

Model-Based Deep Learning for Limited View Photoacoustic Computed Tomography.

Joep van de Weem*, Ben Luijten† and Ruud van Sloun‡
Signal Processing Systems, Technical University Eindhoven
Eindhoven

Abstract—Photoacoustic Computed Tomography (PACT) is an emerging field in biomedical imaging used in e.g. breast cancer detection. The imaging modality uses light to excite the object under study, which then transforms this light into acoustic waves via thermoelastic expansion. A well-known problem in PACT is its reliance on big lasers combined with large ultrasonic sensor arrays to acquire an image. Using more versatile and less expensive options like a probe with a built-in light source could contribute to making PACT a standard technique for clinical applications. However, the signal acquired from a probe is not sufficient for making a sufficiently accurate image. In this paper, we propose a joint deep learning and compressed sensing approach for finding the acoustic inverse. It achieves this by finding a learned regularization for the ill-posed inverse problem and then iteratively using this regularization to fill in data acquired from artificial sensors surrounding the subject. Results show that steps in the good direction are taken, but more focus on the neural network is still needed to make it a viable solution to the problem.

I. INTRODUCTION

Photoacoustic Computed Tomography (PACT) is a rapidly growing medical imaging modality that takes advantage of the natural optical contrast of biological tissues combined with a high ultrasonic resolution [1]. It is capable of providing high resolution images that can be used for breast [2], skin [3] and gastrointestinal [4] cancer detection and is also shown to be capable of measuring the metabolic rate of oxygen [5].

The photoacoustic data is acquired by illuminating the biological tissue using a laser. Due to the photoacoustic effect, a part of the optical energy is converted into ultrasonic emissions through the thermoelastic expansion of the tissue. [6]. This expansion causes waves that can then be received with ultrasonic transducers from which the distribution of biological tissue can be reconstructed based on its specific optical absorption. The reliance on contrast in optical absorption causes PACT to be especially effective at imaging blood vessels due to their high optical absorption [7], [8].

As summarized by Choi et al [9], PACT runs into four main practical limitations: Limited view, limited bandwidth, heterogeneous medium, and lossy medium. In this work, we mainly focus on the limited view problem, but the limited bandwidth problem is also worth addressing. The limited view problem is defined as the incapability of detecting signals outside of the ultrasound transducer aperture. This causes the need for large and expensive systems for acquiring accurate data. An example of such a system can be found in [2], where a complete circular array that is embedded in a bed is used

for imaging. Replacing this with a probe similar to the ones used in ultrasound could contribute to make photoacoustics a standard technique for clinical applications such as breast tumour detection, melanoma imaging, and rheumatoid arthritis [10].

The limited bandwidth problem is caused by the bandlimited ultrasonic transducers used to acquire the photoacoustic signal. Transducers with wider bandwidth like the CMUT [11] and Fabry-Perot transducer [12] exist but are often less stable and more expensive [13] than piezoelectric transducers. The instability together with the better penetration depth at lower frequencies causes the piezoelectric transducers to still be a viable and cheaper option.

Recently, reconstruction methods that exploit prior information about the underlying physics model have gained popularity within the field of biomedical imaging. Such compressed sensing techniques are showing results in resolving the limited view and bandwidth problem in PACT. The Iterative Soft-Thresholding Algorithm (ISTA) [14] or Partial Known Support (PKS) [15] algorithm already show great results mainly focusing on the limited bandwidth problem. In seismography, a commonly used algorithm is the Projection onto Convex Sets (POCS) algorithm [16]. This algorithm enforces data consistency while it iteratively interpolates data from unknown traces. In the context of PACT, this algorithm can be used to fill in data acquired from artificial sensors surrounding the subject as compared to only the data acquired using a probe.

These compressed sensing techniques rely heavily on a robust model for reconstruction. However, the photoacoustic inverse problem is ill-posed, causing the need for estimates of this inverse. Common modalities for this are Filtered Backprojection (FBP) [17], beamforming, and time-reversal algorithms. The forward operation is often calculated using K-wave [18] for simulations. Computational time constraints lead to the need for another modality than K-wave due to its computational complexity on larger images. This paper proposes the use of a faster convolution-based algorithm first proposed by Gao et al. [19] for ultrasound imaging.

To counter the ill-posedness of the model, prior knowledge of the input and output can be used to accurately obtain better regularization using Maximum-a-posteriori (MAP) estimation. The prior knowledge of the system is that it is based on data from a limited view as well as limited bandwidth. However, these concepts are often difficult to describe analytically. This is where learned regularization can be used [20]. In

learned regularization, deep learning is used to learn the prior knowledge of the system and regularize the model by applying it.

The fast growing field of deep learning is showing state of the art results in the improvement of contrast e.g. MR [21], Ultrasound [22] and CT [23] imaging. The usage of Deep Learning in PACT is still in its infancy [24]. Current approaches mainly consist of bandwidth extension networks and artifact removal [24], but recently have shifted more to a model-based approach [25]. The limited view problem has also been tackled using deep learning in both model-based [26] and end-to-end approaches [27].

The reconstruction method proposed in this paper tries to use the POCS algorithm with a learned regularization function to iteratively find the underlying source distribution. By using the POCS algorithm, we try to extend the limited view of a probe to that of an entire circular array, effectively cancelling the limited view problem. The learned regularization function tries to remove the reconstruction artefacts and extend the bandwidth.

To train this network, we opt for a supervised approach. Due to the infancy of PACT and the difficulty of getting labelled training data in medical imaging, simulated data is used for training. After training a more qualitative approach is taken to assess how the proposed method compares to other reconstruction methods on *in vivo* images.

In this paper $\|x\|_p$ denotes the p -norm of x , x^* denotes the Hermitian transpose of x , and \mathcal{F} denotes the Fourier transform.

We structure the paper as follows: In section II we describe compressed sensing concepts as well as the measurement model. We describe our method of reconstruction in section III. In section IV we describe the acquisition of the data for training the neural network. After this, we describe the result acquired using the network in section V followed by a conclusion and discussion in sections VII and VII respectively.

II. BACKGROUND

A. Photoacoustics

The goal of PACT is to visualise the contrast of optical absorption within the object under study. PACT achieves this by first emitting light onto the object and subsequently receiving an ultrasonic signal. In research by Cox et al. [8] it is shown that this process can be divided into three different processes:

- 1) *Optical process*: The propagation of light and absorption in its tissue.
- 2) *Thermoelastic Coupling process*: The thermoelastic expansion and formation of the initial pressure.
- 3) *Acoustic Process*: The propagation of ultrasound waves and their detection.

Out of these three processes, this paper will mainly focus on the forward and backward problems of the Acoustic Process. The other two are described briefly below. A more detailed description is given in [8], [28].

1) *Optical process*: PACT starts with the emittance of light. Often, this is done by a short-pulsed nanosecond laser source transmitting in the near-infrared range (690-1000 nm) to obtain the most contrast for blood vessels [8].

The transport of light is described by the radiative transfer equation (RTE) [28]:

$$\theta(\nabla_r + \mu_a(r) + \mu_s(r))\phi(r, \theta) - \mu_s(r) \int_{4\pi} k(\theta, \hat{\theta})\phi(r, \hat{\theta})d\hat{\theta} = q(r, \theta) \quad (1)$$

Where ϕ is the light radiance at location r moving in direction θ , μ_a is the absorption coefficient, μ_s is the scattering coefficient and $k(\theta, \hat{\theta})$ is the function describing the probability of a photon travelling in direction θ scattering to $\hat{\theta}$. numerically this equation is often implemented using finite element methods or Monte Carlo simulations.

