al., 2009, Cox et al., 2012, Wang and Gao, 2014, Xia et al., 2014, Valluru and Willmann, 2016, Arridge et al., 2016, Hauptmann et al., 2017, Wang et al., 2017, Antholzer et al., 2017].

One of the unsolved problems in PAI is the reconstruction of PA raw sensor data for limited view sensor geometries. Linear US transducer in a handheld PA devices, in contrast to 360 degrees scans with PAT, can only provide data from a limited field of view. Therefore established, direct reconstruction algorithms like DAS beamforming [Veen and Buckley, 1988] can only be applied to the acoustic inverse problem with limitations because the reconstruction typically suffers from circular artefacts and systematic loss of contrast with depth [Hauptmann et al., 2017].

In this master thesis the application of novel machine learning algorithms on the PA limited view problem is investigated. Machine learning algorithms are used in many aspects of modern society: In web search [Agichtein et al., 2006], content filtering on social networks [Lops et al., 2011] and autonomous driving [Chen et al., 2015] as well as in medical image analysis [Cruz and Wishart, 2007, Nasrabadi, 2007, Cruz-Roa et al., 2013, Wang et al., 2016, Sun et al., 2016, Shen et al., 2017, Hauptmann et al., 2017, Isensee et al., 2017a, Isensee et al., 2017b, Jaeger et al., 2017, Litjens et al., 2017, Wirkert et al., 2017]. Due to increased computing power, a new class of intelligent algorithms has emerged in recent years: the deep learning algorithms [LeCun et al., 2015]. They have opened up new possibilities for researchers in life sciences because they often outperform classical machine learning algorithms.

The acoustic inverse problem is one major challenge for limited view PA, which must be overcome for quantitative photoacoustics. Utilizing deep learning algorithms could be one key element on the path to clinical application of PAI.

Objectives

Looking at the breakthrough success of machine learning algorithms, it seems to be a logical next step to use deep learning algorithms for tackling the limited view problem in PAT. In this master thesis a novel approach to reconstruct limited view PA raw sensor data using an established deep learning neural network, the so-called U-Net [Ronneberger et al., 2015], is presented and compared to another state-of-the-art technique.

Outline

In order to tackle the PA limited view problem, two approaches have been pursued. The basic principles of photoacoustic imaging and machine learning are explained in chapter 2. The state-of-the-art is summarized in chapter 3, in which among other methods Delay-and-Sum beamforming is explained, to which the developed methods are compared. Using the mcxyz Monte Carlo simulation framework by Steve Jacques [Jacques, 2014] initial pressure distributions of *in silico* tissue models with several blood vessels were simulated for training and testing our reconstruction approach. This is described in chapter 4. These initial pressure distributions were then used to simulate US waves in tissue and to simulate the raw sensor data with k-Wave. This is described in chapter 5. The two approaches are:

1. **Deep learning post processing**: Delay-and-Sum beamforming was used to reconstruct images from the space-and-time domain to the spacial domain. Deep learning was used to post process these images. This is described in section 4.4.

2. **Direct Deep learning reconstruction**: In this approach only one fully convolutional deep learning network was used to reconstruct the initial pressure distribution directly from the PA raw sensor data. In this approach no prior knowledge was used for the reconstruction. This approach is described in section 4.4.

The conducted experiments and the results for both approaches are shown in chapter 5.3 and they are discussed in chapter 6

Some of the results presented in this master thesis have been published beforehand in the SPIE Proceedings, Photons plus Ultrasound [Waibel et al., 2018a] and at the conference 'Bildverarbeitung für die Medizin 2018' in Erlangen [Waibel et al., 2018b].

2 Principles

To explore the possibilities the photoacoustic effect offers for imaging it is essential to understand the optic and acoustic properties of tissue, since the ability of light and sound to penetrate tissue are the basis of photoacoustic imaging. If photons are absorbed in the tissue they can trigger an acoustic wave in the US regime which can then be recorded with a standard US transducer [Wang and Yao, 2016]. Optic and acoustic tissue properties vary depending on several factors such as the wavelength of light and on the acoustical properties, variations in blood content, water content, collagen content and fibre development [Wang and Gao, 2014]. Since US scatters much less than light in tissue PAI is able to provide higher contrast and resolution images than US [Xia et al., 2014] therefore PAI is a desirable technique for medical and biological applications having all the advantages like US, such as not using ionizing radiation and easy transport. It is cheaper than Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET) [Wang, 2008]. And it does not utilize ionizing radiation such as CT and PET imaging. Although optical-imaging modalities can detect optical absorption by measuring the intensity variation in transmitted or reflected light, they can not measure signals from deep regions in tissue due to the diffusion limit for photons [Hordvik, 1977]. As a rule of thumb in PAT, the achievable spatial resolution is on the order of $\approx \frac{1}{200}$ of the desired imaging depth, which can be up to $7cm$ in tissue [Wang and Hu, 2012].

2.1 Historical Abstract of the Photoacoustic effect

> *"I have heard articulate speech by sunlight! I have heard a ray of the sun laugh and cough and sing! ... I have been able to hear a shadow and I have even perceived by ear the passage of a cloud across the sun's disk."*
>
> Alexander G. Bell, in a letter to his father, dated 26 February 1880

Alexander Graham Bell first described the PAE in 1880 [Bell, 1880]. As he interrupted focused sun light with a fast spinning shutter wheel he heard faint sounds from substances hit by those light pulses. He correctly reasoned that these sounds are caused by heating of the material, because the sounds were louder if the light hit dark material. He hoped the "photophone" to be his greatest invention. Although the PA phenomena was first reported by Bell a century passed before Bowen proposed to use this phenomena for imaging with ionizing radiation for

excitation [Bowen, 1983]. In the 1990s laser induced excitation was proved to work [Kruger, 1994] which was another breakthrough for PAI. A short history of PAI is sketched in 2.1

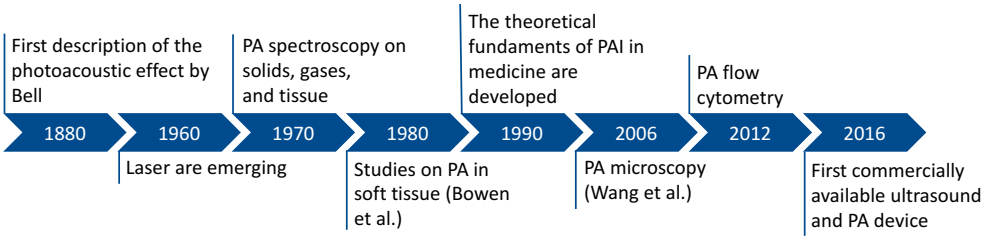


Figure 2.1: A short history of photoacoustic imaging: It took several decades from Bell's first description of the photoacoustic effect to the development of clinical applications.

The Photoacoustic Effect

Due to the impact of thermal energy released by photons when scattered or absorbed in tissue, the tissue thermoelastically expands which results in a pressure change. If the illumination pulse is much shorter than the stress relaxation time τ_s this pressure change takes the form of a elastic shock wave which can be measured with a hydrophone or a piezo-electric US transducers [Ku et al., 2004]. This effect is illustrated in figure 2.2. The photoacoustic effect can be measured after illumination with a short laser pulse. The local fractional volume expansion $\frac{dV}{V}$ can be expressed as:

$$\frac{dV}{V} = -\kappa \cdot p_0 + \beta \cdot T \tag{2.1}$$

where κ is the isothermal compressibility and β is the thermal coefficient of volume expansion, p_0 is the initial pressure change and T the temperature change. The pulse of a laser in photoacoustic imaging is in the order of nanoseconds while the thermal confinement time is in the order of milliseconds for 100μm spatial resolution. Therefore the volume expansion is negligible and the initial acoustic pressure can be written as:

$$p_0 = \frac{\beta \cdot T}{\kappa} = \Gamma(T_0) \cdot \mu_a \cdot \Phi \tag{2.2}$$

Where μ_a is the absorption coefficient of the local absorber and Φ is the optical fluence or irradiance of the absorber, Γ is the temperature dependent Grüneisen parameter:

$$\Gamma_w(T_{\circ C}) = 0.0043 + 0.0053 \cdot T_{\circ C} \tag{2.3}$$

At 37 $^\circ C$ body temperature the Grüneisen parameter is approximately $\Gamma_w(37^\circ C) \approx 0.2$. Photoacoustic measurements are typically induced with a pulsed laser in the infrared light range. The absorption coefficient at around

800nm for blood is approximately $\mu_a = 4cm^{-1}$. Assuming a fluence of I = $\approx 10mJcm^{-2}$ we get an initial pressure change of:

$$\Delta p = 0.2 \cdot 4cm^{-1} \cdot 10mJcm^{-2} = 8kPa \tag{2.4}$$

For estimating the temperature rise in tissue one can approximate $\kappa \approx 5 \times 10^{-10} Pa^{-1}$ and $\beta \approx 4 \times 10^{-4} K^{-1}$ using equation 2.2 on finds that each rise in temperature of about 1mK generates a 800 Pa rise in pressure.

The time evolution of PA wave fields can be modelled using the equations of linear acoustics and it is in the same order of magnitude as in US imaging which makes PAI with conventional US transducers possible [Cox et al., 2012].

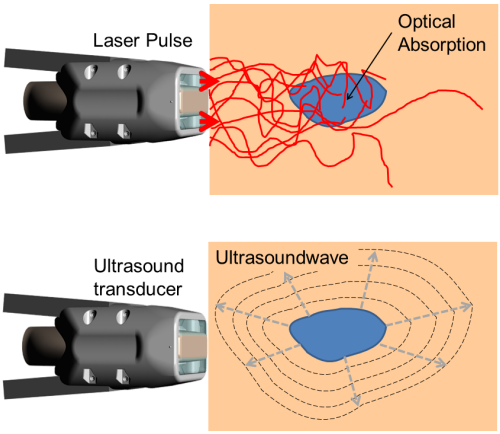


Figure 2.2: The Photoacoustic effect: If a pulsed laser illuminates the tissue, the photons are scattered and absorbed which heats the tissue (top image). The resulting thermoelastic expansion results in a acoustic shock wave in the US regime (image below).

Because the PA conversion is proportional to the optical absorption and individual tissue chromophores absorb light at different wavelengths, they can be selectively imaged with PAI.

2.2 Photoacoustic Imaging

"Photoacoustic imaging is arguably the most exciting biomedical imaging technique of the decade"

Professor Srirang Manohar, University of Twente and Professor Daniel Razansky, Helmholtz Zentrum Munich, 2016 [Manohar and Razansky, 2016]

Making use of the PAE for a medical imaging device promises fine spatial resolution, high sensitivity and good specificity, this has led to a rapid growth of interest in PA in the last decade. With an US transducer, the pressure wave resulting from the initial pressure distribution in tissue can be measured. This could then be used to estimate the absorption in tissue which is the property of interest for researches and physicians as it can indicate tissue properties such as oxygenation saturation. There are several possibilities how the laser source and the transducers can be constructed. Two commonly used geometries are shown in figure 2.3. For a handheld device a linear transducer is used, which can only measure the US signal from a limited point of view, resulting in the limited view problem, this is illustrated in figure 2.3 (a). A more advantageous geometry is shown in figure 2.3 (b) where the transducers surround the tissue and the signal is measured from a 360 degrees scan. With this setup standard algorithms such as a radon transformation or filtered back projection [Kruger, 1994, Webb and Kagadis, 2003, Zhao and Zhang, 2011] can be used, but these are not applicable for PA handheld devices. Photoacoustic tomography offers several advantages as compared to other medical imaging devices. It is faster and less expensive than Magnetic Resonance Imaging. Compared to Optical Tomography, PAT offers a higher imaging depth while sustaining a high spatial resolution within the entire field of view [Xia et al., 2014]. The imaging depth is limited by the distance photons travel in tissue, typically imaging is possible in PAT for depths between 1 and 5 cm [Xia et al., 2014]. While PAT is not limited by the optical diffusion limit, which is at about 1 mm in tissue. It is limited by dissipation to a few centimeters in soft tissue, because the optical fluence is attenuated due to absorption and scattering. The ultrasound attenuation in soft tissue is dominated by absorption instead of scattering. The resolution is limited to about half the acoustic wavelength for US waves, which physically limits the acoustic resolution to about 0.1 mm, not considering aperture. The acoustic attenuation is proportional to acoustic frequency, limiting the penetration depth for higher frequencies. By choosing the transducers frequency, a trade-of between high resolution and high penetration depth is made.

Light in Tissue

Photons are electromagnetic waves which interact with electromagnetic fields of atoms and molecules depending on the photons wavelength and size of molecules and atoms.

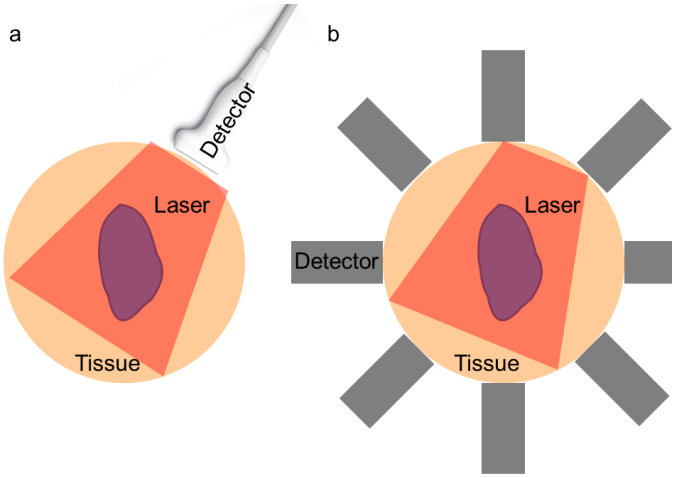


Figure 2.3: Two different ways to construct the laser illumination and US transducers in a PA system. (a) The PA imaging is performed with a conventional US transducer where the signal is measured with a linear transducer leading to the necessity to overcome the acoustic inverse problem with a limited field of view. (b) A ring of transducers is construced around the object of investigation.

In tissue photons experience scattering and absorption caused by interactions with the electromagnetic field of atoms and molecules. Since the Eigenstates of photon energies are quantized, depending on their wavelength, these interactions follow distinct rules described in the Jablonsky diagram shown in figure 2.4. The energy of a photon is:

$$E_{photon} = h \cdot \nu = h \cdot \frac{c}{\lambda} \tag{2.5}$$

Where h is the Planck constant, ν the frequency, c the speed of sound and λ the wavelength of light. Optical properties of tissue are characterized by the absorption coefficient μ_a and the scattering coefficient μ_s.

Scattering of Photons

In tissue, photons can be scattered whereby two types of scattering can be distinguished: elastic scattering where the photons energy is conserved and inelastic scattering where the photon loses or gains energy. Maxwell's equations describe the behaviour of electromagnetic waves. Gustav Mie was able to find an exact solution to these equations for scattering [Mie, 1908]. Elastic Mie scattering is the dominant effect in tissue.

Elastic scattering can be described as the absorption and re-emission of a photon on a structure without loss of photon energy but possible change of direction. For light in tissue scattering in forward direction dominates as the anisotropy factor is close to 1. Accumulating multiple scattering events can randomize the propagation direction of light in tissue, this can be calculated by looking at the anisotropy in scattered light distribution:

$$\mu_s^{'} = \mu_s \cdot (1 - g) \tag{2.6}$$

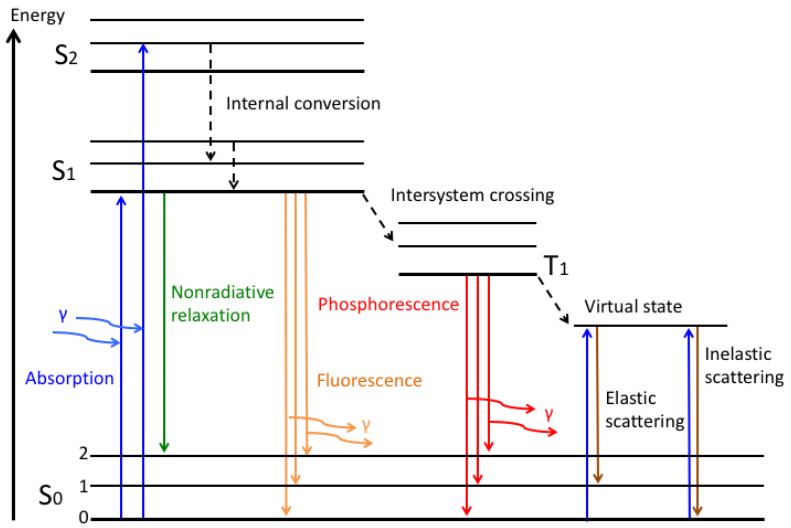


Figure 2.4: Jablonski diagram of possible light tissue interactions. Absorption in a generic molecule, including different relaxation mechanisms which are shown with associated energy levels. The states are arranged vertically by energy and grouped horizontally by spin multiplicity. The ground states of each electronic state are indicated by thick lines, the higher vibrational states with thinner lines.

The anisotropy factor g is defined as the average cosine of the scattering angle after a single scattering event. This means that the closer g is to 1, the more photons are scattered in forward direction. Isotropic scattering is expressed by $g = 0$.

Absorption of Photons

Tissue consists of molecules which interact with the photons optically. Absorbing molecules are called chromophores. Different chromophores have different absorption spectra, which means that the absorption is wavelength dependent. Molecules can store energy as internal energy in different forms, for example asrotational kinetic energy or vibrational energy. This energy is quantized in energy levels which correspond to the eigenstates of the molecule, corresponding to the absorption spectra of a molecule. The sum of all these types of molecular energy is the molecules internal energy. The kinetic internal energy is associated with the thermal motion, which is related to the temperature on a macroscopic level. Because the Eigenenergies of molecules are quantized only photons can be absorbed, which have energy equal to these Eigenenergies. The energy then changes the Eigenzustand of the molecule.

The absorption wavelengths of molecules are not singular wavelengths but broadened due to two broadening effects: Natural broadening is a consequence of Heisenberg's uncertainty principle in quantum mechanics [Heisenberg, 1927]. Collision broadening occurs because collisions with other molecules can abruptly cause the excited state to end, as the energy is transferred to the other molecule. Shortening the lifetime of the excited state in this way gives

rise to a corresponding increase in the spectral line-width [Mukamel, 1982].

As most biological media contains several absorber (for example: blood, lipids, water, melanin), the total absorption coefficient of a material consisting of K chromophores is the sum of the molar absorption coefficients weighted by their concentrations as a function of wavelength can be described by the linear mixture model:

$$\mu_{absorption}(\lambda) = \sum_{n=0}^{K} \alpha_n(\lambda) \cdot C_n \tag{2.7}$$

Where C_n is the concentration of the $n\text{-}th$ chromophore.

When photons are attenuated in matter they increase the internal energy of the absorber. The attenuation is related to the material properties by Lambert-Beers law [Beer, 1852].

The absorption coefficient μ_a inverse characterizes the mean free path l_a in tissue which corresponds to the penetration depth. Both are usually given in units of cm^{-1}. Therefore the probability that a photon is absorbed travelling a distance of dz is μ_a·dz. With Φ being the light intensity it can be written as the differential equation:

$$\frac{d\Phi}{dz} = -\mu_a \cdot \Phi \tag{2.8}$$

with the solution:

$$\Phi = \Phi_0 \cdot exp(-\mu_a \cdot z) \tag{2.9}$$

where Φ_0 is the light intensity on the surface $z = 0$.