When using biological tissues, the radiative transfer equation in tissue is usually described with four parameters. The optical absorption μ_a , scattering coefficient μ_s , scattering anisotropy g and refractive index n . From these g describes the amount of forward direction retained after a scattering event, and n describes how fast light travels through the medium as compared to a vacuum.

For the thermoelastic coupling, the absorbed energy H is needed. This can be calculated using

$$H(r) = \mu_a(r)\Phi(r) = \mu_a(r) \int_{4\pi} \phi(r, \theta)d\theta \quad (2)$$

where $\Phi(r)$ denotes the fluence rate at r .

2) *Thermoelastic coupling process*: Thermoelastic coupling or photoacoustic efficiency describes the process of transferring the absorbed energy to acoustic pressure waves. It is based on the following equation:

$$p_0(r) = \hat{\Gamma}(r)H(r) \quad (3)$$

where $\hat{\Gamma}$ is the photoacoustic efficiency, which depends on the thermal coefficients of the local tissue. Often this parameter is not known, and therefore it is not taken into account in this paper. In the rest of this paper, we assume that $\Gamma = 1$.

B. Compressed Sensing

The problem of estimating an x given measured data y from

$$y = \Phi x + n \quad (4)$$

where $\Phi \in \mathbb{R}^{N_y \times N_x}$ the transformation matrix, which in our case would be the transformation from the initial pressure (p_0) to pressure over time at the sensors ($p(r_0, t)$), and n denoting the noise. The modality of compressed sensing tries to solve this problem using only part of the data, effectively assuming that the acquired data is a compressed version of the actual data, leading to the equation:

$$y = A\Phi x + n \quad (5)$$

where A is a subsampling matrix that samples certain points of x . In the case of PACT, this would point to the acquired data

channels. The ill-posedness of the model causes the solution (x) to be non-unique and sensitive to changes in the input (y). Using prior knowledge of the model could regularize this function. This makes the Maximum-A-Posteriori (MAP) estimate a strong method for finding the original object under study based on this prior knowledge.

The MAP estimation is based on finding the ideal forward measuring model ($p(x|y)$). This forward measuring model is the best fit for the reconstructed image x given the measured data y . such a MAP estimate is described as:

$$\hat{x} = \arg \max_x p(x|y) = \arg \max_x p(y|x)p(x) \quad (6)$$

The likelihood is described as normally distributed ($p(y|x) \sim \mathcal{N}(x, \hat{\sigma})$). Using this, the function can further be derived using $\arg \max(ab) = \arg \min(-\log(a) - \log(b))$ leading to:

$$\hat{y} = \arg \min_y \left(\frac{1}{2} \|x - A\Phi y\|_2^2 + \lambda R(y) \right) \quad (7)$$

here λ is a regularization parameter, and $R(\cdot)$ is a regularization function based on the prior knowledge of the system. The goal of finding the actual true underlying intensity map relies heavily on correct assumptions of the underlying physics model. in other words, a sufficiently accurate estimate of the backward problem, and an accurate choice of the regularizer.

When reconstructing a signal using compressed sensing it is important that the measurement matrix A is incoherent with the transform Ψ . In traditional compressed sensing uses a random measurement matrix to enforce this incoherence. [] In the case of PACT, this random sampling would lead to random sampling in the sensor direction, leading to an array with randomly placed sensors. Preferably, we opt for a more practical application of a probe, which only uses part of a circular array. This would lead to a more coherent measurement matrix, and thus a decrease in accuracy in the reconstruction.

Filling in the data to form a complete circular array around the subject could lead to a more complete view of the object under study. In the fields of geophysics [16], and Magnetic Resonance [29] POCS is commonly used algorithm for solving the problem of incomplete data.

The POCS algorithm iteratively reconstructs missing sensor data by including prior information about the underlying object under study x . It thereby iterates between two steps: 1) reconstruction of a visually plausible (prior-consistent) solution for x from measurements y , and 2) projection of that solution onto the set of physically feasible answers by making them measurement-consistent as can be seen in figure 1. These two steps can be described mathematically as 1):

$$\hat{h}_n = \Phi P(\Phi^{-1} \hat{x}_n) \quad (8)$$

and 2):

$$\hat{x}_{n+1} = \begin{cases} x & \text{if } a \in A \\ \hat{h}_n & \text{else} \end{cases} \quad (9)$$

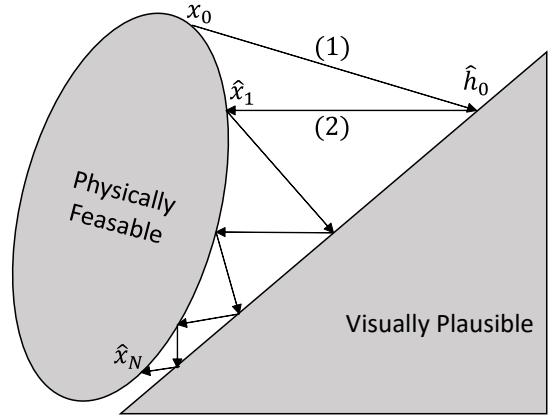


Fig. 1: POCS algorithm searching for the optimal image in the physically feasible set where (1) transforms the physically feasible estimate to a visually plausible (prior-consistent) solution and (2) maps it back to a physically feasible solution.

Where $P(\cdot)$ is a proximal step that is taken to minimize the regularization function $R(\cdot)$, and \hat{h}_n describes the visually plausible image.

In the case of PACT, the model can be described as forward ($\Psi : \mathcal{Y} \rightarrow \mathcal{X}$) and backward ($\Theta : \mathcal{X} \rightarrow \mathcal{Y}$) transforms that transform data from the sensor domain \mathcal{Y} to the image domain \mathcal{X} and vice versa. The POCS algorithm relies heavily on a good model for accurate reconstruction, this will further be described in chapters II-C and II-D. Implementing these transforms in the POCS algorithm leads to:

$$\hat{x}_{n+1} = \begin{cases} x & \text{if } s \in S \\ \Theta P(\Psi \hat{x}_n) & \text{if } s \notin S \end{cases} \quad (10)$$

where S is the real sensor array.

C. Forward Model

The goal for the forward model is to model a function that maps data from the image domain \mathcal{X} to the sensor domain \mathcal{Y} . Here we assume that the speed of sound (c_0) and thermal distribution are homogeneous everywhere. The acoustic pressure $p(r_0, t) \in \mathcal{X}$ can be derived from the photoacoustic wave equation using Green's function leading to:

$$p(r_0, t) = \frac{\beta I_0}{4\pi C_p} \iint_V \frac{A(r)}{\|r_0 - r\|} dr \quad (11)$$

where β is the isobaric volume expansion coefficient, C_p is the specific heat, V is the surface, I_0 is the thermal profile, and $A(r) \in \mathcal{Y}$ is the absorbed energy deposition. From here on we will refer to $p(r_0, t)$ as $x \in \mathcal{X}$ and $A(r)$ as $y \in \mathcal{Y}$. Then, the exact forward problem is described as $y = \Theta x$.

A common method for this transformation is simulating the image using K-wave [18], an open-source Matlab toolbox for numerically solving the photoacoustic wave equation. However, this is often a timely process for larger-scale images. Alternatively, a method that originates from ultrasound is the COLE algorithm by Gao et al. [19]. This convolution-based

algorithm assumes homogeneity in the speed of sound, density, and heat distribution for a faster mapping from space to time. It achieves this using:

$$I_s[y] = \sum_{g=1}^G a_{s,g} w_{s,g} y_{s,g} \delta[\Delta T(r_s, r_g)] \quad (12)$$

where $I_s[t]$ is the mapping from image to sensor domain if we assume that every point in the image is a delta function, $a_{s,g}$ is the apodization used in the backward pass, $w_{s,g}$ is the weighing that is based on distance, and T describes the delay between sensor and grid point using:

$$T(r_s, r_g) = \frac{\|r_s - r_g\|_2}{c} + t_0 \quad (13)$$

Where $T(r_s, r_g)$ propagation time from each sensor at r_s to a grid point at r_g , c is the speed of sound in the medium and t_0 is a fixed delay.

To mitigate aliasing, both linearly upsampling the data in the image domain to create a denser grid, and by adding noise to the grid points to add uncorrelated spatial aliasing to the forward transform.

According to the algorithm the function is then also convolved with the point spread function (PSF) for accurate reconstruction leading to:

$$y = I \circledast F \quad (14)$$

The filter F can be described by the transfer between the input (I) and output (y) of the system.

1) *Filter*: The input (I) in this model can be described as the estimated frequency components in the input. By generating sufficient vascular structures according to section IV, we can estimate the frequency components using a Monte Carlo simulation.

$$\hat{F}_I = \frac{1}{N_{sim}} \sum_{n=1}^{N_{sim}} |\mathcal{F}(\Theta(y_{sim}))| \quad (15)$$

where N_{sim} denotes the number of simulated images, and y_{sim} the generated vascular structures.

The desired response for the output can also be estimated by averaging the frequency components of the output of the system using the in vivo data:

$$\hat{F}_y = \frac{1}{N_{invivo}} \sum_{n=1}^{N_{invivo}} |\mathcal{F}(x_n)| \quad (16)$$

where N_{invivo} denotes the amount of in vivo data and x_{invivo} denotes the acquired in vivo data. Leading to an estimate of F :

$$\hat{F} = \frac{\hat{F}_y}{\hat{F}_I} \quad (17)$$

D. Backward model

Due to the ill-posedness of the forward problem, the inverse does not exist. The reasons for the ill-posedness are described in detail in [28] and are briefly summarized here:

- 1) *noisy measurements*: A true mapping from the sensor to image domain does not exist due to the presence of noise.
- 2) *inaccurate knowledge of the forward process*: inaccuracy can also come from the simplification or lack of knowledge from the forward process, an example of that would be the assumption that the speed of sound is constant.
- 3) *limited measurements*: This is a big part of the problem in photoacoustics. Due to sensor size constraints, aperture size constraints, and acquisition bandwidth, we have incomplete data.

Leading to the need for an estimate of the inverse of Θ . A common modality used for this backward problem is Beamforming. Beamforming is a technique that is used to acquire signals from a specific direction using multiple sensors. To reconstruct an image, we want to know the value at each pixel according to each sensor in the array. A common and robust way to achieve this is by using Delay And Sum (DAS) beamforming.

Using DAS, the data is first delayed by applying a time-delay that can be calculated using (13). This delay is then used for reconstruction using:

$$x_g = \sum_{s=1}^S w_{s,g} y_g \delta[T(r_s, r_g)] \quad (18)$$

where $w_{s,g}$ is the apodization for sensor s and grid point g . The weight is in the simplest cases equal to a rectangular window or a more refined shape like a Hanning window. Adaptive apodization methods that use the statistics of an image are often used to improve this.

Another method that can be used to improve this is by interpolating the sensor data y in the sensor direction to counteract the spatiotemporal aliasing [30].

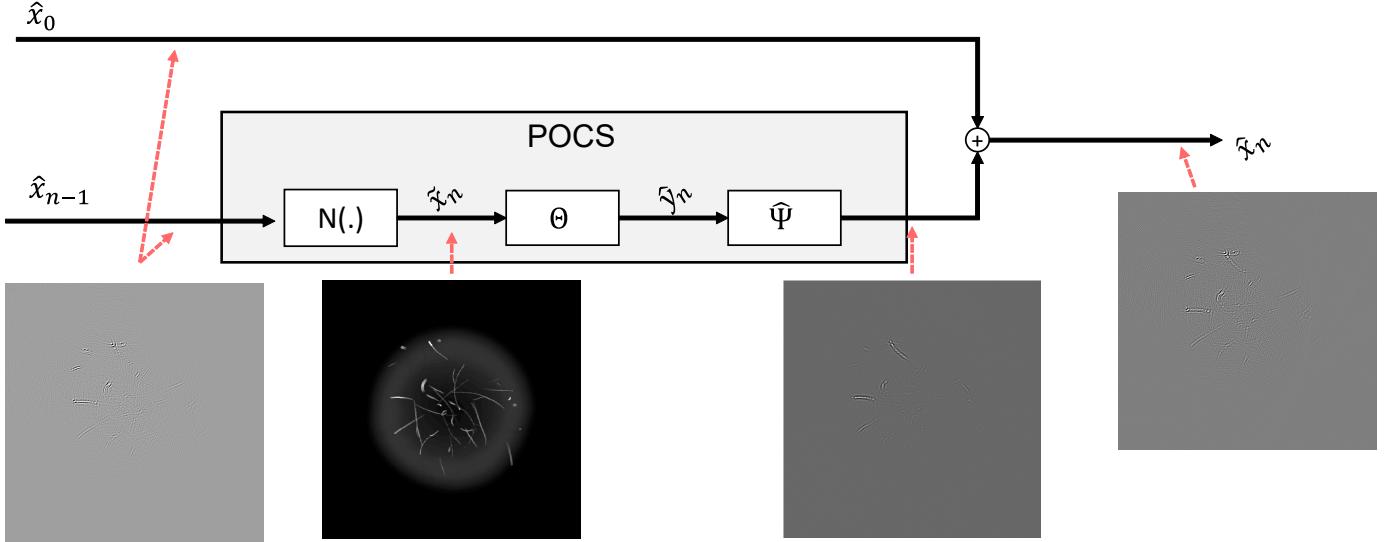
The backward operator $\hat{\Psi}$ can thus be defined as:

$$y = \hat{\Psi}_s x \rightarrow y[r_g] = \sum_{s=1}^S w_{s,g} y_s [T(r_s, r_g)] \quad (19)$$

The fact that an estimation for the inverse of Θ is used leads to the POCS step being less accurate. The artefacts caused by this inaccuracy will need to be compensated in the regularization term.

III. METHODS

In the previous sections the POCS algorithm that iteratively tries to solve our ill-posed problem using prior knowledge of the system. As equation (7) and (8) state, the regularization and proximal mapping are derived from the prior knowledge of the system. Conceptually, we know that the ill-posedness stems from the limited number of noisy measurements and the used forward and backward model. This knowledge is very difficult to describe mathematically. What we propose is that



(a) single iteration of the neural network, with the example images showing the first iteration.

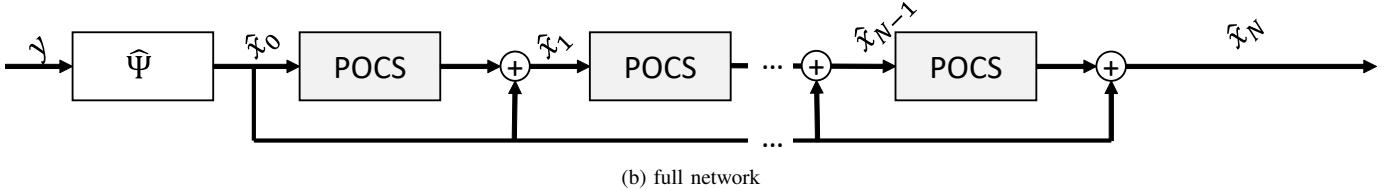


Fig. 2: the network structure where (a) shows a single iteration of the network with images from the first iteration as example. (b) shows the full iterative structure of the network.