The wavelength dependent absorption of major endogenous contrast agents in biological tissue at normal concentrations are shown in figure 2.5. The photons energy can be transferred to the molecule and deposit their energy as

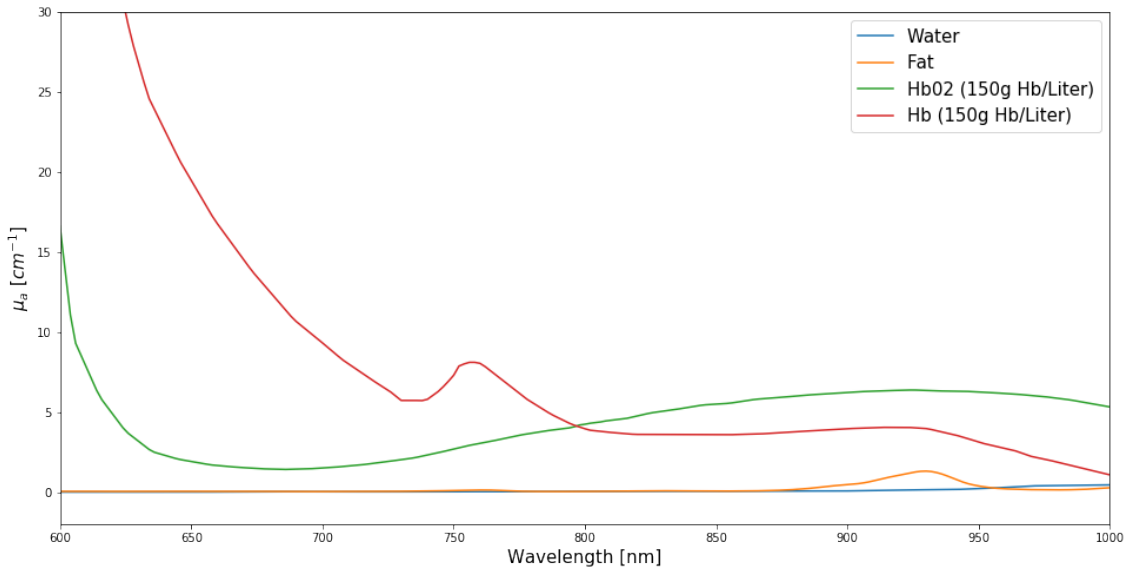


Figure 2.5: Absorption spectra of major endogenous contrast agents in biological tissue at normal concentrations. HbO2 denotes Oxyhemoglobin and Hb deoxyhemoglobin. Data from: https://omlc.org/spectra/

heat. The light source can be described by the fluence rate $\Phi(\vec{x}, t)$, which is dependent on the position x and time t, therefore it is possible to calculate the absorbed power density $H(\vec{x}, t)$, which is the loss in fluence per unit length:

$$H(\vec{x}, t) = -\nabla \cdot \Phi(\vec{x}, t) \tag{2.10}$$

with Lambert-Beers law using the absorption coefficient μ_a can be derived:

$$H(\vec{x}, t) = \mu_a(\vec{x}) \cdot \Phi(\vec{x}, t) \tag{2.11}$$

The energy ΔE deposited in the absorber by a single laser pulse is the integral over time of the whole pulse:

$$\Delta E(\vec{x}) = \int_0^\infty H(\vec{x}, t) \cdot dt = \mu_a(\vec{x}) \cdot \int_0^\infty \Phi(\vec{x}, t) \cdot dt = \mu_a(\vec{x}) \cdot I(\vec{x}) \tag{2.12}$$

$(I\vec{x})$ is the local irradiance in the absorber.

Sound in Tissue

Sound waves are mechanical oscillations in solids, liquids or gases. If the frequency is higher than 20 kHz, the waves are referred to as US waves. Biological tissue such as muscles and fat is often modelled as a liquid in US imaging. This approximation holds true because treating tissue as a fluid is equivalent to ignoring shear waves for which there are good reasons: Shear waves are usually not generated by US transducers and are strongly absorbed by soft tissue. This not true for solids, like bones.

Acoustic waves propagate in tissue by longitudinal motion, not by traverse motion [Millner and Cobet, 1987]. Acoustic waves in the compressible medium tissue are thermodynamic fluctuations in pressure, density, temperature, particle velocity, etc. These can be described by three basic equations:

1. Conservation of mass:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \vec{u}) \tag{2.13}$$

2. Conservation of momentum:

$$\frac{\partial \vec{u}}{\partial t} = -\frac{1}{\rho_0} \cdot \nabla(p) \tag{2.14}$$

3. Pressure density relation:

$$p = c_0^2 \cdot \rho \tag{2.15}$$

where $\vec{u}$ is the acoustic particle velocity, p is the acoustic pressure, ρ is the acoustic density and ρ_0 is the ambient density and c_0 is the speed of sound.

These can be combined to the second order wave equation:

$$(\frac{1}{c_0^2} \cdot \frac{\partial^2}{\partial t^2} - \nabla^2) \cdot p_a = 0 \tag{2.16}$$

Absorption of Acoustic Waves

Acoustic waves travel in a medium with a medium specific speed of sound dependent on a bulk modulus κ and the mediums density ρ, the speed of sound is given by:

$$c = \sqrt{\frac{\kappa}{\rho}} \tag{2.17}$$

As an acoustic wave propagates, it generally looses energy to random thermal motion resulting in acoustic absorption. This absorption can be described if the medium is treated as a sound-absorbing fluid [Bamber and (C. R. Hill, 2004] by:

$$\alpha = \alpha_0 \cdot \omega^y \tag{2.18}$$

The Intensity decreases with deph x as follows:

$$I = I_0 \cdot \exp^{\alpha \cdot x} \tag{2.19}$$

where α is the absorption coefficient, α_0 is the power law prefactor and y is the power law exponent. In the following tabular properties of typical tissue used in US are shown [Bamber and Hill, 1979].

Material	Density (ρ) [kgm^3]	Speed of sound (c) [$\frac{m}{s}$]	Attenuation (α) [$\frac{dB}{cmMHz}$]
Water	1000	1484	0.0022
Air	1.2	330	1.7
Blood	1060	1570	0.18
Fat	920	1450	0.63
Muscle	1070	1570	1.3-3.3

Ultrasound Imaging

Ultrasound waves are referred to sound waves in the range from 20 kilohertz to gigahertz. In US imaging systems, piezo-electric crystals are sensitive to US waves in the range of 2 to 18 megahertz. Either by the shape of the transducer, by a lens in front of the transducer or electronically the US wave is arc-shaped shaped and can be focused on a desired depth. Recent US devices use phased array techniques which enable a change of direction and depth of focus. To generate a 3D image the US beam is swept by sweeping, swinging or rotating the transducer manually or electronically [Carovac et al., 2011, Debus et al., 2012]. In tissue, these waves are reflected or scattered on borders between tissues or structures of different acoustic impedance dependent on the wavelength of the US wave. The absorption and reflection increases with frequency. These waves can be measured with piezo crystals this data is then processed and used to reconstruct an image [Repacholi and Benwell, 1982, Carovac et al., 2011]. Handheld 2D array for real-time 3D imaging have become state-of-the-art in US imaging since in the last decade [Fenster et al., 2001]. Therapeutic US can be sub-divided into firstly anatomical US which refers to imaging organs

or other structures while secondly functional US combines imaging of information such as movement and velocity of tissue or blood with anatomical US. Ultrasound devices can operate in different modes of which some are listed here:

1. **A-mode:** In the amplitude mode the echoes are plotted on a screen as a function of depth.

2. **B-mode:** In the brightness mode (or 2D mode) a linear array of transducers scans a plane through the body which can be viewed as a 2D image on screen.

3. **M-mode:** In the motion mode a rapid sequence of B-mode scans are measured which enables doctors to see an measure motion as the organs move relative to the US probe.

4. **Doppler-mode:** Utilizing the Doppler-effect it is possible to measure if structures are moving, for example blood. This is done by calculating the frequency shift of a particular sample volume.

Therapeutic US can be used to modify or destroy tissue with US waves, for example by moving or pushing tissue, heating tissue, dissolving blood clots or delivering drugs [Carovac et al., 2011].

Possible Clinical Applications of Photoacoustic Imaging

PAI is a technique combining a laser and a US device. Therefore, a hardware unification of a clinical US scanner with an appropriate laser source will essentially result in a combined US and PAI device [Niederhauser et al., 2005, Harrison and Zemp, 2011, Merčep et al., 2015]. Because the PA conversion is proportional to the optical absorption and dependent on the tissue absorption properties light at different wavelengths can be selectively imaged with PAI. Emerging techniques such as US molecular imaging and US elastography which are currently available as add-on features for US scanners have shown promising results in application for early and accurate cancer detection as compared to standard anatomical US.

In-vivo imaging of optical properties with a non-invasive and safe method which combines the advantages of high resolution of US with the large optical contrast of optical tissue properties is a desirable goal of researchers.

There is growing interest in the clinical community on PAI and its possible clinical applications. Advantages are affordability, availability, and portability of US devices. Clinical US is used as a valuable diagnostic tool for resource-limited settings by the ministries of health in low and middle-income countries, several nongovernmental organizations and the World Health Organization [Sippel et al., 2011]. Combining the advantages of US imaging with PAI, both technologies can potentially be adopted widely and more quickly compared to imaging technologies that are costlier. Some promising applications of PAI are listed here:

1. **Monitoring of Cancer**: One of the most promising features of PAI is its ability to characterize tissue components such as hemoglobin, lipids, melanin, collagen and water among many others [Valluru and Willmann, 2016]. In cancer research this is of interest for characterization and monitorization of cancer. One of the hallmarks of cancer [Hanahan and Weinberg, 2011] is neoangiogenesis, which can be detected based on the differences in the absorption of hemoglobin in cancerous tissue compared to the surrounding normal tissue, which could be one of the applications for spectroscopic PAI [Wilson et al., 2014]. Another approach is to use cancer biomarkers and molecular targeted contrast agents for early and accurate cancer detection [Bernsen et al., 2015], these can also be applied with PAI.

2. **Breast Cancer Imaging**: Breast cancer is the most common cancer in women and the second leading cause of cancer death in women. For cure and survival an early diagnose is crucial. Conventional screening methods rely on X-ray mammography and ultrasonography. X-ray mammography's disadvantages are the use of ionizing hazard and insensitivity to tumours in dense breasts, while ultrasonography is insensitive to breast cancer at an early stage owing to low acoustic contrast. There are three representative instruments, the LOIS-64 PAT system, the Philips iU22 system and the Twente PA Mammoscope [Wang and Gao, 2014]. All three systems require patients to lie prone on a modified hospital bed with her breast pendant through the aperture in the bed for imaging.

3. **Sentinel Lymph Node**: As lymph nodes play a central role in metastatic cancer spread they are of large clinical interest. Abnormal node vascularization, morphology and size can be indicators of disease and may be imaged with a PA system. *Ex-vivo* experiments showed that hemoglobin based contrast reveals nodal vasculature and lipid-based contrast reveals the exterior node size, shape and boundary integrity. These two sources of complementary contrast potentially allow indirect observation of cancer, which is one promising application of PAI [Guggenheim et al., 2015].

4. **Assessment of Crohn's Disease Activity**: Because it is possible to quantify hemoglobin-dependent tissue perfusion and oxygenation with non-invasive PAT this modality is suited for monitoring chronic inflammatory diseases [Knieling et al., 2017]. A study was performed on 108 patients with Crohn's disease and compared with endoscopic scoring, histologic scoring as reference tests and conventional US. In the study it is stated that PAT assessment of hemoglobin levels in the intestinal wall can potentially be used to distinguish active disease from remission patients with Crohn's disease without the need for invasive procedures [Knieling et al., 2017].

Quantitative photoacoustic tomography (qPAT) is a technique emerging to clinical appliance [Wang, 2008]. The aim is to determine tissue quantities. Since contrast in PAI is due to chromophores, their concentration is a fundamental quantity to investigate with qPAT. Examples for chromophores which could be investigated are: endogenous

molecules such as oxyhemoglobin or deoxygenated hemoglobin, melanin, lipids and water as well as exogenous contrast agents and absorbing enzymes or other proteins resulting from the expression of reporter genes linked to the expression of a gene of interest [Cox et al., 2012].

Outlook: Other Clinical Applications of the Photoacoustic Effect

The PAE is widely used in other applications. Because this is not the focus of this thesis only two examples are listed here.

1. **Photoacoustic Microscopy**: US microscopes are limited in resolution by the high attenuation of acoustic waves in soft tissue. Since a laser can be focused more tightly than ultrasound PA microscopy can achieve higher resolution [Zhang et al., 2006]. To release the dependence of resolution on acoustic frequency, Optical-Resolution Photoacoustic Microscopy (OR-PAM) utilizes focused light to spatially confine the excitation, resulting in an optical diffraction-limited resolution in the lateral direction [Wang and Gao, 2014]. The excitation laser is focused on an object to excite acoustic waves and the ultrasonic transducer is adopted and aligned confocal with the optical lens. Normally the object is raster scanned in the traverse plane to achieve volumetric imaging. The depth information is encoded by time of flight of the PA wave. Because one laser pulse is send to tissue at each scanning position the imaging speed is limited to the laser pulse rate. Compared to optical microscopy OR-PAM has an advantage in providing high contrast for endogenous chromophores without staining, enabling label-free capability of OR-PAM at different scales [Wang and Gao, 2014]. Acoustic Resolution Photoacoustic Microscopy (AR-PAM) follows a similar principle but the laser is used to illuminate the whole field of view of the acoustic detection volume, in this case the ultrasonic focusing determines the lateral resolution.

2. **PA flow cytometry**: Flow cytometry is a technique which uses suspended cells in as stream of fluid passing through a detector. It is routinely applied in diagnostics of health disorders, especially blood cancers and is used for cell counting, cell sorting, biomarker detection and protein engineering. Commonly flow cytometry is linked to a sorting device to physically separate particles. PA flow cytometry (PAFC) refers to the time-resolved detection of circulating absorbing objects (e.g., tumour cells). Because for PAFC it is not necessary to extract the cells from a living system long-term study of cells in their natural biological environments are possible as compared to conventional flow cytometry. *In vivo* PAFC is using laser pulses on selected vessels with for excitation, followed by the time-resolved measurement of PA signals with an ultrasonic transducer. The blood or lymph vessel function as natural tubes with cell flow. The positive or negative

contrast created by the irradiation of blood vessels can be used: The positive background PAFC can be used to detect higher absorption targets, such as melanoma cells, or cells labelled with nanoparticles, whereas the negative contrast can be used to sense lower absorption targets such as platelets or white blood cells [Wang and Gao, 2014]. One possible application is the detection of circulating tumour cells which play a critical role in evaluating metastasis.

2.3 Machine Learning

"If a typical person can do a mental task with less than one second of thought, we can probably automate it using Artificial intelligence either now or in the near future"

Professor Andrew Ng, Stanford University, 2016 [Andrew Ng, 2016]

Historical Background of Machine Learning

The idea that human abilities can be replicated by machines goes back to the ancient philosophers who developed structured methods of formal deduction in the first millennium BCE. In the 17th century Gottfried Wilhelm Leibniz, Thomas Hobbes and René Descartes investigated if all rational thought could be formulated as systematic as algebra or geometry [McCorduck, 2004]. The foundation of modern mathematics was set by George Boole [Boole, 1854]. Inspired by Bool Bertrand Russel and Alfred North Whitehead presented a formal foundation of mathematics in their Principia Mathematica in 1913 [Bertrand Russel and Alfred North Whitehead, 1913]. Alan Turing developed a formalism with the concepts of algorithms and computation which he applied in the Turing machine, which is considered as the first general computer. He also proposed a 'Learning Machine' in 1950 that could learn and become artificially intelligent. Research in neurology had shown that synapses worked like a network firing electric impulses, Norbert Wiener's cybernetics described control and stability in electrical networks and the information theory by Claude Shannon's described digital signals. Combining these ideas, the construction of an electronic brain was suggested [McCorduck, 2004]. In 1951 Marvin Minsky and Dean Edmonds build the first neural network machine, the Stochastic Neural Analog Reinforcement Calculator called SNARC [Daniel Crevier, 1993]. On the Dartmouth Conference in 1956 Allen Newell and Herbert Simon presented the Logic Theorist, a program with which 38 of the 52 theorems of Russel and Whitehead's Principia Mathematica were proofed or more elegant proofs were found. Until the 80s a golden age of machine learning began, in which many ground-breaking discoveries were made but further developments were not possible due to lacking computing power among other reasons, which lead to the 'first AI Winter'. In 1981 the Japanese government funded the 'fifth Generation project' with the goal to develop

machines which could carry on conversations, translate languages, interpret pictures and reason like human beings, goals which are not achieved today. Some goals have been achieved in recent years however: In 1997 IBMs computer 'Deep Blue' bet the world chess champion, Garry Kasparov and in 2005 a robot from Stanford was able to drive autonomously for 131 miles. In 2011 IBMs Watson was able to win a quiz show against the two greatest quiz champions. In 2016 Google DeepMinds AlphaGo was the first computer to beat professional Go players. There are countless other examples of the success of deep learning approaches in numerous fields [Vinyals et al., 2014, Andrew Ng, 2016, Samek et al., 2017, Andrew Ng, 2016, Chen et al., 2015, Wang et al., 2016, Isensee et al., 2017b, Isensee et al., 2017a, Wirkert et al., 2017].

Basics of Machine Learning

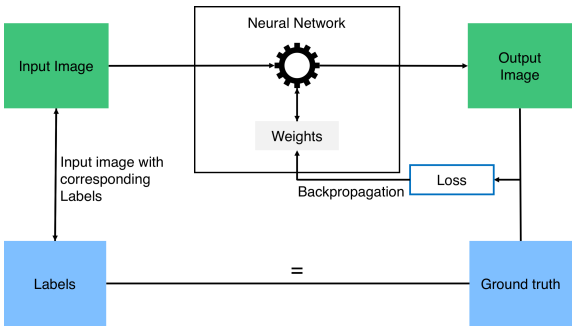
Machine learning is essentially a statistical method that allows a machine to classify patterns or abstract a model from data. They all work similar: First they are trained with data to find underlying structures or hidden models in the data, secondly the model the machine has learned is evaluated on previously unseen data. With enough data of good quality and machine learning algorithms which use several transformations and abstractions, very complex functions can be learned [LeCun et al., 2015].

Typically, two types of learning algorithms can be distinguished:

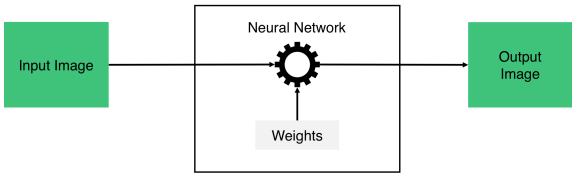
1. **Unsupervised learning** is referred to machine learning algorithms which tries to learn structures from data without corresponding labels.

2. **Supervised learning** is the most common form of machine learning, for supervised learning a large dataset is required with corresponding labels. With this data, the algorithm is trained. Therefore, the algorithm is trying to reproduce the label and a loss function is calculated which measures the error between the machines output and the label, also called ground truth. This error is then used to modify the internal adjustable parameters, often called weights, to reduce this error.

Crucial for a machine learning algorithms performance is how it performs on previously unseen data, this is called generalization. This is evaluated with data which the algorithm has not seen in the training process. The algorithms generalize well if he is able to predict the ground truth on data where he has not seen the ground truth, meaning that the error in the generalization process is similar to the error in the training process. On the opposite, a learning algorithm can perform very well on known data but poorly on previously unseen data, this indicates that the algorithm does not generalize well but only memorizes known solutions without any abstraction, this is called overfitting. Overfitting is a common problem and can be avoided by using more data, data augmentation or by using a different neural network. Sometimes is occurs when the chosen networks architecture is too complex for the problem.



(a) Training the network: A Neuronal network starts with randomly initialized weights and calculates a first output image. This output image is compared to the ground truth using a loss function. The loss is back propagated to update the networks weights. This is done multiple times to optimize the networks performance.



(b) When testing the networks performance no labels are used. The network with the previously trained weights is evaluated on data not seen by the network before. The weights determine what the network predicts.

Figure 2.6: The training (a) of a supervised learning algorithm and its evaluation (b).

Training of the Algorithm

One possible visualization of the machine learning algorithms function is a hilly landscape in high dimensional space as shown in 2.10. The optimal solution is the minima of this landscape. The gradient vector indicates the direction of the steepest decent as shown in figure 2.11, following this gradient vector leads to the minimum where the loss is minimal and the solution is to the problem is optimal.

In the first layer's features of the of the neural network typically represent the edges in an image, these edges with their orientation or position in an image. The second layer typically detects motifs by spotting particular arrangements of edges. The third layer can assemble motifs into larger combinations which then correspond to parts of familiar objects, and subsequent layers would detect objects as combinations of these parts. The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure [LeCun et al., 2015].