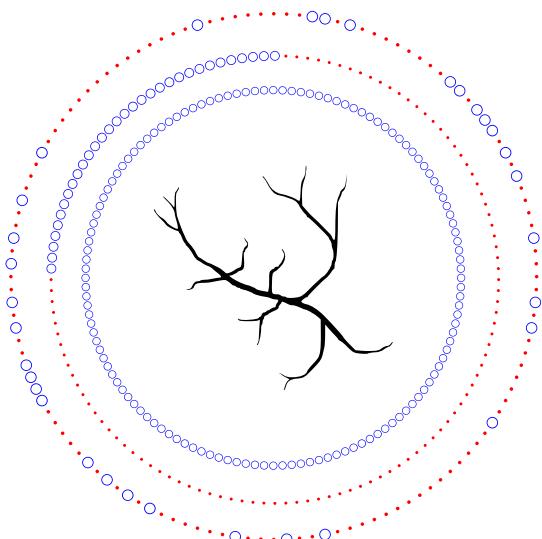


Fig. 3: The various array structures used for reconstruction. The innermost circle shows the full array, the middle shows a partial array and the outermost circle shows a random subset of the array.

the prior knowledge is learned by a neural network and applied as proximal mapping in a learned proximal step.

In equation (7) and (8) we see that the proximal step is a function that is derived from prior knowledge. Learning this step would mean that the prior knowledge is included in the final model.

A. Learned Proximal Step

The ideal proximal step should be able to extend the bandwidth of the image, recover data lost due to the limited view and compensate for the inaccuracies of the initial estimate of the inverse ($\hat{\Psi}$). Previous research shows that neural networks are very strong in this field [20], [27], [28], [33]. Implementing a neural network ($N(\cdot)$) as a proximal mapping ($P(\cdot)$) in equation (10) leads to:

$$\hat{y}_n = \begin{cases} y & \text{if } s \in S \\ \Theta N(\hat{\Psi} \hat{y}_{n-1}) & \text{if } s \notin S \end{cases} \quad (20)$$

Since DAS beamforming sums in the sensor direction, we can further simplify this to:

$$\hat{x}_{n+1} = \hat{\Psi}_s y|_{s \in S_{real}} + \hat{\Psi}_s \Theta_s N(\hat{x}_n)|_{s \in S_{art}} \quad (21)$$

S_{real} are the real sensors and S_{art} are the artificially reconstructed sensors. This network structure will from here on out be referred to as the POCSNet. Schematically, this network is shown in 2.

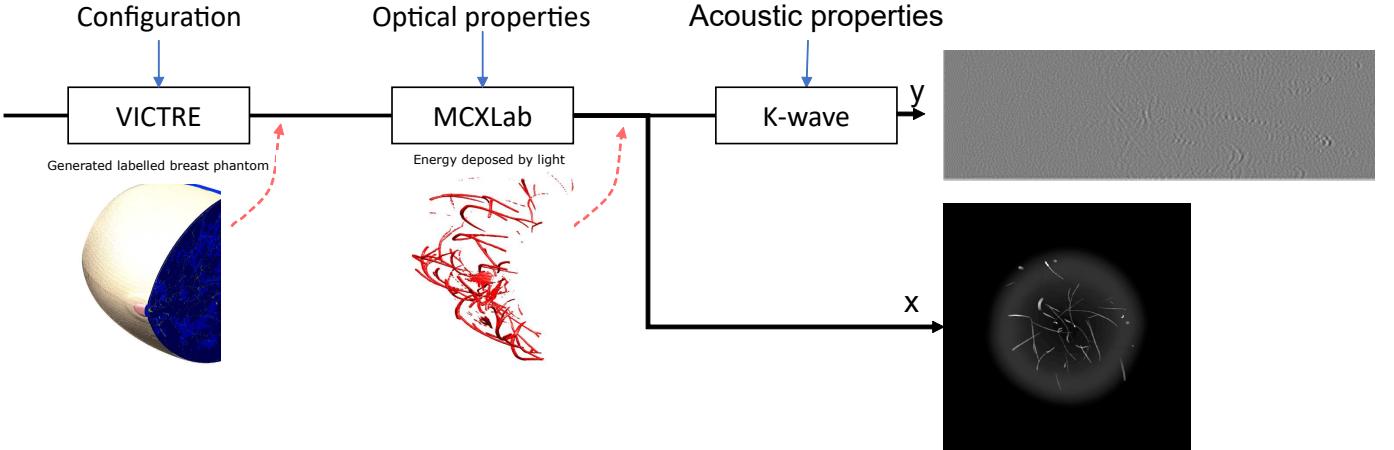


Fig. 4: simulation pipeline for generating training data pair x (ground truth) and y (generated acoustic pressure) using VICTRE [31] for phantom generation, MCXLab [32] for photon simulation, and K-wave [18] for acoustic simulations

B. Neural network

Previous research shows that The UNET [34] is a popular architecture that is proven to be strong in biomedical signal reconstruction. In PACT it is shown to extend the bandwidth [28], remove limited view artefacts [33] and even extend the aperture [27]. As such, we opt for a UNET structure as the learned proximal mapping.

C. Arrays

The network is evaluated with two different types of arrays, these can be seen in Figure 3. The first is a randomly sampled set of elements from the original circular array. This approach shows the effectiveness of the POCS algorithm in its intended purpose in compressed sensing having a more incoherent sampling matrix.

The second array design is a quarter circular part of the original circular array. This shows the effectiveness of the network at filling in the entire image from data acquired from a probe. This approach has a more coherent sampling matrix since the data is only acquired from one side.

In this network, the real sensor array consists of 128 sensors and 384 sensors are artificially reconstructed leading to a total of 512 sensors surrounding the object under study.

D. Training

Due to memory constraints, the network is trained greedily. Each proximal step is trained to find the simulated intensity map based on the output of the previous iteration. In total 5 iterations are used.

The network is implemented using Pytorch [35] and trained using an Adam optimizer with a learning rate of $\mu = 0.001$. For the loss function, the mean squared error loss. The network is trained on simulated images with input being $(S_{real} \times N_t = 128 \times 4000)$ with labels of size $856 \times 856 (N_x \times N_y)$, and tested on 100 test images to assess performance during training. Each proximal layer is trained for 15 epochs.

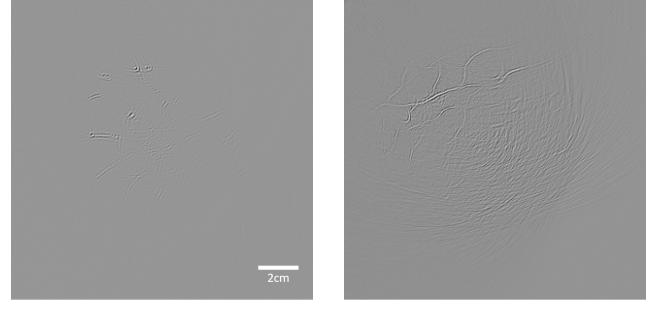


Fig. 5: comparison between simulated (a) and in vivo (b) data

IV. DATA GENERATION AND ACQUISITION

Deep learning methods rely strongly on sufficient and realistic training data. However, due to the nature of the problem, it is difficult to obtain pairs of in vivo measurements and ground truth data. As such, we opted for a training strategy relying on simulations and testing the performance using in vivo data.

A. Simulated

To give an accurate representation of the in vivo data, multiple steps are taken to ensure an accurate model that simulates the acquisition of in vivo data, several steps are taken. These steps are described below and can be seen in 5.

1) Object under study: Digital Breast Phantoms generated using the code in [40] are used to generate simulated photoacoustic data. This method is proven to be an effective method for photoacoustics as shown in [41]. An example of such a phantom is shown in figure 5.

2) Light Distribution : The state of the art for light absorption simulation is currently using a Monte Carlo simulation [42]. In such a Monte Carlo simulation, the path of billions of photons going through the tissue are simulated and the energy deposited in the tissue can be found. In this paper the GPU-accelerated Monte Carlo program MCXLab [32] is used.

To accurately simulate the path of the variables through the tissue, the optical attenuation μ_a , scattering coefficient μ_s , scattering anisotropy factor g and refractive factor n . These parameters for the various tissues found in the breast can be found in table I. The light distribution is then simulated for 10^9 photons. An example of the deposited energy distribution is shown in figure 5.

3) *Acoustic properties:* Since the light simulated in the previous chapter is transformed into an acoustic wave, the acoustic properties of the tissues also need to be taken into account. These properties are the acoustic attenuation α , speed of sound c , density ρ and measure of nonlinearity B/A . The properties for the different tissues as well as their sources can be found in table I.

These properties, the deposited energy from IV-A2 and the labelled breast map from IV-A1 are sliced in the z-direction (the direction ranging from the chest to the nipple) into 32 images each and acoustically simulated using K-wave [18].

4) *Speckle:* Speckle in photoacoustics is caused by having a non-uniformly distributed number of scatterers. K-wave, however, uses a uniform grid for simulations. To achieve the non-uniformity, noise is added to the sound speed and density maps, effectively moving the scatterers randomly.

5) *Filter response:* The same filter as described in section II-C1 can be used but then without the point spread function caused by K-wave, Leading to:

$$F_{sim} = \frac{F}{F_{PSF}} \quad (22)$$

where F_{PSF} refers to the point spread function used in K-wave that can be acquired by simulating a single point in space.

6) *Size:* For grid spacing in simulation, a grid size of 1536x1536x32 (Nx, Ny, Nz) is used due to its sufficient spatial frequency and its maximum prime factor being low, which is preferable for K-wave.

7) *Ground Truths:* A Gaussian window is applied to the intensity maps in the Z-direction and summed to get the 2-dimensional label for training.

An example of the simulated images as compared to in vivo images can be seen in figure 4.

A total of 45 breast phantoms are generated using slightly randomized configurations, these are split up into 600 for training and 120 for testing. To increase the amount of data, each of the training and testing images is rotated eight times, leading to 4800 for training and 960 for testing.

Tissue	Optical Properties ($\lambda = 1064\text{nm}$)				Acoustic Properties			
	μ_a mm^{-1}	μ_s mm^{-1}	g	n	c m/s	ρ kg/L	α Db/cm/MHz	B/A
Water	0.014	0	1	1.33	1482.3	994	0.0022	4.96
fat	0.0054	22.15	0.975	1.4	1440	911	0.3783	10.07
skin	0.0028	15.72	0.8	1.4	1624	1109	1.8365	n.a.
blood	0.45	47.69	0.98	1.4	1578	1050	0.2055	6.11
glandular	0.005	14.87	0.955	1.4	1505	1041	0.7495	9.63

TABLE I: Tissue properties used for the simulation of the light and acoustic properties of the tissues based on [36]–[39]

Network structure	Array Type	Iter.	SSIM ($\mu(\sigma)$)	PSNR ($\mu(\sigma)$)
Stacked UNET	Random Array	1	0.9477 (0.0435)	50.49 (8.12)
		2	0.9293 (0.0571)	38.73 (18.00)
		5	0.9091 (0.0764)	43.19 (13.23)
	Quarter Array	1	0.9215 (0.0461)	47.77 (5.82)
		2	0.9175 (0.0552)	42.69 (11.2)
		5	0.8527 (0.0350)	41.22 (5.79)
POCSNet	Random Array	1	0.9443 (0.0275)	50.06 (6.13)
		2	0.9351 (0.0523)	48.89 (6.51)
		5	0.9323 (0.0506)	48.60 (13.23)
	Quarter Array	1	0.9215 (0.0461)	47.77 (5.82)
		2	0.9397 (0.0485)	50.67 (8.86)
		5	0.9360 (0.0351)	48.37 (5.35)

TABLE II: the results for each of the network with each of the array structures after various iterations. The numerical results are denoted as: mean (standard deviation) or $\mu(\sigma)$.

B. In Vivo

In this case, PACT is used for angiography in the breast using a circular ultrasonic receiver array like the one described in [2]. In this acquisition, a full array of 512 ultrasonic transducers with a central frequency of 2.25MHz and a bandwidth of 90% are used.

V. RESULTS

The results are described using the two different array structures and two different network structures: The POCSnet proposed in chapter III, and a stacked UNET approach to view the results of a similar network without data consistency. In this stacked UNET approach five UNETs are connected sequentially and greedily trained in the same manner as described in chapter III-D.

The networks are evaluated on their capability of improving the image after the first iteration since the output of both the POCSnet and stacked UNET at iteration one uses the same network structure.

Next to the visual evaluation, they are evaluated using the relevant metrics: Structural Similarity Index (SSIM) and the Peak Signal to Noise Ratio (PSNR):

$$SSIM(x, y) = \frac{(2 * \mu_x * \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (23)$$

where μ_x denotes the mean of x, and σ_x^2 denotes the variance of x and σ_{xy} denotes the covariance between x and y.

$$PSNR(x, y) = 20\log_{10}\left(\frac{\max(x)}{\sqrt{MSE(x, y)}}\right) \quad (24)$$

TABLE I: Tissue properties used for the simulation of the light and acoustic properties of the tissues based on [36]–[39]

where $MSE(x, y)$ denotes the mean squared error between x and y .

The reconstructed image will be evaluated in the Fourier domain to see if any information is added in the directions perpendicular to the sensor in the partial array case.

In addition, it is also evaluated using a directionality histogram as described in [43] which shows the preferred orientation of the data by placing it into bins between -90 and 90 degrees.

In figure 6, it can be seen that the initially received data (x_0) in (a, e), is partially reconstructed in the following iterations in (b-d, f-h) for the simulated data. The simulated data shows improvements in specifically the lower spatial frequency ranges.

Figure 6b also shows that the network is capable of reconstructing vessels that are perpendicular to the sensor array. The vessel marked by the red arrow in 6b is not visible in the initial input data shown in 6a.

In figures 6(i - p) it can be seen that this technique also works on in vivo data, as the spatial Fourier domain images show more data in the perpendicular directions. This is further seen in the directionality graph in 6q, where the initial input shows no information between -80 and -20 degrees, while the reconstructed images do.

The results also show that the network has the most improvement in the initial step. This makes the initial guess that the network does in the first iteration very important. This counters the expectation that the network would iteratively make a prior-consistent guess and then "check" that guess with the measured data since the neural artefacts get more reinforced into the image instead of removed. This can especially be seen in the in vivo results where the first iteration 6j marked by the red arrow shows artefacts in the further field as circular artefacts. In the subsequent iterations (k, l) the artefacts are more and more defined.

The result of the neural proximal step can be seen in figure 7 divided in results from the POCSNet (a-h) and stacked UNET (i-p). These results show the output of networks tested on a partial array (a-d, i-j) and a random array (e-h, m-p). numerical results can be found in (s). The actual acoustic p_0 map can be seen in (q), and a reconstruction of the in vivo data using a full-array can be found in (r).