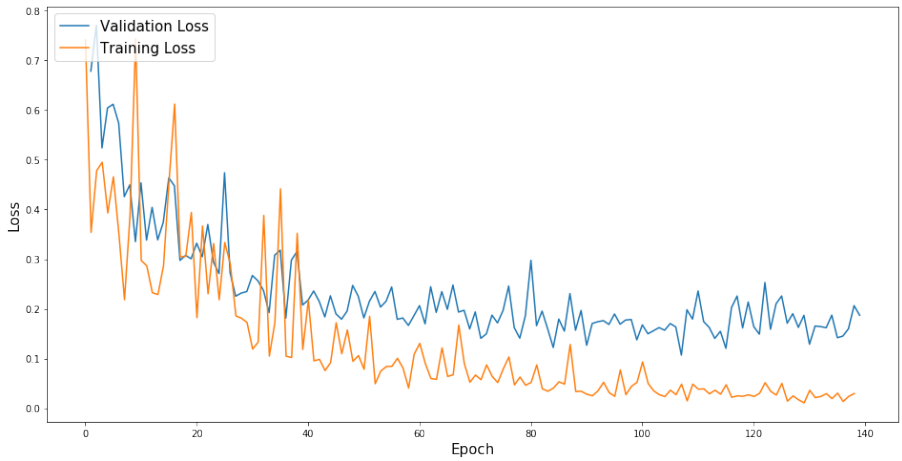
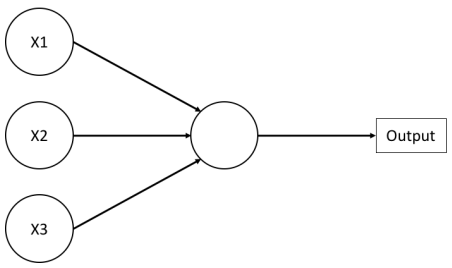


Figure 2.7: The training and validation losses are plotted over the epochs for the training of a U-Net. The training loss is calculated while training the algorithm and the validation loss during validation. One expects the validation loss to be generally higher than the training loss because the algorithm optimizes itself on the training images. The training and validation loss decreases over the epochs. An early split in the curves of the training and validation loss can indicate overfitting.

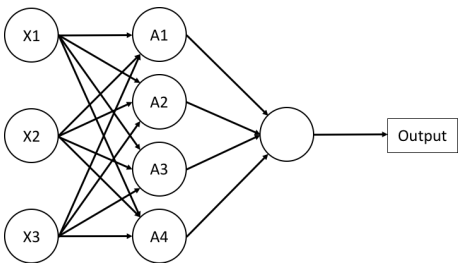
Sigmoid Neurons

A neuron is a computational unit that takes an input and (X_i) and gives an output (f(z)) as shown in figure 2.8. Neurons can be seen as an activation function which can be On and Off, this then is described by a heavy side function. More advanced neuron, such as the sigmoid neuron can take any number between 0 and 1, they are described by the sigmoid function visualized in figure 2.9. They can take multiple inputs (x_i) and have a weight for each input (w_i) and an overall bias b. The output of a sigmoid neuron is defined as:

$$\sigma(z) = \frac{1}{1 + exp(-\sum_j w_j \cdot x_j - b)} \coloneqq \frac{1}{1 + e^{-z}} \tag{2.20}$$



(a) The sigmoid neuron has the inputs (X_i), these can take any values between 0 and 1. The neuron has weights (w_i) for each input indicated with the arrows.



(b) Several sigmoid neurons form a neural network. This one has three input layers (X_i), four hidden layers (A_i) and one output layer.

Figure 2.8: The neuron is the basic structure in a machine learning algorithm. It can be understood as an activation function for artificial neurons. Several parallel neurons form a layer and several layers form a neural network.

Using the sigmoid function, it is possible to vary the neuron by small changes of the input. This makes it possible

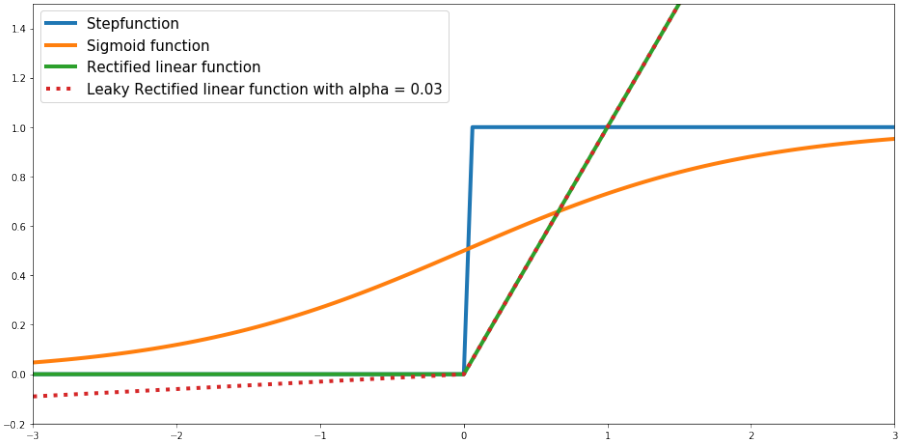


Figure 2.9: Visualization of a heavy side function, a sigmoid function, a ReLu function and a Leaky-ReLu function. The heavy side and sigmoid functions can take any number as input and output a number between 0 and 1. Sigmoid functions, ReLu functions and Leaky-ReLu functions are used in state-of-the-art machine learning algorithms. The exact formula is described in 2.20 and 2.21 and 2.22.

to gradually change the networks behaviour. For deep learning, recent research [LeCun et al., 2015] has shown that rectified linear functions often work better in practice than sigmoid functions for deep neural networks, they can take numbers from a larger range than sigmoid neurons. The basic rectified linear function is described by:

$$\sigma(z) = max(0, x) \tag{2.21}$$

A development of the ReLu is the Leaky ReLu (LReLu) function [Andrew L Maas et al., 2013]. A small leak of size α is introduced, which improves the neurons performance.

$$\sigma(z) = max(0, x) + \alpha \cdot min(0, x) \tag{2.22}$$

Cost Function

The goal in machine learning is to find the best parameters with which a mathematical problem can be represented. In order to quantify how good these parameters are the cost function which describes the loss is minimized:

$$Minimize : C = \frac{1}{2 \cdot m} \sum_{i=1}^{m} (P(i) - GT(i))^2 \tag{2.23}$$

Where P is the algorithms prediction, GT the ground truth and i denotes an image.

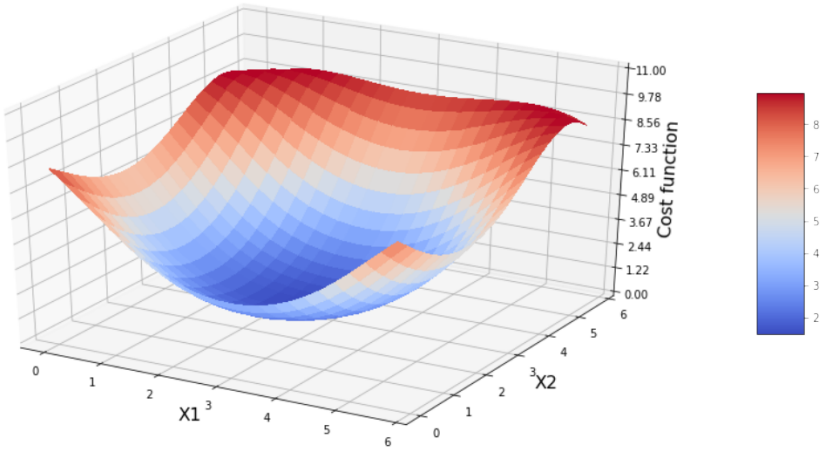


Figure 2.10: The Cost function describing the loss can be imaged as a minimum problem, here displayed with two variables (X1 and X2). In practice, it is a high dimensional problem which cannot easily be visualized.

Gradient Descent

The Cost functions minimum is searched and an easy way to find a minimum is using gradient descent. Therefore, the Cost function has to be differentiable. In order to adjust the weights which are calculated by the gradient descent the learning algorithm computes a gradient vector that, for each weight, indicates by what amount the error would increase or decrease if the weight were increased or decreased by a tiny amount. By updating the weights step-by-step the Cost functions minimum is approached.

$$\theta_{i,j} = \theta_{i-1,j} - \alpha \cdot \frac{\partial}{\partial \theta j} \cdot C(\theta_0, ..., \theta m) \tag{2.24}$$

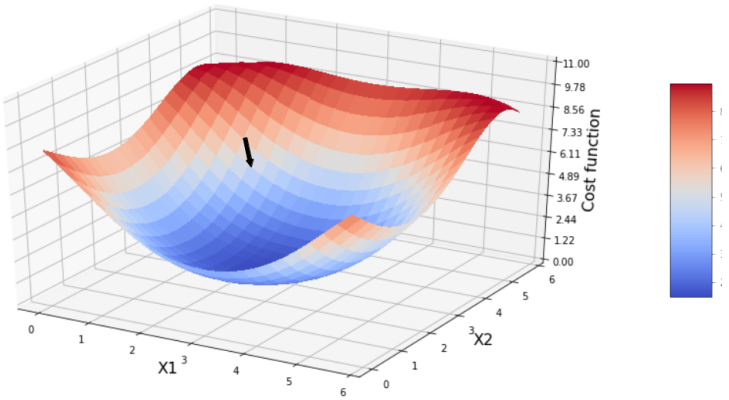


Figure 2.11: A visualization of the Gradient descent: The Gradient at one point of the cost function points in direction of the steepest descent, which is indicated with the black arrow. Like a ball in a sink rolling down to the lowest point, gradient descent is trying to find the global minima of the cost function. The minima of the cost function is the lowest possible loss and therefore the best solution to the task.

A learning rate is introduced to make the algorithm work more efficient. This learning rate is multiplied to cost function and thereby determines the step size for each iteration. If the learning rate is chosen too low the algorithms will take a long time to converge to the minima, if it is chosen to large there is a possibility to overshoot the minima. In state-of-the-art algorithms, a flexible learning rate is used which adapts [Jacobs, 1988]. A lot has been done to optimize gradient descent algorithms in recent years [Ruder, 2016]. One state-of-the-art algorithm called Adam [Kingma and Ba, 2014] which is based on adaptive estimates of lower order moments, it is computationally efficient and can deal with large datasets [Kingma and Ba, 2014]. The step size in which the gradient is descending can be calculated as:

$$\Delta C = \nabla C \cdot \Delta x \tag{2.25}$$

Where the learning rate η defines the stepsize for each iteration.

$$\Delta x = -\eta \cdot \Delta C \tag{2.26}$$

Iterating this will give a minimum of the Costfunction. This results in the update rules for the learning algorithms, which updates the weights and biases:

$$w_k \rightarrow w_k^{'} = w_k - \eta \cdot \frac{\partial C}{\partial w_k} \tag{2.27}$$

$$b_l \rightarrow b_l^{'} = b_l - \eta \cdot \frac{\partial C}{\partial b_l} \tag{2.28}$$

By repeatedly applying this update rule the optimal weights and biases can be found, which means that the Costfunction is minimized.

Backpropagation

Backpropagation is a method used to calculate how much each neuron has contributed to the error after a batch of data was processed. With backpropagation, the weights of neurons can be adjusted after calculating the loss function and by propagating the error backwards. The backpropagation algorithm was originally introduced in the 1970s, but its importance wasn't fully appreciated until 1986 [Rumelhart et al., 1986]. In this paper, several neural networks using backpropagation are described. They are faster than earlier approaches, making it possible to use neural networks to solve problems which had previously been unsolvable. Today, the backpropagation algorithm is the most used optimization approach in neural networks [LeCun et al., 2015]. During backpropagation, the networks weights are adapted and thereby the network learns the best parameters [LeCun et al., 2015]. Recent advances by intelligent algorithms applied on data show a great potential for clinical appliance. Especially deep learning is a rapidly emerging to a state-of-the-art technique in medical information [Hauptmann et al., 2017].

Deep Learning

Deep learning is a method in the family of machine learning algorithms, which has evolved in recent years. Enabled by increasing computational power larger neural networks can now be used [LeCun et al., 2015]. Deep learning algorithms combine multiple layers in a network for feature extraction and transformation, the more layers such a network has, the deeper it is said to be. Like neuronal networks deep neuronal networks have an input and an output layer but deep neuronal networks have several hidden layers in between. Deep learning algorithms are characterized by using a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the previous layers output as input. Different levels of abstraction are used by Deep learning algorithms to abstract models from data.

Layers in a Neural Networks

A convolutional neural network consists of an input, output and several hidden layers. Layers which have been used in this implementation of the U-Net are listed below:

1. **Convolutional layers** are inspired by biological processes, the connectivity pattern between neurons inspired by the organization of the visual cortex. Convolutional layers apply a convolution operation to the input.

2. **Batch norm layer** performs a normalization of the whole image batch. The mean and standard deviation are calculated over the batch, while γ and β are learnable parameters. The formula is given by:

$$Y = \frac{x - mean(x)}{\sqrt{var(x) + \epsilon}} \cdot \gamma + \beta \tag{2.29}$$

3. **Rectified linear units (ReLu)** are activation functions. Leaky ReLus allow a small, non-zero gradient when the unit is not active as described in section 2.3.

4. **Pooling layers** combine the outputs of neuron clusters at one layer to a single neuron in the next layer. Max Pooling for example uses the maximum value from each cluster of neurons at a prior layer.

5. **Upsampling layers** performs an upsampling using nearest neighbour, linear, bi-linear and tri-linear interpolation.

6. **Fully connected layers** connect every neuron in one layer to every neuron in another layer.

Improving the Networks Performance

1. **Batch computing** is used to increase computational performance. A set of data is consolidated to a batch which improves computational performance since most libraries are optimized for array computing.

2. **Data Augmentation** is used to increase the data on which the algorithms is learning. An increase in data on which the algorithms learns leads to an increase in prediction accuracy. Therefore, data augmentation can improve an algorithms performance [Perez and Wang, 2017]. Due to the special geometry of PA images the only reasonable data augmentation for PA US data is mirroring along the z-axis, which is doubling the dataset size on which the algorithm learns.

3. **GPU Computing** is a technique to increase computational speed by using the graphics processing unit (GPU), which typically only handles computations only for computer graphics but can be used to compute tasks normally executed by the central processing unit (CPU). Because GPU's normally have more cores than CPU's they and can process far more graphical data per second than CPU's can. Therefore, transferring the data to the GPU and processing it there can lead to a significant speedup.

The U-Net

The U-Net is a deep neural network and a state-of-the-art deep learning method. It consists of the same number of contracting and expansive layers. Where the number of feature channels is the same in the first and last layer, the number of feature channels is doubled in the second and second-last layer and so forth while halving the resolution in each contraction step while doubling it in each expansion step. All layers have a large number of feature channels which allows the network to propagate context information to higher resolution layers. This symmetry makes the network look symmetrical and yields the U-shape architecture. Contracting layer consist of unpadded convolutions followed by a rectified linear unit and a pooling operation for down sampling. At each down sampling step the number of feature channels is doubled [Ronneberger et al., 2015]. The expansive layers consist of up sampling feature map followed by an up-convolution which halves the number of feature channels, each is followed by a rectified linear unit. The architecture of the used U-Net is shown in figure 2.12 as presented in the U-Nets first publication.

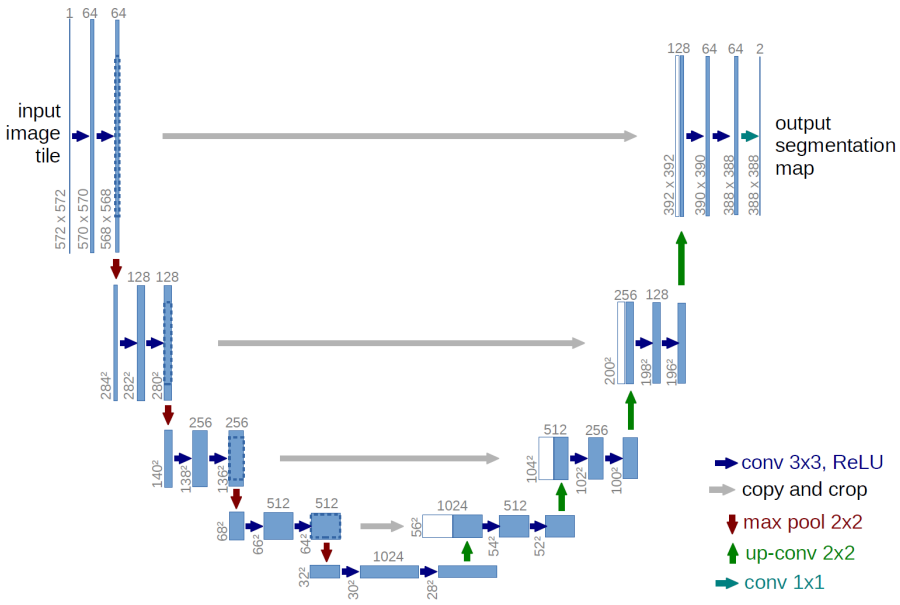


Figure 2.12: Architecture of the U-Net as published by Ronneberger [Ronneberger et al., 2015]. The blue boxes correspond to multi-channel feature maps. Their number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations [Ronneberger et al., 2015] [reprinted with permission].

3 State-of-the-Art

Due to the similarities in PA and US image reconstruction, algorithms from US imaging are applied in PA imaging with some modifications [Wei et al., 2015]. In clinical practice the established, technical state-of-the-art method for reconstruction of PA images from raw sensor data is Delay-and-Sum (DAS) beamforming and advancements of it [Thomenius, 1996]. For a given initial pressure distribution the acoustic wave travel at the speed of sound in tissue and reach the sensor at different sensor locations depending on their origin. This leads to delays in time and phase which correspond to the distance the wave has travelled from their origin. In DAS beamforming the signals, measured at different locations of the linear sensor, are delayed according to their position and then summed. This results in a beamformed image. DAS beamforming has the purpose to yield the origins of the measured sound waves [Mozaffarzadeh et al., 2017]. It is a simple and fast algorithm but not able to provide quantitative estimations of the initial pressure distribution [Veen and Buckley, 1988]. Caused by the geometry of the photoacoustic US probe and the arising limited view problem, they give rise to artefacts [Thomenius, 1996] [Treeby and Cox, 2010]. One advancement of DAS beamforming is Delay-Multiply-and-Sum (DMAS) beamforming [Matrone et al., 2015] and Double-Stage Delay-Multiply-and-Sum (DS-DMAS) beamforming [Mozaffarzadeh et al., 2018].

From corresponding elements of an array DAS and DMAS beamformers calculate delays and sums. DMAS was first used in confocal microwave imaging for breast cancer detection [Lim et al., 2008] and was then applied to PAI. The DMAS algorithms calculates combinatorically samples which are coupled and multiplied before summation. This algorithm is able to enhance the PA images in terms of side lobe levels and resolution [Park et al., 2016]. DS-DMAS is a recently introduced algorithm which is a modified version of the DMAS beamformer by using the summation of the DAS terms which can lead to noise reduction and artefact suppression [Mozaffarzadeh et al., 2018]. These established reconstruction algorithms which are based on DAS beamforming are not optimal to obtain high quality reconstructions as they can give rise to artefacts or adverse noise amplification [Thomenius, 1996, Treeby and Cox, 2010, Hauptmann et al., 2017]. Advances in these algorithms have been made [Huang et al., 2013, Arridge et al., 2016] but a drawback of these is their high computational complexity and the difficulty to handcraft prior constraints that capture the spatial structure of the target accurately [Hauptmann et al., 2017]. All beamforming algorithms assume ideal point-like US transducers with a large acceptance angle and an infinite temporal-frequency bandwidth, which is in in practice not achievable [Xia et al., 2014], it was found that the bandwidth limits both axial

and lateral resolution, while the detector aperture mainly affects the lateral resolution. The transducer response acts as an additional filter to the original pressure [Wang et al., 2003].