When comparing the results of the partial array from the POCSNet (a-d) to those of the Stacked UNET (i-l), it can be seen that the proposed POCSNet shows few improvements when more iterations are used, as the output on the simulated results in (a) is indistinguishable from that of (b) except for the vessel marked with the red arrow. However, this vessel cannot be seen in (q), meaning that it is a neural artifact that is not removed by the iterations. In the in vivo results of (c, d) it can even be seen that the first iteration shows better results than the fifth, as more vessels that can be seen in (r) are removed.

The results for the Random array show that the more traditional compressed sensing approach shows similar results than the partial array. The POCSnet shows in 7(e-h) that the network does not improve the data after the initial guess, and

instead further cements the artefacts into the image. The results of the stacked UNET show similar results, as there is virtually no difference between 7(m) and (n).

looking at the numerical results in table II we can see that the network does show slight improvements in the reconstruction using the POCSNet as compared to the stacked UNET. The stacked UNET shows in both the random and quarter array case a decreasing quality in the reconstruction over the iterations using both metrics. The POCSNet shows improvements as compared to the first iteration.

From figure 6 and 7 it can be concluded that using the data consistency proposed in equation 21 is beneficial to the image. As in figure 6(b) we can see vessels that were perpendicular to the sensor array, like the one marked by a red arrow. In this image, the vessel marked by white is barely visible, and could be recognised as an artifact.

To conclude, the results in figure 6 show that the network is capable of filling in data from different directions after one iteration. In addition, numerical results show an improvement in using the POCSNet as compared to the stacked UNET, as using more iterations show an increase in both metrics used for the partial array case. However ,it also adds artefacts. In the subsequent layers, it can be seen that these artefacts are further reinforced into the final image. This conclusion is further corroborated in figure 7 where the POCSNet does not show improvements between the first and fifth iteration, a result that is similar to that of the stacked UNET.

VI. DISCUSSION

These results show promising steps in the good direction with the numerical results and the application of the data consistency steps. However, in reconstructions of the neural network, it can be seen that it fails to filter out the neural artefacts and instead, further reinforces these in the image. In this chapter, we first focus on what works in the current state of the network. After which we look at problems and potential solutions.

Initial tests show that, when no prior knowledge is available of the tissue of the object under study, the COLE algorithm for the forward pass shows similar results to the K-wave simulations, while being up to 80 times quicker. Meaning that this implementation of the forward pass could also be used with other model-based deep learning PACT reconstruction approaches, leading to a significant decrease in computation times. Using this together with the network proposed in [26] should give promising results.

Next to that a new image generating pipeline is proposed in this paper that generates very accurate images while paying attention to the entire physical process of generating an image, as well as using an accurate breast phantom generator.

The results show that the image in 6b have more information than that of 6a. Which should also mean that the proximal mapping in the subsequent layer should be able to get more information. This is one of the main problems in the reconstruction, as it is shown in the results of 7(a-h) that the POCSNet is not. Adding some kind of data consistent

loss based on the reconstruction, or using a different kind of network structures like in [44] and [45] could lead to a better result.

One of the main problems with the proposed structure is the sampling matrix and transformation matrix are not incoherent enough. This would mean that an accurate representation for the data cannot be found.

For the training data, it is stated that the generator still needs data to better tune the parameters [40]. Leading to slightly inaccurate breast maps.

Another part of the generation that is approximated in this paper, is the noise. In Figure 4, it can be seen that the noise in the in vivo data differ from the noise in the simulated data. in the simulated data, white Gaussian noise was assumed, however the in vivo data shows more coherent noise. Leading to the need for a more accurate noise model.

VII. CONCLUSION

To enhance the reconstruction process of PACT, this paper proposes a model based approach that uses a compressed sensing algorithm to reconstruct the image using learned prior knowledge of the system. The system iterates between a visually feasible set of data based on the prior knowledge and projects that back onto the physically plausible set. It achieves this by using learned regularization in which a neural network is used to learn the prior knowledge of the system based on simulated data. This approach is then tested on both the simulated and in vivo data and compared to a computationally similar approach of stacked neural networks.

These tests show that the proposed neural network shows promise when looking at the numerical results when compared to the stacked neural network, together with an increase in information when using the initial data-consistency step. However the subsequent iterations show that the neural artefacts generated by the first iteration are not removed, but instead further reinforced into the output of the network. This leads to the conclusion that, while steps are taken in the positive direction, the network still needs work to data-consistently reconstruct limited view photoacoustic images.

REFERENCES

- [1] L. V. Wang, "Tutorial on Photoacoustic Microscopy and Computed Tomography," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, pp. 171–179, Jan. 2008. Conference Name: IEEE Journal of Selected Topics in Quantum Electronics.
- [2] L. Lin, P. Hu, J. Shi, C. M. Appleton, K. Maslov, L. Li, R. Zhang, and L. V. Wang, "Single-breath-hold photoacoustic computed tomography of the breast," *Nature Communications*, vol. 9, no. 1, p. 2352, 2018.
- [3] U. Dahlstrand, R. Sheikh, A. Merdasa, R. Chakari, B. Persson, M. Cinthio, T. Erlöv, B. Gesslein, and M. Malmsjö, "Photoacoustic imaging for three-dimensional visualization and delineation of basal cell carcinoma in patients," *Photoacoustics*, vol. 18, May 2020.
- [4] L. V. Wang, "Prospects of photoacoustic tomography," *Medical Physics*, vol. 35, no. 12, pp. 5758–5767, 2008. _eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1111/1.3013698>.
- [5] X. Wang, X. Xie, G. Ku, L. V. Wang, and G. S. D.v.m., "Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography," *Journal of Biomedical Optics*, vol. 11, p. 024015, Mar. 2006. Publisher: International Society for Optics and Photonics.
- [6] J. Yao and L. V. Wang, "Photoacoustic tomography: fundamentals, advances and prospects," *Contrast media & molecular imaging*, vol. 6, pp. 332–345, Sept. 2011.
- [7] J. Yao, J. Xia, K. I. Maslov, M. Nasirianaki, V. Tsytarev, A. V. Demchenko, and L. V. Wang, "Noninvasive photoacoustic computed tomography of mouse brain metabolism in vivo," *NeuroImage*, vol. 64, pp. 257–266, Jan. 2013.
- [8] B. T. Cox, J. G. Laufer, and P. C. Beard, "The challenges for quantitative photoacoustic imaging," in *Photons Plus Ultrasound: Imaging and Sensing 2009*, vol. 7177, p. 717713, International Society for Optics and Photonics, Feb. 2009.
- [9] W. Choi, D. Oh, and C. Kim, "Practical photoacoustic tomography: Realistic limitations and technical solutions," *Journal of Applied Physics*, vol. 127, p. 230903, June 2020. Publisher: American Institute of Physics.
- [10] K. Daoudi, P. J. v. d. Berg, O. Rabot, A. Kohl, S. Tisserand, P. Brands, and W. Steenbergen, "Handheld probe integrating laser diode and ultrasound transducer array for ultrasound/photoacoustic dual modality imaging," *Optics Express*, vol. 22, pp. 26365–26374, Oct. 2014. Publisher: Optical Society of America.
- [11] O. Oralkan, A. Ergun, J. Johnson, M. Karaman, U. Demirci, K. Kaviani, T. Lee, and B. Khuri-Yakub, "Capacitive micromachined ultrasonic transducers: next-generation arrays for acoustic imaging?," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, pp. 1596–1610, Nov. 2002. Conference Name: IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.
- [12] E. Zhang, J. Laufer, and P. Beard, "Backward-mode multiwavelength photoacoustic scanner using a planar Fabry-Perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues," *Applied Optics*, vol. 47, pp. 561–577, Feb. 2008. Publisher: Optical Society of America.
- [13] C. Lutzweiler and D. Razansky, "Optoacoustic Imaging and Tomography: Reconstruction Approaches and Outstanding Challenges in Image Performance and Quantification," *Sensors*, vol. 13, pp. 7345–7384, June 2013. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.
- [14] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby, "On the adjoint operator in photoacoustic tomography," *Inverse Problems*, vol. 32, p. 115012, Oct. 2016. Publisher: IOP Publishing.
- [15] J. Meng, L. V. Wang, L. Ying, D. Liang, and L. Song, "Compressed-sensing photoacoustic computed tomography in vivo with partially known support," *Optics Express*, vol. 20, p. 16510, July 2012.
- [16] S. Gan, S. Wang, Y. Chen, X. Chen, W. Huang, and H. Chen, "Compressive sensing for seismic data reconstruction via fast projection onto convex sets based on seislet transform," *Journal of Applied Geophysics*, vol. 130, pp. 194–208, July 2016.
- [17] M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Physical Review E*, vol. 71, p. 016706, Jan. 2005. Publisher: American Physical Society.
- [18] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, p. 021314, Apr. 2010.
- [19] H. Gao, H. F. Choi, P. Claus, S. Boonen, S. Jaecques, G. H. V. Lenthe, G. V. D. Perre, W. Lauriks, and J. D'hooge, "A fast convolution-based methodology to simulate 2-D/3-D cardiac ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, pp. 404–409, Feb. 2009. Conference Name: IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.
- [20] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, May 2019.
- [21] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, pp. 102–127, May 2019.
- [22] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE*, vol. 108, pp. 11–29, Jan. 2020. Conference Name: Proceedings of the IEEE.
- [23] I. Domingues, G. Pereira, P. Martins, H. Duarte, J. Santos, and P. H. Abreu, "Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET," *Artificial Intelligence Review*, vol. 53, pp. 4093–4160, Aug. 2020.
- [24] J. Gröhle, M. Schellenberg, K. Dreher, and L. Maier-Hein, "Deep learning for biomedical photoacoustic imaging: A review," *Photoacoustics*, vol. 22, p. 100241, June 2021.

- [25] A. Hauptmann and B. T. Cox, "Deep learning in photoacoustic tomography: current approaches and future directions," *Journal of Biomedical Optics*, vol. 25, p. 112903, Oct. 2020. Publisher: International Society for Optics and Photonics.
- [26] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, "Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1382–1393, June 2018. Conference Name: IEEE Transactions on Medical Imaging.
- [27] N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning photoacoustic tomography with sparse data," *Nature Machine Intelligence*, vol. 1, pp. 453–460, Oct. 2019. Number: 10 Publisher: Nature Publishing Group.
- [28] Y. E. Boink, "Data-driven reconstruction methods for photoacoustic tomography," Feb. 2021.
- [29] A. A. Samsonov, E. G. Kholmovski, D. L. Parker, and C. R. Johnson, "POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1397–1406, 2004. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20285>.
- [30] S. Hu and L. V. Wang, "Photoacoustic imaging and characterization of the microvasculature," *Journal of Biomedical Optics*, vol. 15, no. 1, 2010.
- [31] Y. Lou, W. Zhou, T. P. Matthews, C. M. Appleton, and M. A. Anastasio, "Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging," *Journal of Biomedical Optics*, vol. 22, p. 041015, Jan. 2017. Publisher: International Society for Optics and Photonics.
- [32] L. Yu, F. Nina-Paravecino, D. R. Kaeli, and Q. Fang, "Scalable and massively parallel Monte Carlo photon transport simulations for heterogeneous computing platforms," *Journal of Biomedical Optics*, vol. 23, p. 010504, Jan. 2018. Publisher: International Society for Optics and Photonics.
- [33] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Limited-View and Sparse Photoacoustic Tomography for Neuroimaging with Deep Learning," *Scientific Reports*, vol. 10, p. 8510, May 2020. Number: 1 Publisher: Nature Publishing Group.
- [34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015. arXiv: 1505.04597.
- [35] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [36] T. Lister, P. A. Wright, and P. H. Chappell, "Optical properties of human skin," *Journal of Biomedical Optics*, vol. 17, p. 090901, Sept. 2012. Publisher: International Society for Optics and Photonics.
- [37] V. G. Peters, D. R. Wyman, M. S. Patterson, and G. L. Frank, "Optical properties of normal and diseased human breast tissues in the visible and near infrared," *Physics in Medicine and Biology*, vol. 35, pp. 1317–1334, Sept. 1990. Publisher: IOP Publishing.
- [38] S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine and Biology*, vol. 58, pp. R37–R61, May 2013. Publisher: IOP Publishing.
- [39] P. Hasgall, F. Di Gennaro, C. Baumgartner, E. Neufeld, B. LLoyd, M. Gosselin, D. Payne, A. Klingenberg, and N. Kuster, "IT'IS Database for thermal and electromagnetic parameters of biological tissues," 2018. Medium: .h5, .xml, .txt type: dataset.
- [40] C. G. Graff, "A new, open-source, multi-modality digital breast phantom," in *Medical Imaging 2016: Physics of Medical Imaging*, vol. 9783, p. 978309, International Society for Optics and Photonics, Mar. 2016.
- [41] Y. Bao, H. Deng, X. Wang, H. Zuo, C. Ma, and C. Ma, "Development of a digital breast phantom for photoacoustic computed tomography," *Biomedical Optics Express*, vol. 12, pp. 1391–1406, Mar. 2021. Publisher: Optical Society of America.
- [42] C. Zhu and Q. Liu, "Review of Monte Carlo modeling of light transport in tissues," *Journal of Biomedical Optics*, vol. 18, p. 050902, May 2013. Publisher: International Society for Optics and Photonics.
- [43] J.-Y. Tinevez, "FIJI Directionality - <https://github.com/fiji/Directionality>," July 2021. original-date: 2014-04-07T22:20:23Z.
- [44] H. Lan, D. Jiang, C. Yang, F. Gao, and F. Gao, "Y-Net: Hybrid deep learning image reconstruction for photoacoustic tomography in vivo," *Photoacoustics*, vol. 20, p. 100197, Dec. 2020.
- [45] T. Vu, M. Li, H. Humayun, Y. Zhou, and J. Yao, "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," *Experimental Biology and Medicine*, vol. 245, pp. 597–605, Apr. 2020. Publisher: SAGE Publications.