Another approach to limited view PA reconstruction is time reversal. Time reversal is a method where a numerical model of the forward problem is simulated backwards in time and delivers approximations of the initial pressure distribution [Treeby and Cox, 2010]. This can be done by re-transmitting the measured PA raw sensor data in to the domain in time-reversed order [Treeby and Cox, 2010]. This is done by solving the wave equation time-reversed using the measured data as a boundary condition. This approach is able to compensate the acoustic absorption and dispersion on the measured time domain signals by reversing the absorption coefficient and utilizing an absorbing equation of state and regularizing it by filtering the absorption and dispersion terms in the spatial frequency domain using a Turkey window. This approach was validated by Treeby and Cox et al. on a carbon fibre phantom and the vasculature in the abdomen of a mouse. This algorithm is integrated in the k-Wave toolkit which is very versatile: It is possible to tune all relevant tissue parameters, simulate different sensor geometries and experimental setups. One disadvantage is the computational intensity [Xia et al., 2014]. Like for the beamforming approaches time reversal assumes an idealized point-like sensor with a large acceptance angle which is not achievable.

While the vast majority of other methods in the field is model-based, first attempts have been made to apply deep learning approaches - which have yielded breakthrough success in a number of medical image processing challenges [Isensee et al., 2017b, Shen et al., 2017]. Hauptmann et al. have presented a novel approach using a deep learning neural network trained layer-by-layer to reconstruct 3D photoacoustic images with high resolution [Hauptmann et al., 2017]. Their network incorporates the physical model into the reconstruction procedure to iteratively reduce artefacts. In on of their approaches they have used a U-Net to post process data from a direct reconstruction but have come to the result that this approach is limited by the quality of the initial reconstruction [Hauptmann et al., 2017]. It has been tested on *in silico* data and on 1200 volumes of vessels from segmented CT of the human palm. Each volume had a size of $40 \times 120 \times 120$ pixels. They come to the conclusion that in limited-view PA reconstruction it is essential to incorporate the physical model to reduce artefacts. They state that incorporating the gradient of the data fit and learning an iterative algorithm consisting of several convolutional neural networks improves reconstruction performance and can lead to a speed up in computational time [Hauptmann et al., 2017]. As they have applied this approach at a human wrist they were not satisfied with their results and state that their network had not learned to threshold noise-like artefacts in the low absorption regions. They want to tackle this problem by adding Gaussian noise to the training data. They also want to acquire images of the human hand, wrist and palm to train a transfer learning function which improved the results.

Reiter et al. have used a convolutional network with feed-forward fully connected layers to identify locations of point sources such as brachytherapy seeds in 2 dimensions [Reiter and Bell, 2017]. They used pre-beamformed data for their learning algorithm which they simulated with k-Wave. They used various medium speed of sounds

(1440 - 1640 $\frac{m}{s}$) and target locations between 5 and 10 mm in tissue and point sources of size 1 to 5 mm. On 2412 randomly selected images the mean axial error was 0.28 mm and the lateral error was 0.37 mm which is the average imaging systems resolution, according to their paper. They were also able to identify the location of an artefact in the simulated image with a axial error of 2.1 mm and an lateral error of 3.0 mm. In future, they aim at detecting multiple point sources, partial wave fields and identification of artefacts for removal using convolutional neural networks. They hope to replace traditional beamforming algorithms for identification of point sources in future which could also lead to an elimination of reflection artefacts [Reiter and Bell, 2017].

Antholzer et al. have developed a deep learning algorithm based on the U-Net for a circular measurement geometry in 2 dimensions. It uses a linear standard reconstruction algorithm, the filtered back projection, before applying the deep learning algorithm on sparse sensor data to remove artefacts. They used two datasets with each 1000 images of size 128 $\times$ 128 pixels, one without noise and one with random noise. They simulated 1 to 5 ellipses in each initial pressure images. Then they used k-Wave to simulate the PA raw sensor data with 30 detector positions. In their paper, they state that their approach performs equally good on state-of-the-art iterative approaches for PAT from sparse data but in less computation time. In future, they want to test their approach in real-world scenarios including realistic phantoms and different measurement geometries [Antholzer et al., 2017].

4 Methods

To evaluate the feasibility of this deep learning approach to the acoustic inverse problem for handheld PAI, it was decided to work with *in silico* data for several reasons: Firstly this gives the possibility to simulate the ground truth and the PA raw sensor data, which is essential for training the deep learning networks. Using *in vitro* data would have meant to measure the phantom with a CT scanner or a 3 dimensional ultrasound. This approach would have been costlier and it would have taken months to create a large dataset for deep learning. The optical forward simulation which created the ground truth was done with the mcxyz toolkit [Jacques, 2014] implemented in the Medical Imaging Interaction Toolkit (MITK) [Wolf et al., 2005]. The acoustical forward simulation which created the PA raw sensor data was done with the k-Wave toolkit in Matlab [Treeby and Cox, 2010]. For reconstruction the DAS beamforming algorithm in MITK was used. The deep learning approaches for reconstruction were implemented in python. In total 3600 images with multiple vessels and 2000 images with singlevessels were simulated. The steps we have taken in this approach are sketched in figure 4.1.

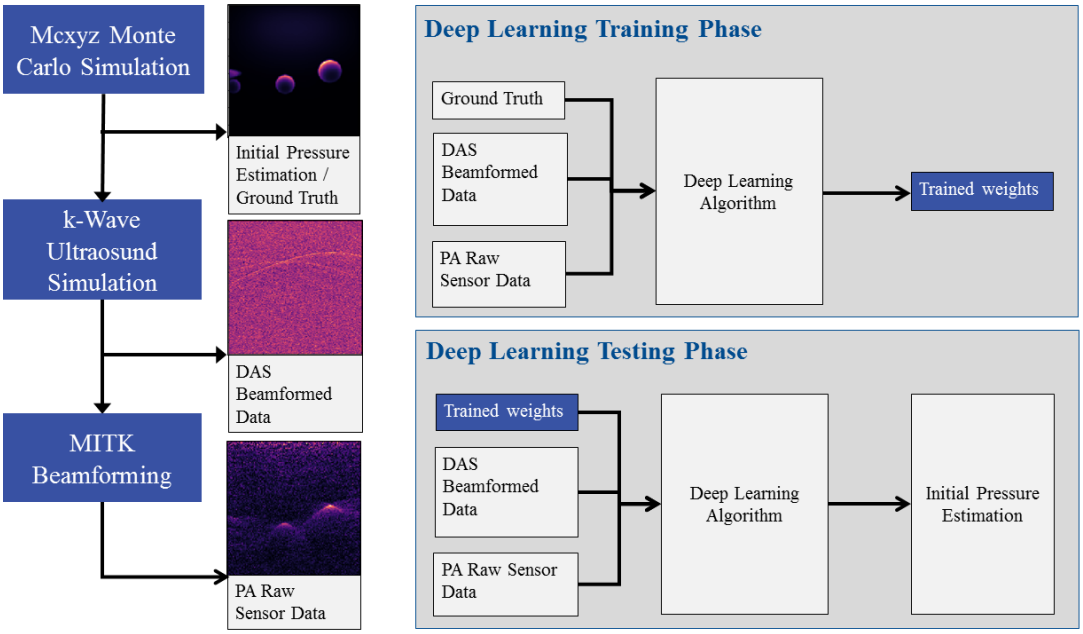


Figure 4.1: The design of this experiment's pipeline: First the initial pressure images of single and multiple blood vessels were generated with mcxyz Monte Carlo simulations (top left). These images were used to simulate pressure waves in tissue with k-Wave (middle left). This tool also allows to simulate a transducer which can measure the simulated pressure wave using a simulated piezoelectric sensor. This gave the PA raw sensor data, which is the data an US transducer measures. The third step was to beamform the sensor data, this was done with the DAS beamforming plugin in Medical Imaging Interaction Toolkit (bottom left). In the next step this data was used to train the deep learning neural network, for this step the ground truth images are needed (top right). The state of the network was saved in the networks weights. For testing the networks performance, the weights were used to predict the initial pressure distribution from first the PA raw sensor data and secondly the DAS beamformed images (bottom right).

4.1 Optical Forward Simulation

Images of initial pressure distributions were simulated given a specific distribution of optical parameters using an adaptation of the widely used Monte Carlo simulation framework mcxyz [Jacques, 2014]. The simulation pipeline is integrated in the MITK [Wolf et al., 2005]. With this toolkit, it is possible to simulate photon fluence dependent on optical absorption properties of tissue in 3 dimensional volumes. A context image was introduced, in which the relevant information of the measured signal voxel with characteristics of the imaging system are combined in one feature vector. Then a machine learning algorithm was trained to estimate the fluence in every voxel based on this feature vector. With this method the fluence corresponding to an absorption was simulated [Kirchner et al., 2017]. As in this master thesis 2 dimensional images were used slices from these 3 dimensional volumes were taken out. Multiplying the fluence and absorption one gets the initial pressure distribution.

4.2 Acoustical Forward Simulation

The k-Wave toolbox is a simulation framework for the time domain simulation of acoustic wave fields in Matlab. It is an open source toolbox which can be used to simulate time domain acoustic waves in tissue [Mast et al., 2001]. This toolbox was used to simulate the sensor data, which is the data the sensor measures from an initial pressure distribution in the media, resulting in simulated PA raw sensor data where one axis is corresponding to the number of transducers and the other to the time resolved pressure measured by the transducer. k-Wave solves the system of coupled acoustic equations described in the previous sections using the k-space pseudo-spectral method. This combines the spectral calculation of spatial derivatives (in this case using the Fourier collocation method) with a temporal propagator expressed in the spatial frequency domain or k-space [Bojarski, 1982, Mast et al., 2001].

Photoacoustic Wave Propagation

The k-Wave simulation of the propagation of acoustic waves in tissue is governed by these equations: The Wave equation 2.16, the absorption equation 2.7, the conservation of mass 2.13, the conservation of momentum 2.14 and the pressure density relation 2.15. These equations are numerically optimized for fast simulations in k-Wave.

Pseudo Spectral and k-Space Methods

For solving partial differential equations in acoustics, the finite-difference (FD), finite-element, and boundary-element methods are most commonly used. Because of the high number of calculation steps for high frequencies this approach is too slow for modelling ultrasonic waves with k-Wave. For calculating a FD scheme, the gradient of the field is

linearly interpolated between its values at the grid points. A better way to estimate the gradient could be to fit a higher-order polynomial to a greater number of nodes and then calculating the derivative of the polynomial. The higher the degree of polynomial, the more points can be used, which can improve the accuracy estimation of the derivative. The Pseudo spectral (PS) method takes this further and fits a Fourier series to all of the data which is efficient when using the fast Fourier transformation [Treeby and Cox, 2010]. While the PS method can improve efficiency, conventional FD schemes are used to calculate the gradients in the time domain. The FD approximation introduces instability into the numerical simulation that can only be controlled by limiting the size of the time-step, which is done by the Courant-Friedrichs-Lewy (CFL) number 4.4 [Treeby and Cox, 2010].

Perfectly Matched Layer

The simulating of propagating wave fields requires an efficient numerical scheme to compute the derivatives near the grid boundaries. For PS or k-space methods, using the FFT to compute the spatial derivatives causes waves leaving one side of the domain to reappear at the opposite side. Implementing a perfectly matched layer (PML) can reduce this [Treeby and Cox, 2010].

Including the PML the first-order acoustic equations given in 2.2 become:

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0} \cdot p - \alpha \cdot u \tag{4.1}$$

$$\frac{\partial p_x}{\partial t} = -\rho_0 \cdot \frac{\partial u_x}{\partial x} - \alpha_x \cdot \rho_x \tag{4.2}$$

$$p = c_0^2 \cdot \sum \rho_{x,y,z} \tag{4.3}$$

Staggered Grids

The numerical solution of equation 4.3 is computed in iterations. First the pressure distribution within the computational domain is used to calculate the spatial deviates in each direction $(\partial x, \partial y, \partial z)$. These are used to compute the corresponding velocity values using the first-order FD. In the next step the spatial derivatives of the velocity for each Cartesian direction are computed. These are used to update the values of the acoustic density within the domain, using a first-order FD. In the last step, the pressure is calculated using the appropriate equation of state.

Parameters used in k-Wave

kGrid: The kGrid is an grid on which the individual wave fiels per time step are calculated at each grid point. It's size is naturally chosen to have the same dimensions as the image. The number of grid points therefore corresponds to the number of pixels in the image.

Absorption: The Absorption is imported into Matlab as a .nrrd file

Fluence: The Fluence is imported into Matlab as a .nrrd file

Initial Pressure: The initial pressure is the absorption multiplied with the fluence.

CFL Number: A useful number when discussing stability of the simulation is the Courant-Friedrichs-Lewy (CFL) number, which is defined as the ratio of the distance a wave can travel in one time step to the grid spacing [Courant et al., 1928], it is defined as:

$$CFL = \frac{c_0 \cdot \partial t}{\partial x} \tag{4.4}$$

The CFL number can be thought of as a non-denationalized time step, and for that reason it is useful for defining the maximum permissible time step without reference to a specific grid spacing, in this case 0.3.

Time steps: In order to guarantee stability of the simulation the CFL number is used to calculate the time steps:

$$\partial t = \frac{CFL \cdot \partial x}{c_{max}} \tag{4.5}$$

4.3 Beamforming Approach

For reconstructing PA raw sensor data several algorithms have been published as mentioned in the state-of-the-art section 3. Most medical US devices on the market use a Delay-and-Sum (DAS) beamforming [Thomenius, 1996] algorithms to beamform B-mode images. Caused by the lack of complete information which is due to the limit in aperture size and operating frequency these algorithms are limited in resolution and contrast [Matrone et al., 2015]. DAS beamforming was used to process the PA raw sensor data for the deep learning post processing approach and as the state-of-the-art reconstruction method.

4.4 Deep Learning Approach

The huge success of deep learning methods in medical image processing described in previous chapters made it feasible to apply deep learning to the acoustic inverse problem. Since the task of removing artefacts in images

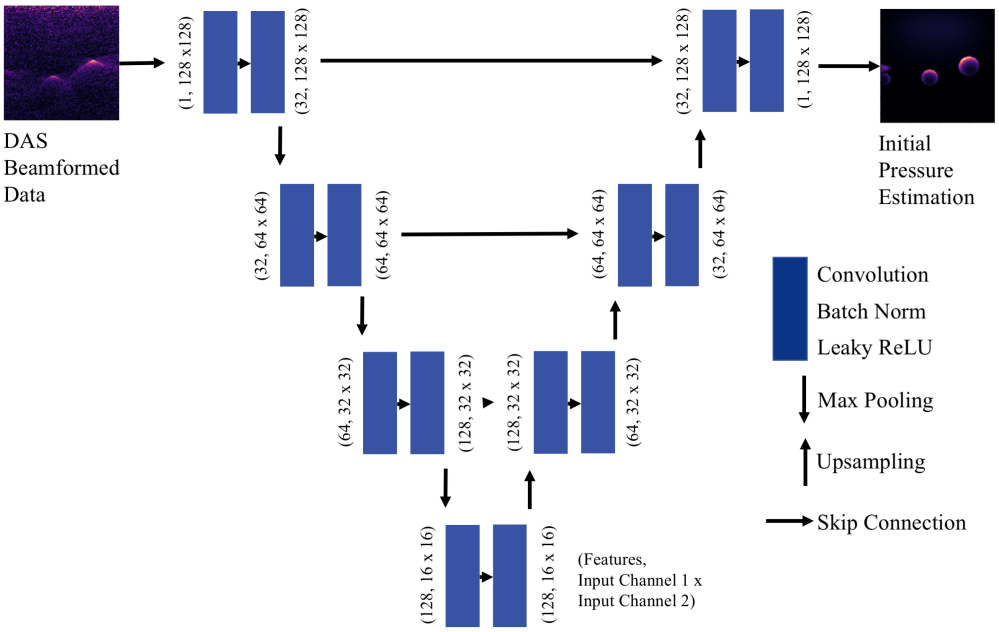


Figure 4.2: Architecture of the U-Net used for post processing the DAS Beamformed data. This U-Net consists of four layers. The contracting steps consist of the repeated application of two convolutional layers with a kernel size of 3x3 and stride of 1x1, each followed by a Leaky ReLu layer and a batch normalization layer.

requires simple processing steps and a general concept can be abstracted from the given data this seemed to be a reasonable task for a deep learning algorithm. If the direct reconstruction from the PA raw sensor data was equally learnable was an open question. The deep learning network was implemented in Python with the PyTorch [Pytorch, 2018] toolbox which offers operations with a GPU acceleration using NVIDIA-CUDA [Cuda, 2013]. From the PyTorch toolbox the tensor computation and the Adam optimizer has been used. For the neural network the U-Net was chosen, which was published by Ronneberger et al. in 2015. It is a fully convolutional neural network. It has the strength to work with few training data and is a state-of-the-art deep learning network [Ronneberger et al., 2015].

Deep Learning Post Processing

Deep learning was used for post processing previously DAS beaformed PA raw sensor data. The input images had a size of 128 × 128 pixels, the output images had the same size. Therefore, a standard U-Net was used. The U-Nets architecture is shown in figure 4.2.

In this implementation, the U-Net consists of four layers. The contracting steps consist of the repeated application of two convolutions with kernel size 3x3 and stride 1x1, each followed by a leaky rectified linear unit (LReLu) and a batch normalization. The down sampling is performed by a 2x2 max pooling operation. The expansive units consist

of an up-sampling followed by the same convolutions, ReLus and batch normalization steps as in the contracting units. Information from the contracting units is concatenated via the skip connections to the corresponding expanding units. The skip connections were introduced to maintain the high frequency information, which are otherwise lost due to the contractions. For optimization PyTorch's implementation of Adam with a base learning rate of 10^{-4} and a weight decay of 10^{-5} as well as a L1 loss function for back propagation was chosen.

Deep Learning Reconstruction

A direct reconstruction approach of PA raw sensor data was performed with a modified U-Net architecture. In this case, the PA raw sensor data was not pre-processed with DAS beamforming but the PA raw sensor data was directly used. The major challenge was to transform the data from a space and a time domain to a 2-dimensional space domain. In the chosen implementation, the modified U-Net consists of four layers. The contracting and expanding units in the architecture is the same as described in the subsection above 4.4 and shown in figure 4.2 but the skip connections were modified by introducing a layer in each skip connection, which reduces the image dimensions in one direction by using an anisotropic kernel size of 20×3 and striding of 20×1. The modified U-Net maps the information from the time axis to a spatial dimension and thereby reduces the resolution of the input PA raw sensor data from 2560×128 pixels to the target output resolution of 128×128 pixels. The architecture of the modified U-Net is shown in figure 4.3. This architecture was chosen because it is expected to loose less information when transforming the image from the time and space domain the the spacial domain in the skip connections as compared to doing it before or after the U-Net.

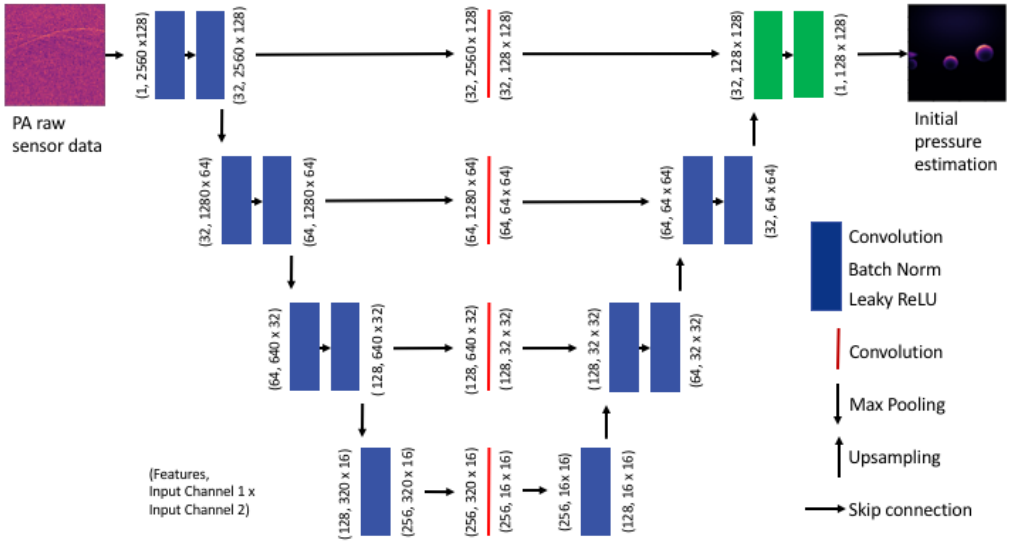


Figure 4.3: Architecture of the direct reconstruction U-Net. This modified U-Net is trained to estimate the initial pressure distribution directly from the PA raw sensor data. The skip connections of the U-Net include a layer which reduces the image dimensions in one direction by using an anisotropic kernel size of 20 × 3 and striding of 20 × 1. In each skip connection, the information is mapped from the time axis to a spatial dimension and thereby the dimensions of the input PA raw sensor data are reduced from 2560 × 128 pixels to the target output resolution of 128 × 128 pixels. The contracting and expansive layers are the same as used in the post processing U-Net.

5 Experiments and Results

Having decided which methods to use in this master thesis suitable experiments and datasets were created for evaluating the approaches feasibility. This validation is crucial to investigate if further work should be invested in applying deep learning for PA image reconstruction and to identify advantages and disadvantages of this approach and possibly future challenges. The experiments were conducted to test if the state-of-the-art can be outperformed in reconstruction accuracy with the deep learning networks. The first question was if deep learning can be used to post process DAS beamformed data. The the second question was if a deep learning network could be used to directly reconstruct the initial pressure distribution from the raw sensor data. *In silico* data was used to train and validate these algorithms, which has several advantages: First and most importantly, a large dataset for training was needed with corresponding initial pressure images as ground truth. Since the currently used methods beamforming and time reversal have drawbacks in reconstructing the initial pressure distribution from limited view PA measurements these are not suitable for creating a dataset. It was decided to use *in silico* data because the generation of simulated data is fast and easy and the sources of noise and artefacts are limited and well known when compared to *in vitro* or *in vivo* data. This approach has the additional benefit that it was possible to generate data with highly controllable parameters and complexity. Two datasets were generated, one with only one singlevessel in each image and one with multiple vessels per image. All parameters were chosen to be realistic for human tissue. The initial pressure images were generated using mcxyz Monte Carlo simulations in C++. Using this data the k-Wave simulation gave us the PA raw sensor data an ultrasound transducer would record. With MITK beamforming this sensor data was beamformed. The initial pressure data, the PA raw sensor data and the beamformed data was then used for the two deep learning approaches implemented in python.

5.1 Experimental datasets

Two experimental data sets were generated, one with a single vessel per image, containing 2000 images and one with multiple vessels per image, containing 3600 images, all were created in 2 dimensions. Having two data sets allows to perform a cross-validation between them, and the generalization capabilities of the algorithm trained on the singlevessel data set can be evaluated on the mulitvessel data set. The initial pressure distributions of single and

multiple vessels, which were generated with the mcxyz Monte Carlo simulation, used the settings for the simulation, which are shown in table 5.1:

Property	Value range	Unit
Reduced Scattering	15	cm^{-1}
Background Absorption	0.1	cm^{-1}
Vessel Absorption	U(2-8)	cm^{-1}
Vessel Radius	U(0.5 - 4)	cm
Number of Vessels	U(1 - 7)	

Table 5.1: Tissue properties used in the mcxyz Monte Carlo simulation. U denotes a uniform random distribution between the stated limits.

The data given in table 5.1 was chosen in order to simulate realistic tissue data. The following assumptions on tissue properties were chosen: The number of vessels in the singlevessel dataset was maximal one and it was randomly chosen between 1 and 7 vessels for the multivessel dataset. The vessel radius was chosen to have a large variation in radius but no vessel which is too small or too large. The vessel absorption was chosen to be in the range of the absorption properties of blood as they are shown in figure 2.5. The reduced scattering coefficient was set to 15 cm^{-1} and the background absorption was set to 0.1 cm^{-1} as described in [Jacques, 2014].

The initial pressure distribution images were processed with k-Wave to generate PA raw sensor data. The k-Wave framework with the k-Wave 2D FFT pseudo-spectral domain method for a line sensor was used for these simulations. The properties set in the k-Wave simulation are listed in table 5.2.

Property	Value	Unit
Tissue Parameters:		
Speed of sound in tissue	1500	$\frac{m}{s}$
Medium density	1000	$\frac{kg}{m^3}$
Absorption	0	cm^{-1}
Image properties:		
Input image size	128 × 128	$Pixel$
Pixel spacing	0.3	mm
Simulation properties		
Sensor recording time	$3.84 \cdot 10^{-5}$	s
Time spacing	$1.5 \cdot 10^{-8}$	s
Number of time steps	2560	$Pixel$
Simulation depth	57.6	mm
Number of transducer Elements	128	
Transducer center frequency	$7.5 \cdot 10^{6}$	Hz
Transducer bandwidth	80	%
Perfectly matched layer size x-direction	20	$Pixel$
Perfectly matched layer size y-direction	20	$Pixel$

Table 5.2: Properties used in the k-Wave simulation.

All properties were chosen in order to simulate human tissue as realistic as possible. The speed of sound in tissue varies between 1450 $\frac{m}{s}$ for fat to 1580 $\frac{m}{s}$ for muscle it was chosen to 1500 $\frac{m}{s}$ for this simulation. The medium

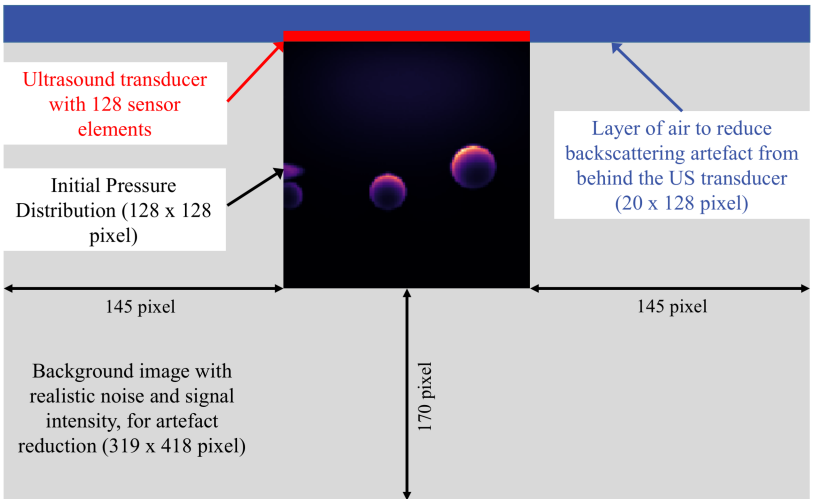
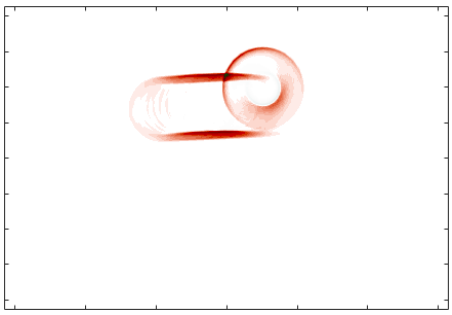


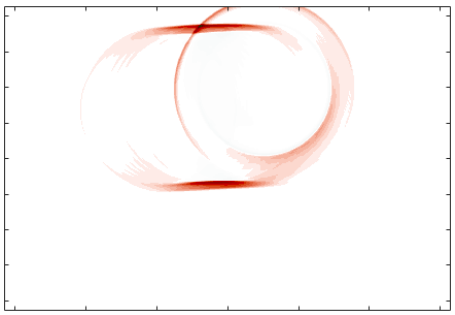
Figure 5.1: For simulating the PA raw sensor data with the k-Wave toolkit the aim was to simulate the tissue properties as realistically as possible. Therefore, the image was embedded in a tissue realistic background and added a thin layer of air above the sensor.

density was chosen to 1000 $\frac{kg}{m^3}$ with fat being slightly lighter and muscle slightly heavier. Because the ultrasound transducer, which was used for our laboratory system, has 128 transducer elements the images were chosen to be 128 $\times$ 128 pixels with a pixel spacing of 0.3 mm. This resulted in images of a width of 3.84 cm. The time spacing of $1.5 \cdot 10^{-8} s$ is equivalent to a sampling rate of about 67 MHz. This results in PA raw sensor data dimensions of 128 to 2560 pixels and therefore a simulation depth of 5.76 cm. The transducer center frequency was chosen to be 7.5 MHz with a bandwidth of 80 % in order to simulate the US transducer used for our PA system. In order to reduce backscattering artefacts in the simulation a layer of 20 $\times$ 128 pixels of air above the sensor was simulated. In order to reduce artefacts each initial pressure distribution image was embedded in a 319 $\times$ 418 pixel sized background image so that on the left and right border of the image a background of 145 pixels and at the bottom of the image was a background of 170 pixels, not including the 20 pixels of air which also had to be taken into account, which is shown in 5.1.

The k-Wave simulation aims at simulating US waves as realistically as possible. This includes simulating sources of artefacts at the border of an image, for example caused by the phantoms wall or artefacts by a bandwidth-limited US transducer measurement. These artefacts were reduced by choosing an appropriate experimental setup, shown in 5.1.



(a) Visualization of the pressure wave in tissue a few moments after the simulation started.



(b) Visualization of the pressure wave in tissue at a later timestep.

Figure 5.2: Simulating the acoustic pressure in homogeneous tissue with k-Wave. The orange colour indicates regimes of high pressure as it propagates through tissue at subsequent time steps. The waves spread from the initial pressure distribution. The virtual US transducer is simulated in the upper part of the image. The PA raw sensor data recorded by the US transducer.

5.2 Performance Assessment

Of both datasets 20% were used as test set, the rest was spitted in 80% training set and 20% validation set. Random Gaussian noise was added with a standard deviation of 20% of the mean maximum pixel intensity over all simulated image slices in each dataset to the PA raw sensor data. Images in the training set were mirrored for data augmentation to increase the number training samples. Both models were trained with the same amount of data presented. The results reported in the following section were calculated on the test set while the validation set was used to supervise convergence of the training process and hyper-parameter optimization. These hyper-parameters are show in table 5.3. The Number of epochs was set to 1000 since after about 400 epochs the cost function did not decrease further, by learning for 1000 epochs it was made sure that the network had finished learning. The batch size was set to 2, so in every batch were 2 images, considering the limited GPU capabilities and the large model size. The number of batches per epoch was set to 400, in in every epoch the algorithm optimized its weights on 400 randomly chosen images from the training set and took 100 randomly chosen images from the validation set. The base learning rate was set to $1 \cdot 10^{-4}$ with a decay of $1 \cdot 10^{-5}$. This was experimentally evaluated to be the most suitable learning rate. The optimizer's weight decay was chosen to $1 \cdot 10^{-5}$ to prevent overfitting.

Property	Value	Unit
Number of Epochs	1000	
Number of batches	200	per Epoch
Batchsize	2	
Number of validation batches	50	per Epoch
Base Learning Rate	$1 \cdot 10^{-4}$	
Learning Rate Decay	0.985	
Optimizer weight decay	$1 \cdot 10^{-5}$	

Table 5.3: Hyper-parameters setup for the deep learning algorithm.

For reporting of the qualitative results the relative initial pressure estimation error was calculated to the ground truth for each estimated image with this metric:

$$err = \left| \frac{\hat{IP} - IP}{IP} \right| \tag{5.1}$$

where err is the calculated error, $\hat{IP}$ is the initial pressure estimation and IP is the simulated ground truth initial pressure.

5.3 Results

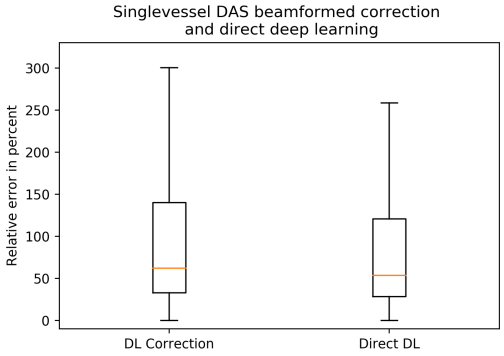
Both, the singlevessel and the multivessel dataset were evaluated quantitatively and qualitatively in order to assess the performance of both approaches. They have been evaluated on the corresponding test sets which contained 20% of the images of each dataset. To the test sets the same random Gaussian noise as in the training datasets was added. The training and test sets were strictly split so the test sets contained images which were never used before by the neural network. In order to investigate how the networks are able to generalize from the given training data a crosswise validation was calculated where the network trained on the singlevessel data was tested on the multivessel dataset and the network trained on the mulivessel dataset was evaluated on the singlevessel dataset. It took longer than three months of running the simulations full time on three computers and four virtual machines on an OpenStack Supercomputer clusters with a total of 160 cores running to generate the datasets, but only a few minutes to evalute the test

Figure 5.4 and 5.9 show qualitative comparisons of DAS beamformend images, of post processed images, directly reconstructed images and the ground truth simulated initial pressure distribution. The images were sorted by the pixel wise relative error to the ground truth on the region of interest. The region of interest is defined by the intensity of the initial pressure distribution, an intensity above a fixed threshold is defined as region of interest. Only images which were not empty were taken into account. Qualitatively it is visible that the DAS beamforming approach is performing worse in greater depth than near the sensor, this is especially visible in figure 5.9 (b). The direct deep learning approach can reconstruct the position correctly but does not reconstruct the correct shape of the pressure distribution.

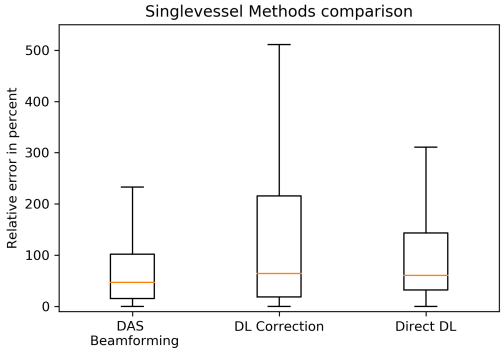
Singlevessel

The quantitative comparison of the pixel wise mean difference on the whole singlevessel images is shown in Figure 5.3(b). The images have been normalized using z-score normalization over the whole dataset before the evaluation. The median relative initial pressure estimation error changed from 47% with an interquartile range (IQR) of [15% - 102%] in case of the DAS beamforming to 64% with IQR [19% - 215%] using the proposed deep learning architecture for correcting DAS beamformed images. The relative error of the direct deep learning approach was 60% with IQR [32% - 143%]. In a direct comparison without prior image normalization, the median relative initial pressure estimation error was 62% with IQR [33% - 140%] in case of the correction of DAS beamformed images and the direct approach yields a median error of 53% with IQR [28% - 120%] as shown in figure 5.3(a).

For the reconstruction approaches of the image containing singlevessels the highest absolute errors occur in the region of the vessel as it can be seen in figures 5.5 (b), 5.6 (b) and 5.7 (b). All three reconstructions also produce errors in the background. For the relative errors as shown in figures 5.5 (b), 5.6 (b) and 5.7 (b) a halo-like structure in the upper half of the image is visible. For the relative error images, larger error is made by the deep learning



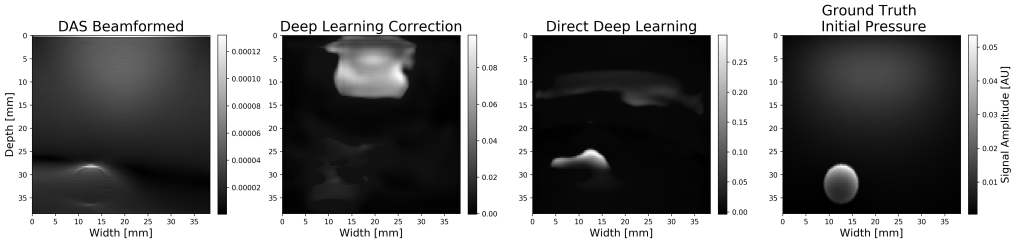
(a) Non-normalized images for direct comparison of the two methods.



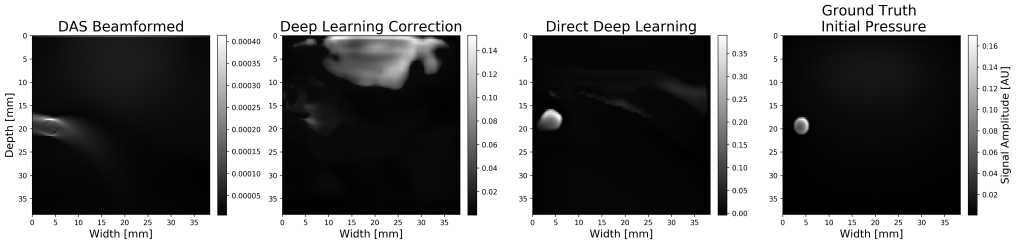
(b) Normalized images using z-score normalization for comparison of the suggested methods.

Figure 5.3: Boxplots of the pixel-wise relative initial pressure distribution estimation errors. Both deep learning (DL) methods do not yield superior results as compared to DAS Beamforming.

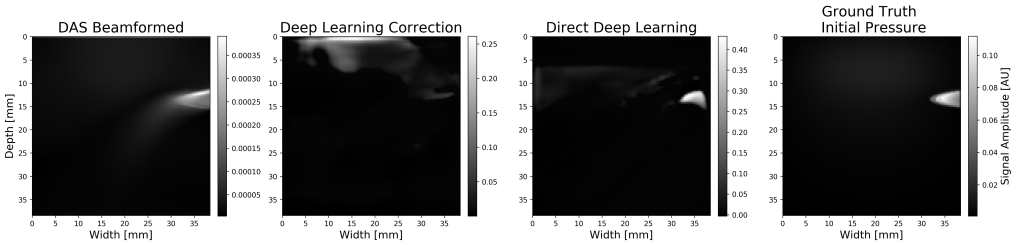
approaches at the borders of the vessel. This might be by either not constructing the size of the vessel or the position correctly. The DAS beamforming approach makes large reconstruction errors in the vessel.



(a) Singlevessel image with the smallest error. The deep learning correction fails in the reconstruction and the direct deep learning estimates the position correctly but fails to reconstruct the exact shape and size.

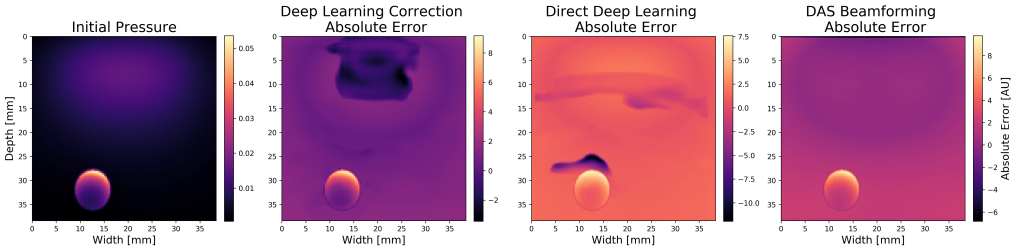


(b) Singlevessel image with a median error. The deep learning correction fails in the reconstruction and the direct deep learning estimates the position correctly but fails to reconstruct the exact shape and size.

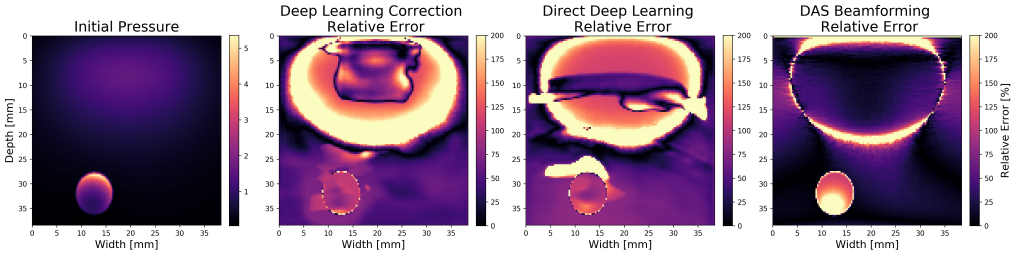


(c) Singlevessel image with the largest error. Both deep learning approaches fail to reconstruct the blood vessel.

Figure 5.4: Sample images from the singlevessel test set. They are sorted by the relative pixel wise error on the region of interest from the ground truth on the direct reconstruction test set. The region of interest was defined by a threshold on the initial pressure image pixelvalues.