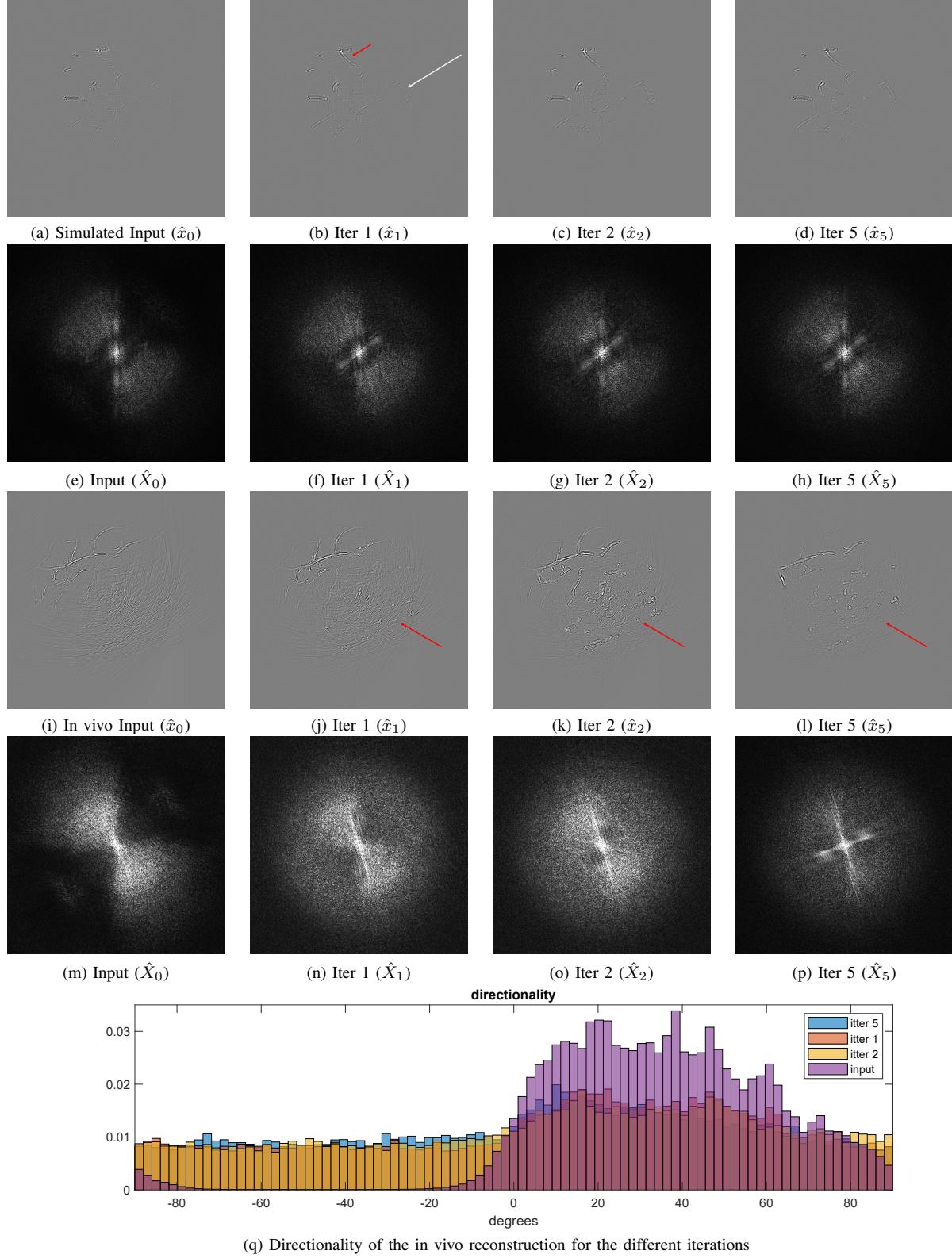


Fig. 6: Results of the POCSNet for a probe as sensor array. the first row (a-d) shows the reconstruction of simulated data from the input data in (a) to iter 1, 2 and 5. the second row (e-h) show the absolute value of the fourier transforms of the images on the first row. the third and fourth row show the same, but then for an in vivo slice. The last row (q) shows the directionality histogram for the different iterations of the in vivo reconstruction

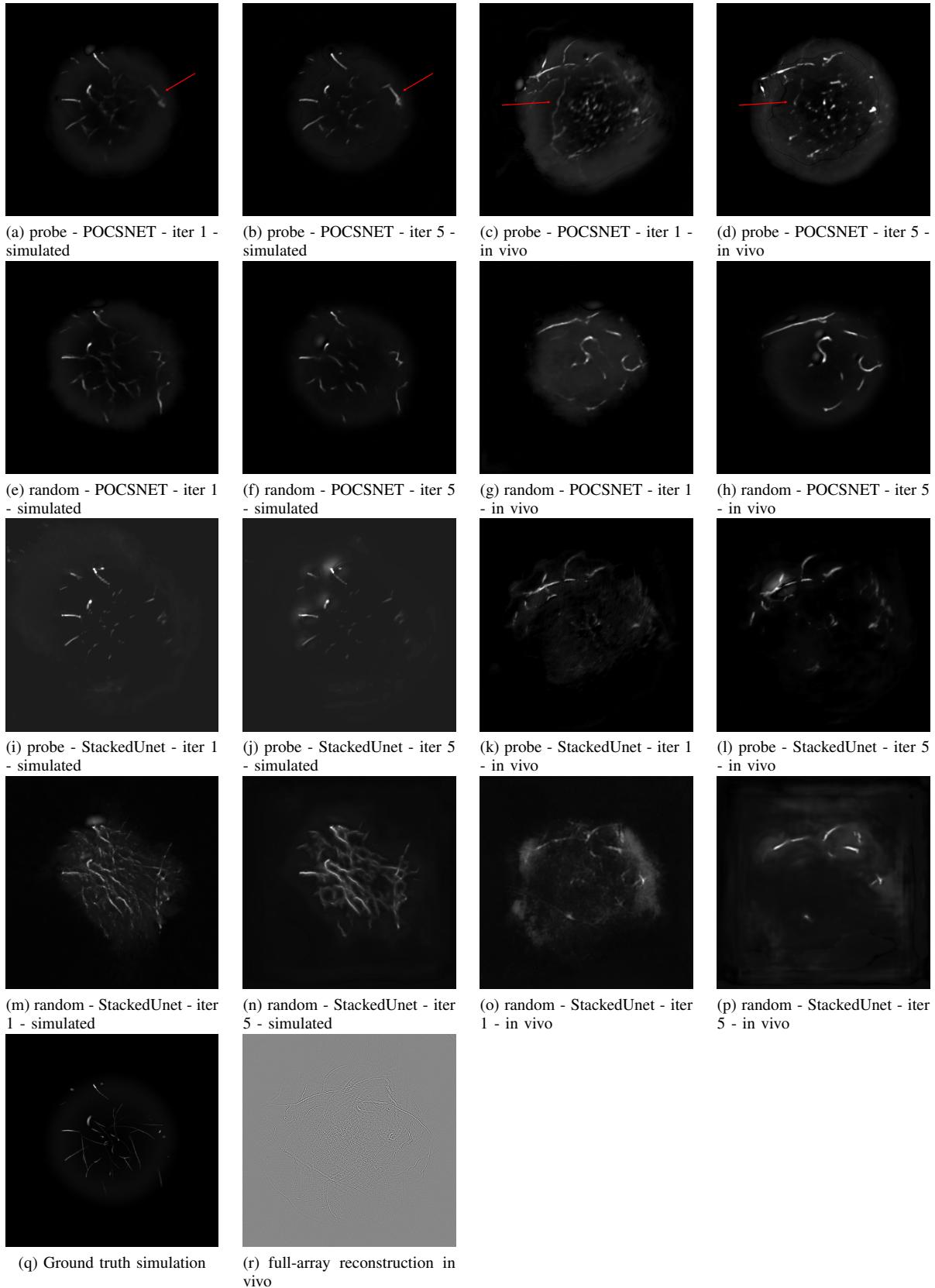


Fig. 7: This figure shows the comparison of the stacked UNET to the POCSNet on a partial and random subarray. In this figure (a-h) show the reconstructions from the POCSNet, with the first row (a-d) being the reconstruction for a probe and the second (e-h) being the reconstruction for a random array. The final row shows the ground truth for the simulated data in (q), the full-array reconstruction of the in-vivo data in (r).