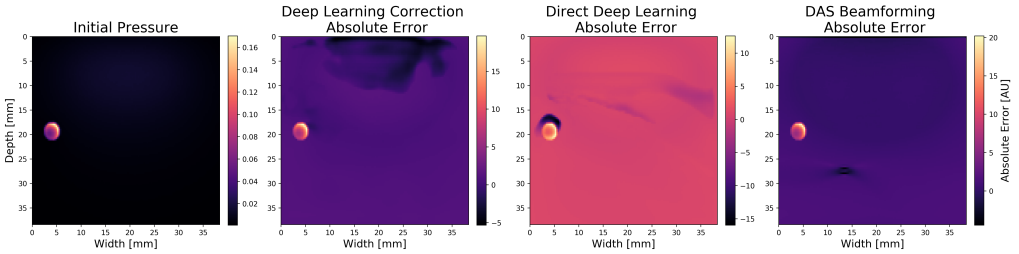


(a) Absolute error images as sorted by the direct deep learning image with the smallest error. All approaches make the highest error in the region of the initial pressure.

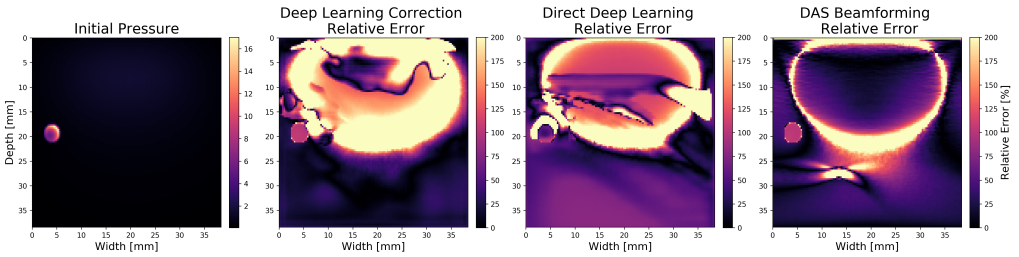


(b) Relative error images as sorted by the direct deep learning image with the smallest error. One can see a halo-like structure in the upper half of the images. Both deep learing approaches make the largest error in the halo-like structure and at the border of the blood vessel. The DAS beamforming approach has the largest error in the vessel.

Figure 5.5: Singlevessel image with the smallest error as evaluated on the direct deep learning images as shown in figure 5.4. (a) shows the absolute error images and (b) the relative error images for the three reconstruction methods.

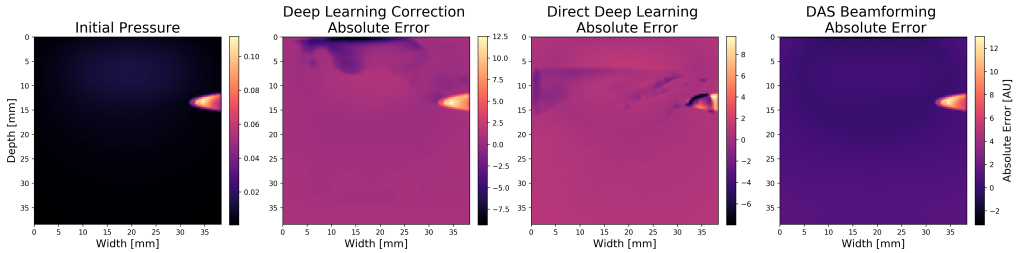


(a) Absolute error images for the direct deep learning image with the median error. All three approaches make the largest error in the vessel structure.

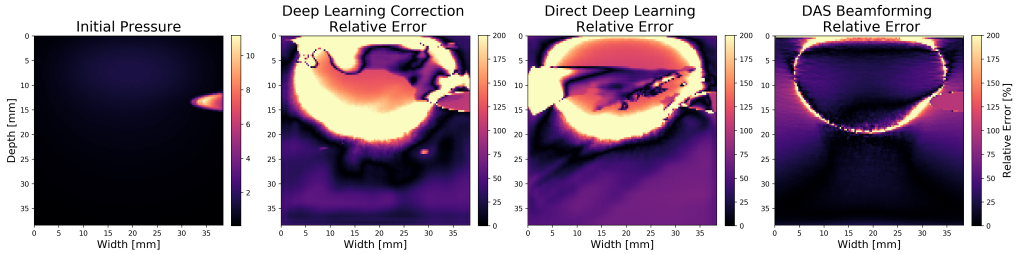


(b) Relative error images as sorted by the direct deep learning image with the median error. The halo-like structure in the upper half of the image is clearly visible in all three approaches. Both deep learning approaches make an error around the vessel and the DAS beamforming approach and the deep learning correction make errors inside the vessel.

Figure 5.6: Singlevessel image with the median error as evaluated on the direct deep learning images and as shown in figure 5.4. (a) shows the absolute error image and (b) the relative error image for the three reconstruction methods.



(a) Absolute error images as sorted by the direct deep learning image with the largest error. All three approaches make large errors in the region of the initial pressure distribution and in the background of the image.

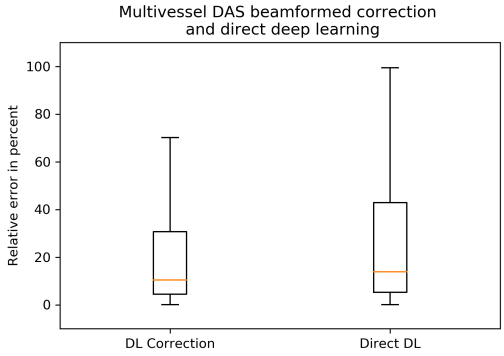


(b) Relative error images as sorted by the direct deep learning image with the largest error. The halo-like structure in the upper half of the image is visible for all three approaches but is appearing brighter at both deep learning approaches.

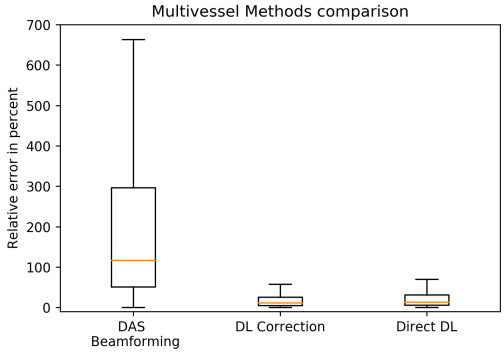
Figure 5.7: Singlevessel image with the largest error as evaluated on the direct deep learning images as shown in figure 5.4. (a) shows the absolute error images and (b) the relative error images for the three reconstruction methods.

Multivessel

The quantitative comparison of the pixel wise mean difference on the whole images is shown in Figure 5.8(b). The images have been normalized using z-score normalization over the whole dataset before the evaluation. The median relative initial pressure estimation error improved from 117% with an interquartile range (IQR) of [52% - 296%] in case of the DAS beamforming to 12% with IQR [0.5% - 26%] using the proposed deep learning architecture for correcting DAS beamformed images. The relative error of the direct deep learning approach was 14% with IQR [0.6% - 32%]. In a direct comparison without prior image normalization, the median relative initial pressure estimation error was 10% with IQR [0.4% - 31%] in case of the correction of DAS beamformed images and the direct approach yields a median error of 14% with IQR [0.5% - 43%] as shown in Figure 5.8(a). These results have been published before in [Waibel et al., 2018a, Waibel et al., 2018b].



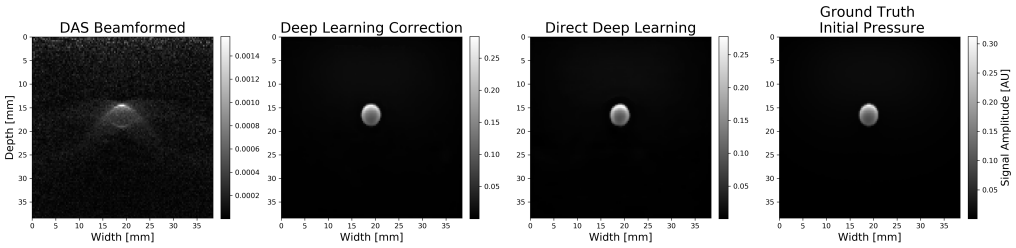
(a) Non-normalized images for direct comparison of the two methods.



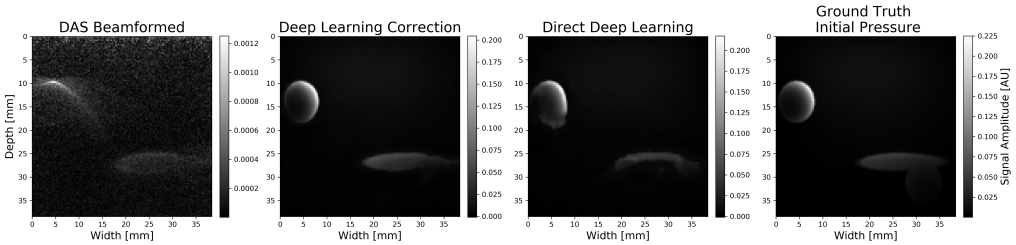
(b) Normalized images using z-score normalization for comparison of the suggested methods.

Figure 5.8: Boxplots of the pixel-wise relative initial pressure distribution estimation errors. Both methods yield huge advantages compared to DAS beamforming. A slightly lower relative error was achieved by using a U-Net to correct DAS beamformed images compared to the direct approach. These have been published before in [Waibel et al., 2018a, Waibel et al., 2018b].

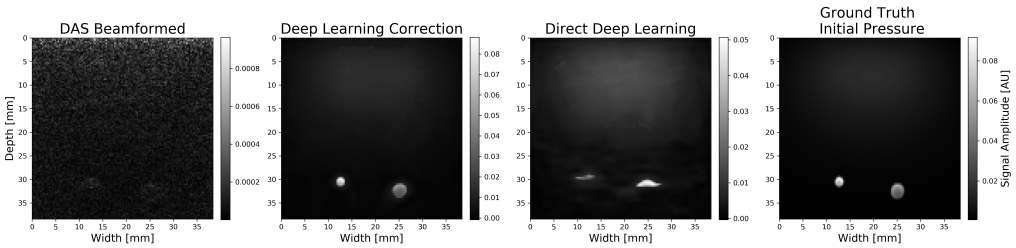
In figure 5.10 (a), 5.11 (a) and 5.12 (a) the absolute error of the normalized images are shown to indicate where



(a) Multivessel image with the smallest error. The signal is very close to the tissue surface and the shape is preserved very well.



(b) Multivessel image with a median error. Even though both different shaped vessels were reconstructed, the direct approach yields some inaccuracies especially in the lower parts.



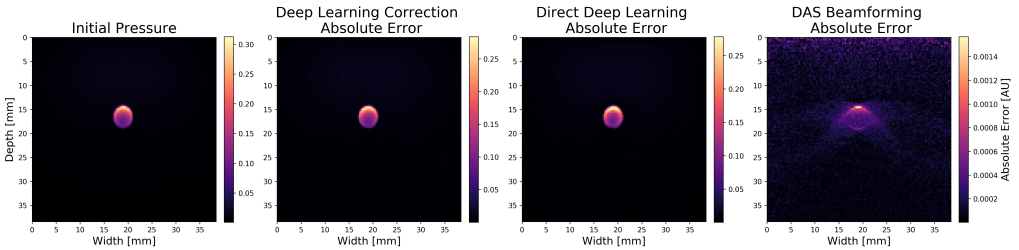
(c) Multivessel image with the largest error. The signal is very deep inside the tissue hardly visible in the DAS beamformed image and with large errors in the direct approach.

Figure 5.9: Sample images from the multivessel test set. They are sorted by the relative pixel wise error on the region of interest from the ground truth on the direct reconstruction test set. The region of interest was defined by a threshold on the initial pressure image pixelvalues. These have been published before in [Waibel et al., 2018a, Waibel et al., 2018b].

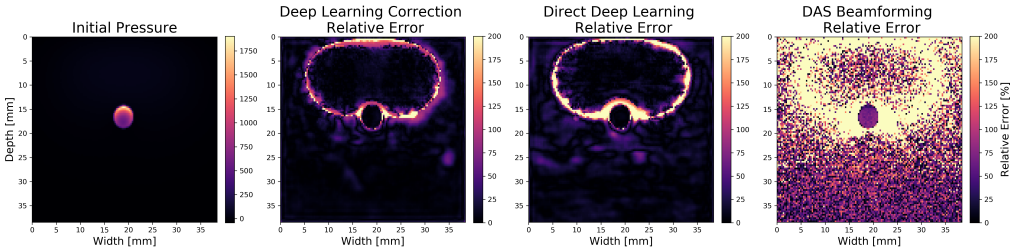
the highest absolute errors in the images errors are. The image on the left is the ground truth and therefore the ideal reconstruction. For all images, the largest errors occur in the area of the largest signal which could indicate a wrong prediction of the signal intensity. The circular artefacts in the DAS beamforming images are clearly visible in figure 5.10 (b), which is a known drawback of this method [Hauptmann et al., 2017], also there are errors on the upper part of the image, which are not visible in the images of both deep learning approaches. These artefacts are even larger than the signal of a vessel deep in tissue as can be seen in figure 5.12.

The relative errors or these images are plotted in figure 5.10 (b), 5.11 (b) and 5.12 (b). These images have been normalized before computing the relative error. The relative error images have been rescaled to the same scale. On the first look, it is obvious that the DAS beamforming approach makes the most relative errors throughout the

image. At all three approaches have a halo-like structure in the upper half of the images. And all three approaches have problems reconstructing the correct intensity in some images, especially if the vessel is covered by vessel with most presumably absorb most photons as it is visible in figure 5.11 (b). Comparing both deep learning approaches it is visible that both make similar errors on or around the borders of the vessels but the direct deep learning method might do slightly less good on the vessels intensity as can be seen in figure 5.12 (b).



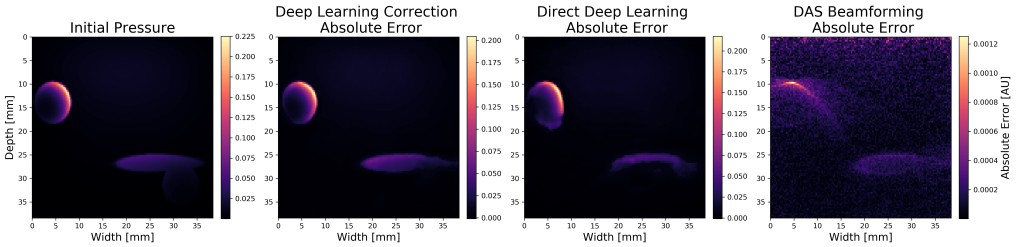
(a) Absolute error images as sorted by the direct deep learning image with the smallest error. For both deep learning approaches the largest errors are at the reconstructed vessel. The DAS beamforming approach does makes the largest error on the border of the vessel. The maximum absolute error is similar high for both deep learning algorithms.



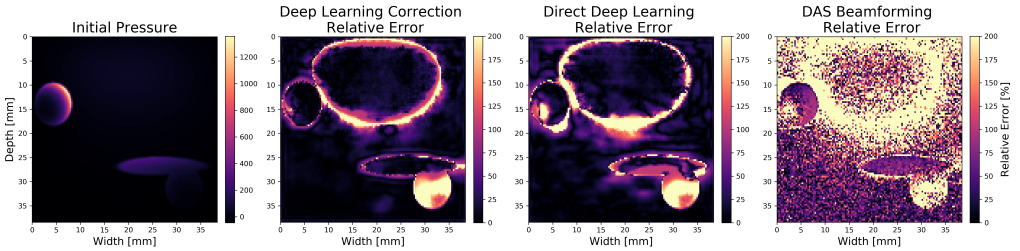
(b) Relative error images as sorted by the direct deep learning image with the smallest error as shown in 5.9 (a). These images have been normalized before computing the relative error. The relative error images have been rescaled to the same scale. All three approaches make errors at a halo-like structure in the upper half of time image. Both deep learning approaches do make errors on the borders of the vessels.

Figure 5.10: Images with the smallest error, calculated as the pixelwise relative error to the ground truth as shown in figure 5.9. (a) shows the absolute error image and (b) the relative error image for the three reconstruction methods.

Absolute error image for the image with the largest error. The largest errors for the deep learning correction approach occur on the borders of the vessels. The direct deep learning and the DAS beamforming algorithm both do not correctly reconstruct the intensity in the vessel. The maximum absolute error is lowest for the deep learning correction approach and similar in the direct deep learning approach and DAS beamforming.

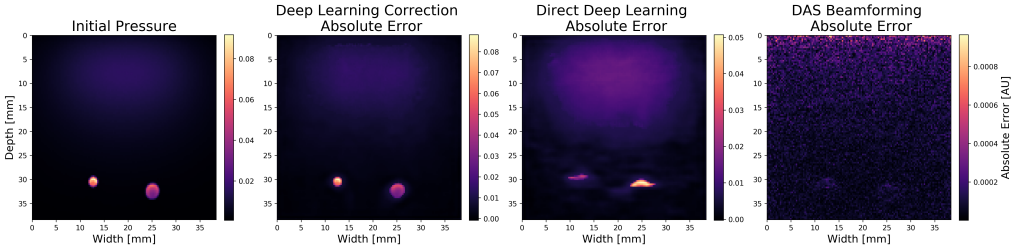


(a) Absolute error images as sorted by the direct deep learning image with a median error. The deep learning correction algorithm does not reconstruct the vessel in great depth under the horizontal vessel correctly. The direct deep learning and the DAS beamforming approach does recognize this vessel. The maximum absolute error is increasing from deep learning correction to direct deep learning and then approximately doubling for DAS beamforming.

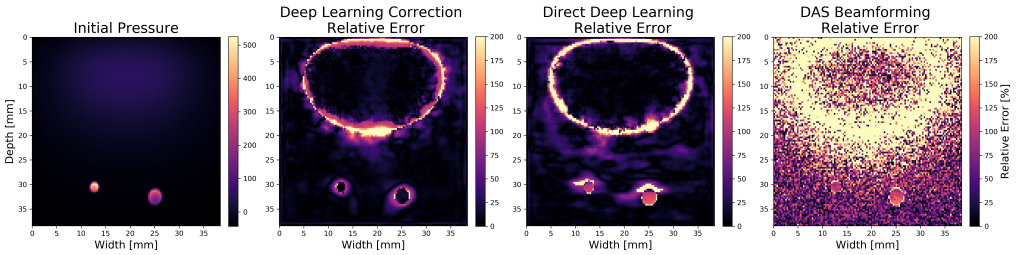


(b) Relative error images as sorted by the direct deep learning image with the median error. Both deep learning approaches do make errors on the borders of the vessel but correctly reconstruct the position of the vessel. For the direct deep learning approach an artefact above the vessels is visible.

Figure 5.11: Images with the median error, calculated as the pixelwise relative error to the ground truth as shown in figure 5.9. (a) shows the absolute error image and (b) the relative error image for the three reconstruction methods.



(a) Absolute error image as sorted by the direct deep learning image with the largest error. The largest errors for the deep learning correction approach occur on the borders of the vessels. The direct deep learning and the DAS beamforming algorithm both do not correctly reconstruct the intensity in the vessel. The maximum absolute error is lowest for the deep learning correction approach and similar in the direct deep learning approach and DAS beamforming.



(b) Relative error images as sorted by the direct deep learning image with the median error. All three approaches miss to reconstruct the vessel deep in the tissue which is located below with the horizontal vessel. The signal of this vessel might be very weak because the vessel is deep in tissue and the signal is might be distracted by the signal from the vessel laying above.

Figure 5.12: Images with the largest error, calculated as the pixelwise relative error to the ground truth as shown in figure 5.9. (a) shows the absolute error image and (b) the relative error image for the three reconstruction methods.

Crossevaluation

In order to assess how the deep learning networks, perform when they have to generalize from the learned task onto another problem the networks which have been trained on singlevessel images was evaluated on multivessel images and the networks which have been trained on multivessel images was evaluated on singlevessel. This was done for both, the DAS beamforming correction and the direct deep learning approach.

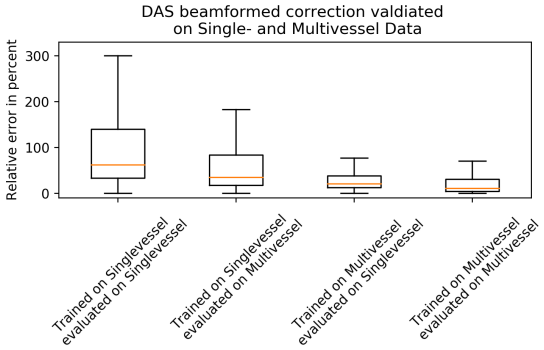
The quantitative comparison of the pixel wise mean difference of the crosswise tested images is shown in Figure 5.13. The images have not been normalized before the evaluation. For the direct DAS Beamforming deep learning approach the median relative errors from the reconstruction images to the ground truth with interquartile ranges (IQR) are listed in table 5.4.

DAS Beamforming Correction Approach	Median	Interquartile Range
Trained on singlevessel dataset and tested on the singlevessel test set	61	[32 - 139]
Trained on singlevessel dataset and tested on the multivessel test set	34	[17 - 83]
Trained on multivessel dataset and tested on the singlevessel test set	20	[12 - 38]
Trained on multivessel dataset and tested on the multivessel test set	10	[0.5 - 30]

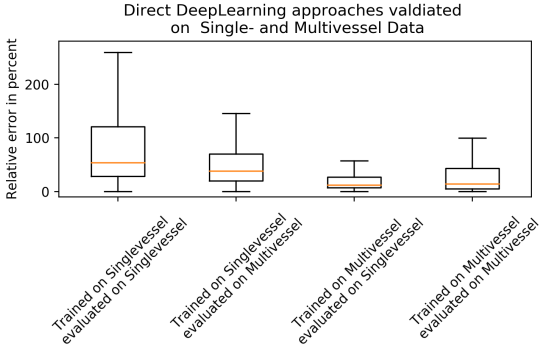
Table 5.4: Median relative errors of the cross tested images for the DAS Beamforming correction approach and their interquartile ranges.

Direct Deep Learning Approach	Median	Interquartile Range
Trained on singlevessel dataset and tested on the singlevessel test set	53	[28 - 120]
Trained on singlevessel dataset and tested on the multivessel test set	38	[20 - 70]
Trained on multivessel dataset and tested on the singlevessel test set	12	[0.7 - 27]
Trained on multivessel dataset and tested on the multivessel test set	14	[0.5 - 43]

Table 5.5: Median relative errors of the cross tested images for the direct deep learning approach and their interquartile ranges.



(a) Boxplots of the relative errors of the four different DAS Beamforming correction deep learning approaches.



(b) Boxplots of the relative errors of the four different direct deep learning approaches.

Figure 5.13: Comparing the all approaches tested on all test sets available. The Networks which was trained on the singlevessel dataset performs better if evaluated on the multivessel test set than on the singlevessel test set for both approaches. The network trained on the multivessel dataset performs better on the multivessel test set and in a similar quality on the singlevessel test set and it performs better than the networks which was trained on the singlevessel data.

6 Discussion and Conclusion

In limited view photoacoustic imaging (PAI), it is essential to develop accurate and fast reconstruction methods. In this master thesis two methods are presented that can be used to either enhance reconstruction of Delay-and-Sum (DAS) beamforming with a post processing step or provide a direct reconstruction of the initial pressure distribution from PA raw sensor data. The results show high accuracy even with noisy raw data, and provide a qualitative improvement compared to DAS beamforming. In future research this approach should be transfered to *in vitro* and *in vivo* data. The two proposed approaches that can improve currently used PAI reconstruction algorithms in limited view settings are:

1. **Deep learning post processing of DAS beamformed images**

2. **Direct deep learning reconstruction of PA raw sensor data**

The first method was used to post process Delay-and-Sum (DAS) beamformed PA raw sensor data. It was shown that systematic errors and artefacts can be detected and corrected with state-of-the-art deep learning methods. The relative error on the multivessel dataset was 12% with an interquartile range (IQR) of [0.5% - 26%]. This was evaluated using the proposed deep learning architecture for correcting DAS beamformed images, where conventional DAS beamforming fails with a relative error of 117% with an IQR of [52% - 296%]. For evaluation these images have been normalized over the whole dataset. Qualitatively it is visible that for example the circular artefacts in the DAS beamformed images can be reduced using deep learning as shown in figure 5.9. The presented post processing method achieved qualitative and quantitative improvements compared to DAS beamforming and yielded highly accurate results even under application of noise to the simulated raw data. DAS beamforming is a method whose aim is to yield the origin of US waves and not to reconstruct the initial pressure distributions [Mozaffarzadeh et al., 2017]. DAS beamforming fails at the reconstruction of initial pressure distributions.

It was shown that initial pressure distributions can also be directly reconstructed from the PA raw sensor data with a modified U-Net without the use of any prior processing steps. This method needs one calculation step less but gives slightly higher relative errors for the multivessel dataset than the deep learning correction method. For a quantitative comparison, the relative error on the normalized multivessel dataset was 14% with IQR [0.6% - 32%]. For the not-normalized multivessel images an error of 10% with an IQR of [0.4% - 31%] was achieved for the deep

learning post processing approach and 14% with an IQR of [0.5% - 43%] for the direct deep learning approach. The higher relative error may be due to the increased model size which is caused by the enlarged input space of 2560 pixels as compared to 128 pixels in axial direction. The main challenge was to transform the image dimensions from the PA raw sensor data to the image dimensions of the initial pressure distribution. This was solved by introducing additional convolutional layers in the skip connections of the U-Net as shown in figure 4.3. The skip connections were originally introduced to retain the entire range of high frequency information. Without skip connections the information contained in high frequencies in the time domain would be lost by a large perceptive field along the time axis. With the modified U-Net the problem of loosing high-frequency information might be re-introduced because the skip connections were modified in order to transfer the image from the time to the space dimension. One future step could be to investigate how the neural network's architecture can be modified in order to improve the direct reconstruction from PA raw sensor data. With an advanced architecture the reconstruction of the deep learning post processing approach might in future be outperformed.

When comparing the qualitative results to the quantitative results the error metric used to calculate the quantitative results plays an important role. For the evaluation the relative pixelwise error over the whole image to the ground truth was used, as shown in equation 5.1. This metric considers all pixels of the image, therefore also errors in image regions with low signal are accounted for. For the evaluation of the algorithms performance the reconstruction accuracy on low signal regions in the image is important, therefore these regions were taken equally into account. For the evaluation of the qualitative results a region of interest was defined which only included regions above a threshold of intensities on the initial pressure images. This was done because the aim was to evaluate the algorithms qualitative performance on the task to reconstruct the blood vessels.

For a qualitative comparison the absolute errors to the ground truth are plotted in figures: 5.5 (a), 5.6 (a) and 5.7 (a) for the singlevessel images and in figures: 5.10 (a), 5.11 (a) and 5.12 a) for the multivessel images. In all cases it becomes clear that the largest errors are made in the region of the largest signal amplitude.

The relative errors to the ground truth pressure distribution on the singlevessel dataset are plotted in the figures: 5.5 (b), 5.6 (b) and 5.7 (b). When evaluating the deep learning correction on the singlevessel dataset it becomes clear, that this method fails to reconstruct the initial pressure distribution. One reason could be the small dataset used. The direct deep learning approach on the singlevessel dataset sometimes is able to reconstruct the position of the initial pressure distribution, but often fails to reconstruct the exact shape or size of the blood vessel. For horizontal blood vessels as one can be seen in figure 5.7 from the singlevessel dataset, the DAS beamforming approach delivers the best results in this example. Both deep learning approaches fail to reconstruct shape and position of the vessel.

Relative errors for multivessel images are plotted in the figures: 5.10 (b), 5.11 (b) and 5.12 (b). While both deep learning approaches are generally able to reconstruct the position of the blood vessel, they make errors at the border of the blood vessels. This might be caused by not getting the size of the vessels correctly. Figure 5.12 (b) shows the

largest errors which the deep learning approaches make when approximating the initial pressure distribution. These approaches show the ability to reconstruct the inner part of the vessels with a low error but large errors occur at the borders of the vessels. The DAS beamforming algorithm estimates the vessel position well but does not reconstruct the intensity in the vessel correctly which can be seen on the wave-like artefacts around the reconstructed initial pressure distribution in figure 5.9.

When evaluating both deep learning approaches on the singlevessel and multivessel test set, one recognizes that both approaches generally perform better on the multivessel test set than on the singlevessel test set. The deep learning correction on the singlevessel test set fails completely and the direct reconstruction has problems reconstructing the exact position, shape and size of the blood vessels. One possible reason is the size of the dataset used for training the algorithms.

The size of the dataset is important for deep learning algorithm to generalize. It was estimated that in the singlevessel dataset every second image showed no vessel whereas in the multivessel dataset about one in ten images showed no vessel. In the simulation of the vessel structures with the mcxyz Monte Carlo simulation, a limited number of vessels, one for the singlevessel images and multiple for the multivessel images were simulated in three dimensions. Slices from these three-dimensional volumes were cut, which by chance contained one, multiple or no vessels. This means that the neural network which was trained on the multivessel dataset, was trained on about 1900 images, which containing relevant information. The network which was trained on the singlevessel dataset, was trained on about 600 images, which contained relavent information for abstracting a model. Additionally the multivessel dataset contained more information per image, since there are multiple vessels in most images. These images have a greater variability and thus the algorithm is more competent in generalizing and therefore predicts the initial pressure distribution with less errors. The importance for deep learning methods to train on a large dataset with a large amount information is shown thereby. It is especially visible when comparing the results of the singlevessel and the multivessel dataset. In future, the possibility to use unsupervised learning could be explored as no ground truth data is needed for this approach and therefore large datasets are more easily available. But since in unsupervised learning no ground truth is available, the assessment of performance is difficult and there is no evidence that this approach is suitable for tackling the acoustic inverse problem.

In all images at the same position in the upper half of the image, an artefact in the shape of a halo-like structure is visible as shown in figure 6.1 and in the figures: 5.5 (b), 5.6 (b) and 5.7 (b) of the singlevessel images and in 5.10 (b), 5.11 (b) and 5.12 (b) of the multivessel images. Both deep learning approaches learn an average of this halo-like structure which they compensate for. As this structure is not the same for all images, due to different fluence rates, the deep learning approaches may produce this errors. To compensate for this artefact more training data could improve the networks performance.

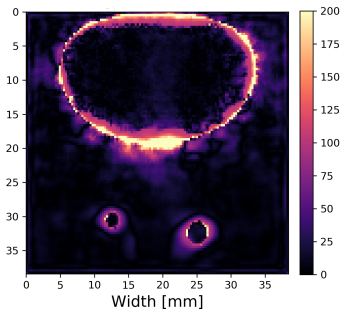


Figure 6.1: The halo-like structure in the relative error images in the upper half of the image. This might be be an artefact which is not correctly compensated in the deep learning approaches.

Generally it is visible that the performance of all presented approaches decreases with increased depth in tissue. A reason is the measurement angle of the US transducer, which decreases with depth. Thereby the limitation of the view increases, as visualized in figure 6.2. The wave front reaching the sensor from a deeper origin in tissue is also flatter than a wave front originating closer to the sensor. Because the shape of a wave contains information about the shape of the signal origin, this information is relevant for the reconstruction. If the wave is originating from a deeper section of the tissue less information is measured by the sensor about the shape of the signal origin. Therefore, it is easier for an algorithm to reconstruct the exact position of the origin, if the vessel is closer to the sensor. Another possible reason is that the initial pressure can be described with spherical waves. The intensity decreases with distance from the origin quadratically. And a signal which travels a longer distance in tissue also experiences stronger attenuation, therefore the resolution decreases with imaging depth. Also, the possibility of artefacts being created, increases with a longer distance travelled. This is difficult to improve because the physical properties of the US transducer can only be changed by the manufacturer.

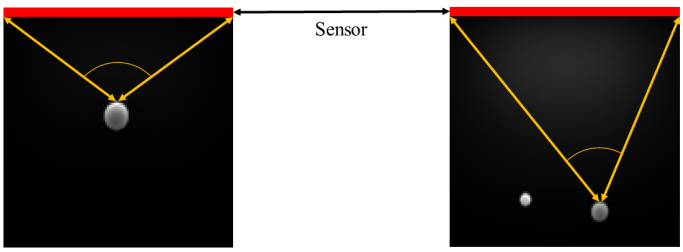


Figure 6.2: Visualization of the angle dependency of the measured signal from the imaging depth. The deeper the imaged blood vessel is located in the tissue, the smaller is the angle under which the signal is detected by the sensor. Therefore the limitation of the sensor's field of view increases.

For evaluating how these approaches generalize on data from another dataset crossevaluations have been performed. In Figure 5.13 it can be seen that the algorithm trained on multivessel data is generally performing better than the algorithm trained on singlevessel data. This could be due to the fact that the number of training data

was larger in the multivessel dataset which contained 3600 images whereas the singlevessel dataset only contained 2000 images. While it is no surprise that the algorithm trained on multivessels is performing well on the singlevessel dataset, since in the multivessel dataset there are images with only one vessel, it is a surprise that the algorithm trained on the singlevessel dataset is performing better on the multivessel dataset than on the singlevessel dataset. The singlevessel test sets contained 400 images, while the multivessel test sets contained 720 images. If the size of the test set and the number of not-empty images contained in it makes a significant difference when evaluating the approaches performances was not tested, but this seems unlikely for large enough test sets.

The proposed methods were not yet applied to *in vivo* or *in vitro* data. Starting directly with *in vivo* or *in vitro* data, presumably much more training images would have been needed and a lot of time would have been spent for denoising and artefact removal before it would have been possible to start training these deep learning algorithms. In future, the deep learning algorithms should be able to remove artefacts and noise automatically. The transfer step from *in silico* to *in vitro* and *in vivo* data could be done by pre-training the neural network on *in silico* data and then generalize it to *in vitro* data. By doing so, one has to save the weights from the network trained on the *in silico* data and then apply these weights to initialize a network which trains on more complex *in vitro* data. This was done in this master thesis with the network trained on singlevessel data, which was applied on multivessel data and vice versa. In future this could be applied to data with a realistic noise model and realistic artefacts. This could iteratively be done while improving the network's ability to generalize with each iteration. For this domain transfer, several specific factors have to be considered such as calibration, correct modelling of noise, angular sensitivity of the US transducer, the presence of artefacts and local variations of speed of sound and acoustic attenuation leading to a generally lower contrast-to-noise-ratio. As such, the next challenge is to generalize the proposed methods to be applicable to realistic data. Improving the deep learning approaches, other state-of-the-art solutions such as [Hauptmann et al., 2017, Antholzer et al., 2017] can be implemented to indicate best practices and to find the most reliable solution.

Especially in medicine, where reliable and accountable diagnoses and treatments are needed, deep learning approaches are sometimes viewed with scepticism, because they appear as 'black boxes' [noa, 2018, Samek et al., 2017, Ribeiro et al., 2016] and because these algorithms can suffer from significant difficulties [Shalev-Shwartz et al., 2017]. Therefore, their application at medical tasks is a matter of ongoing debate. Several strategies have been developed to understand what happens in intermediate levels of neural networks [Karpathy and Fei-Fei, 2014, Zeiler and Fergus, 2014, Springenberg et al., 2014, Nguyen et al., 2016, Montavon et al., 2017, Shalev-Shwartz et al., 2017] but it still remains unclear how the networks' mechanisms work because the underlying mechanisms cannot be directly influenced. This can lead to problems as it is not clear which features or details of an image influence the algorithms' predictions. It was also shown that this opacity can lead to issues of bias [Cathy O'Neil, 2016]. Consequently, the ability of the algorithms to generalize is limited [Marcus, 2018]. It is an ongoing discussion to

what extend these algorithms can be applied to problems, similar to the problems that they were trained for [Kim and Hwang, 2016, Andrew Ng, 2016, Antony et al., 2016, Litjens et al., 2017]. Another limitation of deep learning are tasks where insufficient information in the gradients about the underlying target function is available. A low signal-to-noise ratio, bad conditioning, or flatness of the activation functions are at hand [Shalev-Shwartz et al., 2017]. The huge amount of data, which deep neural networks need to generalize is a problem for many researchers. Several strategies, such as data augmentation [Isensee et al., 2017b, Perez and Wang, 2017] have been developed to address this issue. But for many tasks the amount of medical data in good quality is limited. Another drawback of machine learning is the difficulty to integrate prior knowledge, such as the speed of sound in tissue into the model. Partly this knowledge is represented in the networks features but not in quantified statements [Marcus, 2018]. And deep learning networks can be fooled [Vinyals et al., 2014, Nguyen et al., 2014] as recent real world examples show [Evtimov et al., 2017, Athalye et al., 2017]. Therefore, it is not certain that predictions of the neural network are correct and it is not understandable by what the networks predictions are determined. It is important not to misunderstand the misleading word deep learning, where 'deep' refers to a technical property and not to a conceptual. While human intelligence, even in new-borns [Gervain and Werker, 2008], is able to learn through abstract explicit definition, and implicit means from very little samples in minutes, artificial intelligence currently fails at the same task [Marcus, 2018].

Considering the advantages and disadvantages of using neural networks, one has to decide on the given task if neural networks are suitable for the problem or not. Looking at the overwhelming success of neural networks on some medical imaging tasks, the possible advantages are apparent. But one has to bear in mind that the development of artificial intelligence is nowhere near the human-like ability in solving unfamiliar problems because their abilities are highly specialized while the human intelligence is very general.

In this master thesis it was shown that the application of deep learning on the photoacoustic inverse problem on *in silico* data leads to promising results. Two approaches have been presented: The first utilized Delay-and-Sum beamforming and deep learning for post processing. The second approach shows that it is possible to solve the acoustic inverse problem for limited view PA directly with a single modified deep neural network. The modified U-Net has learned to transfer the PA raw sensor data from the time and space domain to a two dimensional spacial domain, abstracting and learning the underlying physical model. Both approaches were able to detect and remove artefacts and compensate noise. The qualitative and quantitative evaluation of both approaches show promising results and outperform DAS beamforming. In future this approach can be transferred to *in vitro* and *in vivo* data, which is a possible next step on the road to clinical quantitative photoacoustic imaging.

Part I

Appendix

A Lists

List of Figures

List of Tables

B Bibliography

[noa, 2018] (2018). AI diagnostics need attention. *Nature*, 555(7696):285.

[Agichtein et al., 2006] Agichtein, E., Brill, E., and Dumais, S. (2006). Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 19–26, New York, NY, USA. ACM.

[Andrew L Maas et al., 2013] Andrew L Maas, Hannun, A. Y., and Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models.

[Andrew Ng, 2016] Andrew Ng (2016). What AI Can and Can't Do.

[Antholzer et al., 2017] Antholzer, S., Haltmeier, M., and Schwab, J. (2017). Deep Learning for Photoacoustic Tomography from Sparse Data. *arXiv:1704.04587 [cs]*. arXiv: 1704.04587.

[Antony et al., 2016] Antony, J., McGuinness, K., Connor, N. E. O., and Moran, K. (2016). Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks. *arXiv:1609.02469 [cs]*. arXiv: 1609.02469.

[Arridge et al., 2016] Arridge, S. R., Betcke, M. M., Cox, B. T., Lucka, F., and Treeby, B. E. (2016). On the Adjoint Operator in Photoacoustic Tomography. *Inverse Problems*, 32(11):115012. arXiv: 1602.02027.

[Athalye et al., 2017] Athalye, A., Engstrom, L., Ilyas, A., and Kwok, K. (2017). Synthesizing Robust Adversarial Examples. *arXiv:1707.07397 [cs]*. arXiv: 1707.07397.

[Bamber and (C. R. Hill, 2004] Bamber, J. C. and (C. R. Hill, J. C. Bamber, a. G. R. t. H. e. (2004). Attenuation and Absorption in Physical Principles of Medical Ultrasound. In *Attenuation and Absorption in Physical Principles of Medical Ultrasound*, pages pp. 93–166. Chichester: John Wiley.

[Bamber and Hill, 1979] Bamber, J. C. and Hill, C. R. (1979). Ultrasonic attenuation and propagation speed in mammalian tissues as a function of temperature. *Ultrasound in Medicine & Biology*, 5(2):149–157.

[Beer, 1852] Beer (1852). Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten - Beer - 1852 - Annalen der Physik - Wiley Online Library.

[Bell, 1880] Bell, A. G. (1880). On the production and reproduction of sound by light. *American Journal of Science*, 3(20):305–324.

[Bernsen et al., 2015] Bernsen, M. R., Kooiman, K., Segbers, M., van Leeuwen, F. W. B., and de Jong, M. (2015). Biomarkers in preclinical cancer imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, 42(4):579–596.

[Bertrand Russel and Alfred North Whitehead, 1913] Bertrand Russel and Alfred North Whitehead (1913). *Principia Mathematica*. Cambridge University Press.

[Bojarski, 1982] Bojarski, N. N. (1982). The k-space formulation of the scattering problem in the time domain. *The Journal of the Acoustical Society of America*, 72(2):570–584.

[Boole, 1854] Boole, G. (1854). *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. Dover Publications.

[Bowen, 1983] Bowen, T. (1983). *Radiation_ induced thermoacoustic imaging*. Google Patents.

[Carovac et al., 2011] Carovac, A., Smajlovic, F., and Junuzovic, D. (2011). Application of Ultrasound in Medicine. *Acta Informatica Medica*, 19(3):168–171.

[Cathy O'Neil, 2016] Cathy O'Neil (2016). Weapons of Math Destruction, How Big Data Increases Inequality and Threatens Democracy by Cathy O'Neil.

[Chen et al., 2015] Chen, C., Seff, A., Kornhauser, A., and Xiao, J. (2015). DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. In *DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving*, pages 2722–2730.

[Cherry, 2009] Cherry, S. R. (2009). Multimodality Imaging: Beyond PET/CT and SPECT/CT. *Seminars in Nuclear Medicine*, 39(5):348–353.

[Courant et al., 1928] Courant, R., Friedrichs, K., and Lewy, H. (1928). Ueber die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen, Bd. 100 (1928)*, pages S. 32–74.

[Cox et al., 2012] Cox, B., Laufer, J. G., Arridge, S. R., and Beard, P. C. (2012). Quantitative spectroscopic photoacoustic imaging: a review. *Journal of Biomedical Optics*, 17(6):061202.

[Cruz and Wishart, 2007] Cruz, J. A. and Wishart, D. S. (2007). Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*, 2:59–77.

[Cruz-Roa et al., 2013] Cruz-Roa, A. A., Ovalle, J. E. A., Madabhushi, A., and Osorio, F. A. G. (2013). A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013*, Lecture Notes in Computer Science, pages 403–410. Springer, Berlin, Heidelberg.

[Cuda, 2013] Cuda, N. (2013). CUDA Toolkit https://developer.nvidia.com/cuda-toolkit.

[Daniel Crevier, 1993] Daniel Crevier (1993). *AI: The Tumultuous Search for Artificial Intelligence*. NY: BasicBooks.

[Debus et al., 2012] Debus, J., Delorme, Stefan, and Jenderka Klaus-Vitold (2012). *Duale Reihe Sonografie*. Thieme.

[Evtimov et al., 2017] Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., Rahmati, A., and Song, D. (2017). Robust Physical-World Attacks on Deep Learning Models. *arXiv:1707.08945 [cs]*. arXiv: 1707.08945.

[Fenster et al., 2001] Fenster, A., Downey, D. B., and Cardinal, H. N. (2001). Three-dimensional ultrasound imaging. *Physics in Medicine & Biology*, 46(5):R67.

[Gervain and Werker, 2008] Gervain, J. and Werker, J. F. (2008). How Infant Speech Perception Contributes to Language Acquisition. *Language and Linguistics Compass*, 2(6):1149–1170.

[Guggenheim et al., 2015] Guggenheim, J. A., Allen, T. J., Plumb, A., Zhang, E. Z., Rodriguez-Justo, M., Punwani, S., and Beard, P. C. (2015). Photoacoustic imaging of human lymph nodes with endogenous lipid and hemoglobin contrast. *Journal of Biomedical Optics*, 20(5):50504.

[Hanahan and Weinberg, 2011] Hanahan, D. and Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674.

[Harrison and Zemp, 2011] Harrison, T. and Zemp, R. J. (2011). The applicability of ultrasound dynamic receive beamformers to photoacoustic imaging. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 58(10):2259–2263.

[Hauptmann et al., 2017] Hauptmann, A., Lucka, F., Betcke, M., Huynh, N., Cox, B., Beard, P., Ourselin, S., and Arridge, S. (2017). Model based learning for accelerated limited view 3d photoacoustic tomography. *arXiv:1708.09832 [cs, math]*. arXiv: 1708.09832.

[Heisenberg, 1927] Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4):172–198.

[Hell and Wichmann, 1994] Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782.

[Hordvik, 1977] Hordvik, A. (1977). Measurement techniques for small absorption coefficients: recent advances. *Applied Optics*, 16(11):2827–2833.

[Huang et al., 2013] Huang, C., Wang, K., Nie, L., Wang, L. V., and Anastasio, M. A. (2013). Full Wave Iterative Image Reconstruction in Photoacoustic Tomography With Acoustically Inhomogeneous Media. *IEEE Transactions on Medical Imaging*, 32(6):1097–1110.

[Isensee et al., 2017a] Isensee, F., Jaeger, P., Full, P. M., Wolf, I., Engelhardt, S., and Maier-Hein, K. H. (2017a). Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features. *arXiv:1707.00587 [cs]*. arXiv: 1707.00587.

[Isensee et al., 2017b] Isensee, F., Kickingereder, P., Bonekamp, D., Bendszus, M., Wick, W., Schlemmer, H.-P., and Maier-Hein, K. (2017b). Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks. In *Bildverarbeitung für die Medizin 2017*, Informatik aktuell, pages 86–91. Springer Vieweg, Berlin, Heidelberg.

[Jacobs, 1988] Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1(4):295–307.

[Jacques, 2014] Jacques, S. L. (2014). Coupling 3d Monte Carlo light transport in optically heterogeneous tissues to photoacoustic signal generation. *Photoacoustics*, 2(4):137–142.

[Jaeger et al., 2017] Jaeger, P. F., Bickelhaupt, S., Laun, F. B., Lederer, W., Heidi, D., Kuder, T. A., Paech, D., Bonekamp, D., Radbruch, A., Delorme, S., Schlemmer, H.-P., Steudle, F., and Maier-Hein, K. H. (2017). Revealing Hidden Potentials of the q-space Signal in Breast Cancer. In *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Lecture Notes in Computer Science, pages 664–671. Springer, Cham.

[Karpathy and Fei-Fei, 2014] Karpathy, A. and Fei-Fei, L. (2014). Deep Visual-Semantic Alignments for Generating Image Descriptions. *arXiv:1412.2306 [cs]*. arXiv: 1412.2306.

[Kim and Hwang, 2016] Kim, H.-E. and Hwang, S. (2016). Deconvolutional Feature Stacking for Weakly-Supervised Semantic Segmentation. *arXiv:1602.04984 [cs]*. arXiv: 1602.04984.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. arXiv: 1412.6980.

[Kirchner et al., 2017] Kirchner, T., Groehl, J., and Maier-Hein, L. (2017). Local context encoding enables machine learning-based quantitative photoacoustics. *arXiv:1706.03595 [physics]*. arXiv: 1706.03595.

[Knieling et al., 2017] Knieling, F., Neufert, C., Hartmann, A., Claussen, J., Urich, A., Egger, C., Vetter, M., Fischer, S., Pfeifer, L., Hagel, A., Kielisch, C., Görtz, R. S., Wildner, D., Engel, M., Röther, J., Uter, W., Siebler, J., Atreya, R., Rascher, W., Strobel, D., Neurath, M. F., and Waldner, M. J. (2017). Multispectral Optoacoustic Tomography for Assessment of Crohn's Disease Activity. *New England Journal of Medicine*, 376(13):1292–1294.

[Kruger, 1994] Kruger, R. A. (1994). Photoacoustic ultrasound. *Medical Physics*, 21(1):127–131.

[Ku et al., 2004] Ku, G., Wang, X., Stoica, G., and Wang, L. V. (2004). Multiple-bandwidth photoacoustic tomography. *Physics in Medicine & Biology*, 49(7):1329.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[Lim et al., 2008] Lim, H. B., Nhung, N. T. T., Li, E. P., and Thang, N. D. (2008). Confocal Microwave Imaging for Breast Cancer Detection: Delay-Multiply-and-Sum Image Reconstruction Algorithm. *IEEE Transactions on Biomedical Engineering*, 55(6):1697–1704.

[Litjens et al., 2017] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M. v. d., Ginneken, B. v., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.

[Lops et al., 2011] Lops, P., Gemmis, M. d., and Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook*, pages 73–105. Springer, Boston, MA.

[Manohar and Razansky, 2016] Manohar, S. and Razansky, D. (2016). Photoacoustics: a historical review. *Advances in Optics and Photonics*, 8(4):586–617.

[Marcus, 2018] Marcus, G. (2018). Deep Learning: A Critical Appraisal. *arXiv:1801.00631 [cs, stat]*. arXiv: 1801.00631.

[Mast et al., 2001] Mast, T. D., Souriau, L. P., Liu, D. L. D., Tabei, M., Nachman, A. I., and Waag, R. C. (2001). A k-space method for large-scale models of wave propagation in tissue. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 48(2):341–354.

[Matrone et al., 2015] Matrone, G., Savoia, A. S., Caliano, G., and Magenes, G. (2015). The Delay Multiply and Sum Beamforming Algorithm in Ultrasound B-Mode Medical Imaging. *IEEE Transactions on Medical Imaging*, 34(4):940–949.

[McCorduck, 2004] McCorduck, P. (2004). *Machines Who Think*. A K Peters, Ltd. Natick, Massachusetts.

[Merčep et al., 2015] Merčep, E., Jeng, G., Morscher, S., Li, P.-C., and Razansky, D. (2015). Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 62(9):1651–1661.

[Mie, 1908] Mie, G. (1908). Beitraege zur Optik trüber Medien, speziell kolloidaler Metalloesungen. *Annalen der Physik*, 330(3):377–445.

[Millner and Cobet, 1987] Millner, R. and Cobet, U. (1987). *Ultraschalltechnik*. Physik-Verlag.

[Montavon et al., 2017] Montavon, G., Bach, S., Binder, A., Samek, W., and Müller, K.-R. (2017). Explaining NonLinear Classification Decisions with Deep Taylor Decomposition. *Pattern Recognition*, 65:211–222. arXiv: 1512.02479.

[Mozaffarzadeh et al., 2017] Mozaffarzadeh, M., Mahloojifar, A., and Orooji, M. (2017). Medical photoacoustic beamforming using minimum variance-based delay multiply and sum. In *Medical photoacoustic beamforming using minimum variance-based delay multiply and sum*, volume 10335, page 1033522. International Society for Optics and Photonics.

[Mozaffarzadeh et al., 2018] Mozaffarzadeh, M., Mahloojifar, A., Orooji, M., Adabi, S., and Nasiriavanaki, M. (2018). Double-Stage Delay Multiply and Sum Beamforming Algorithm: Application to Linear-Array Photoacoustic Imaging. In *IEEE Transactions on Biomedical Engineering*, volume 65, pages 31–42.

[Mukamel, 1982] Mukamel, S. (1982). Collisional broadening of spectral line shapes in two-photon and multiphoton processes. *Physics Reports*, 93(1):1–60.

[Nasrabadi, 2007] Nasrabadi, N. M. (2007). Pattern Recognition and Machine Learning. *Journal of Electronic Imaging*, 16(4):049901.

[Nguyen et al., 2016] Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., and Yosinski, J. (2016). Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space. *arXiv:1612.00005 [cs]*. arXiv: 1612.00005.

[Nguyen et al., 2014] Nguyen, A., Yosinski, J., and Clune, J. (2014). Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. *arXiv:1412.1897 [cs]*. arXiv: 1412.1897.

[Niederhauser et al., 2005] Niederhauser, J. J., Jaeger, M., Lemor, R., Weber, P., and Frenz, M. (2005). Combined ultrasound and optoacoustic system for real-time high-contrast vascular imaging in vivo. *IEEE Transactions on Medical Imaging*, 24(4):436–440.

[Nobelprize, 2017] Nobelprize (2017). Nobelprize.

[Park et al., 2016] Park, J., Jeon, S., Meng, J., Song, L., Lee, J. S., and Kim, C. (2016). Delay-multiply-and-sum-based synthetic aperture focusing in photoacoustic microscopy. *Journal of Biomedical Optics*, 21(3):36010.

[Perez and Wang, 2017] Perez, L. and Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv:1712.04621 [cs]*. arXiv: 1712.04621.

[Pytorch, 2018] Pytorch (2018). PyTorch.

[Razansky et al., 2009] Razansky, D., Distel, M., Vinegoni, C., Ma, R., Perrimon, N., Koester, R. W., and Ntziachristos, V. (2009). Multispectral opto-acoustic tomography of deep-seated fluorescent proteins in vivo. *Nature Photonics*, 3(7):412–417.

[Reiter and Bell, 2017] Reiter, A. and Bell, M. A. L. (2017). A machine learning approach to identifying point source locations in photoacoustic data. In *A machine learning approach to identifying point source locations in photoacoustic data*, volume 10064, page 100643J. International Society for Optics and Photonics.

[Repacholi and Benwell, 1982] Repacholi, M. H. and Benwell, D. A. (1982). *Essentials of Medical Ultrasound: A Practical Introduction to the Principles, Techniques, and Biomedical Applications*. Medical Methods. Humana Press.

[Ribeiro et al., 2016] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv:1602.04938 [cs, stat]*. arXiv: 1602.04938.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). UNet Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. arXiv: 1505.04597.

[Ruder, 2016] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv:1609.04747 [cs]*. arXiv: 1609.04747.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

[Samek et al., 2017] Samek, W., Wiegand, T., and Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *arXiv:1708.08296 [cs, stat]*. arXiv: 1708.08296.

[Shalev-Shwartz et al., 2017] Shalev-Shwartz, S., Shamir, O., and Shammah, S. (2017). Failures of Gradient-Based Deep Learning. *arXiv:1703.07950 [cs, stat]*. arXiv: 1703.07950.

[Shen et al., 2017] Shen, D., Wu, G., and Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19(1):221–248.

[Sippel et al., 2011] Sippel, S., Muruganandan, K., Levine, A., and Shah, S. (2011). Review article Use of ultrasound in the developing world. *International Journal of Emergency Medicine*, 4:72.

[Springenberg et al., 2014] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for Simplicity: The All Convolutional Net. *arXiv:1412.6806 [cs]*. arXiv: 1412.6806.

[Sun et al., 2016] Sun, W., Zheng, B., and Qian, W. (2016). Computer aided lung cancer diagnosis with deep learning algorithms. In *Computer aided lung cancer diagnosis with deep learning algorithms*, volume 9785, page 97850Z. International Society for Optics and Photonics.

[Talukdar et al., 2013] Talukdar, Y., Avti, P., Sun, J., and Sitharaman, B. (2013). Multimodal Ultrasound-Photoacoustic Imaging of Tissue Engineering Scaffolds and Blood Oxygen Saturation In and Around the Scaffolds. *Tissue Engineering Part C: Methods*, 20(5):440–449.

[Thomenius, 1996] Thomenius, K. E. (1996). Evolution of ultrasound beamformers. In *1996 IEEE Ultrasonics Symposium. Proceedings*, volume 2, pages 1615–1622 vol.2.

[Treeby and Cox, 2010] Treeby, B. E. and Cox, B. T. (2010). k-Wave MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of Biomedical Optics*, 15(2):021314.

[Valluru and Willmann, 2016] Valluru, K. S. and Willmann, J. K. (2016). Clinical photoacoustic imaging of cancer. *Ultrasonography*, 35(4):267–280.

[Veen and Buckley, 1988] Veen, B. D. V. and Buckley, K. M. (1988). Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24.

[Vinyals et al., 2014] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2014). Show and Tell: A Neural Image Caption Generator. *arXiv:1411.4555 [cs]*. arXiv: 1411.4555.

[Waibel et al., 2018a] Waibel, D., Gröhl, J., Isensee, F., Kirchner, T., Maier-Hein, K., and Maier-Hein, L. (2018a). Reconstruction of initial pressure from limited view photoacoustic images using deep learning. In *Photons Plus Ultrasound: Imaging and Sensing 2018*, volume 10494, page 104942S. International Society for Optics and Photonics.

[Waibel et al., 2018b] Waibel, D., Gröhl, J., Isensee, F., Maier-Hein, K. H., and Maier-Hein, L. (2018b). Rekonstruktion der initialen Druckverteilung photoakustischer Bilder mit limitiertem Blickwinkel durch maschinelle Lernverfahren. In *Bildverarbeitung für die Medizin 2018*, Informatik aktuell, pages 201–201. Springer Vieweg, Berlin, Heidelberg.

[Wang et al., 2016] Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. (2016). Deep Learning for Identifying Metastatic Breast Cancer. *arXiv:1606.05718 [cs, q-bio]*. arXiv: 1606.05718.

[Wang, 2008] Wang, L. V. (2008). Prospects of photoacoustic tomography. *Medical Physics*, 35(12):5758–5767.

[Wang and Gao, 2014] Wang, L. V. and Gao, L. (2014). Photoacoustic Microscopy and Computed Tomography: From Bench to Bedside. *Annual Review of Biomedical Engineering*, 16(1):155–185.

[Wang and Hu, 2012] Wang, L. V. and Hu, S. (2012). Photoacoustic tomography: in vivo imaging from organelles to organs. *Science (New York, N.Y.)*, 335(6075):1458–1462.

[Wang and Yao, 2016] Wang, L. V. and Yao, J. (2016). A Practical Guide to Photoacoustic Tomography in the Life Sciences. *Nature methods*, 13(8):627–638.

[Wang et al., 2003] Wang, X., Pang, Y., Ku, G., Xie, X., Stoica, G., and Wang, L. V. (2003). Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain. *Nature Biotechnology*, 21(7):803–806.

[Wang et al., 2017] Wang, Y., Li, J., Lu, T., Zhang, L., Zhou, Z., Zhao, H., and Gao, F. (2017). Combined diffuse optical tomography and photoacoustic tomography for enhanced functional imaging of small animals: a methodological study on phantoms. *Applied Optics*, 56(2):303–311.

[Webb and Kagadis, 2003] Webb, A. and Kagadis, G. C. (2003). Introduction to Biomedical Imaging. *Medical Physics*, 30(8):2267–2267.

[Wei et al., 2015] Wei, C.-W., Nguyen, T.-M., Xia, J., Arnal, B., Wong, E. Y., Pelivanov, I. M., and O'Donnell, M. (2015). Real-Time Integrated Photoacoustic and Ultrasound (PAUS) Imaging System to Guide Interventional Procedures: Ex Vivo Study. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 62(2):319–328.

[Wilson et al., 2014] Wilson, K. E., Bachawal, S. V., Tian, L., and Willmann, J. K. (2014). Multiparametric Spectroscopic Photoacoustic Imaging of Breast Cancer Development in a Transgenic Mouse Model. *Theranostics*, 4(11):1062–1071.

[Wirkert et al., 2017] Wirkert, S. J., Vemuri, A. S., Kenngott, H. G., Moccia, S., Götz, M., Mayer, B. F. B., Maier-Hein, K. H., Elson, D. S., and Maier-Hein, L. (2017). Physiological Parameter Estimation from Multispectral Images Unleashed. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, Lecture Notes in Computer Science, pages 134–141. Springer, Cham.

[Wolf et al., 2005] Wolf, I., Vetter, M., Wegner, I., Böttger, T., Nolden, M., Schöbinger, M., Hastenteufel, M., Kunert, T., and Meinzer, H.-P. (2005). The Medical Imaging Interaction Toolkit. *Medical Image Analysis*, 9(6):594–604.

[Xia et al., 2014] Xia, J., Yao, J., and Wang, L. V. (2014). Photoacoustic tomography: principles and advances. *Electromagnetic waves (Cambridge, Mass.)*, 147:1–22.

[Zeiler and Fergus, 2014] Zeiler, M. D. and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014*, Lecture Notes in Computer Science, pages 818–833. Springer, Cham.

[Zhang et al., 2006] Zhang, H. F., Maslov, K., Stoica, G., and Wang, L. V. (2006). Functional photoacoustic microscopy for high-resolution and noninvasive *in vivo* imaging. *Nature Biotechnology*, 24(7):848.

[Zhao and Zhang, 2011] Zhao, H. and Zhang, L. Q. (2011). Automatic tracking of muscle fascicles in ultrasound images using localized radon transform. *IEEE Transactions on Biomedical Engineering*, 58(7):2094–2101.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 29. März 2018 .....