A Partially Learned Algorithm for Joint Photoacoustic Reconstruction and Segmentation

Yoeri E. Boink, Srirang Manohar, and Christoph Brune

***Abstract*—In an inhomogeneously illuminated photoacoustic image, important information like vascular geometry is not readily available when only the initial pressure is reconstructed. To obtain the desired information, image segmentation is often employed as a post-processing step. In this work, we propose to jointly acquire the photoacoustic reconstruction and segmentation, by modifying a recently developed partially learned algorithm based on a convolutional neural network (CNN). We investigate the stability of the algorithm against changes in initial pressures and photoacoustic system settings to acquire an algorithm that is robust to input and system settings. The CNN can be employed for other other modalities and is easily modified to perform other higher-level tasks than segmentation. Our method is validated on challenging synthetic and experimental photoacoustic tomography data in a limited angle and limited view setup. It is computationally less expensive than non-learned iterative methods and gives higher quality reconstructions and higher quality segmentations than state-of-the-art learned and non-learned methods.**

***Index Terms*—inverse problems, convolutional neural networks, photoacoustic tomography, segmentation, learned iterative reconstruction.**

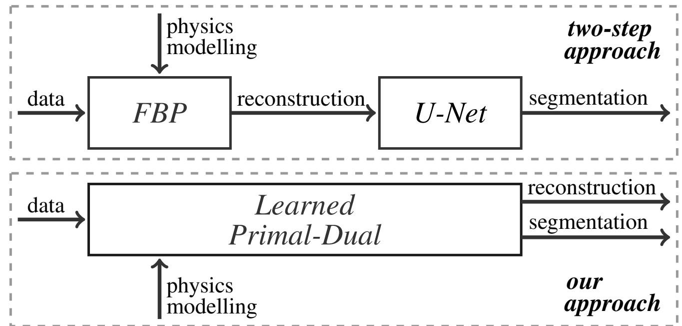

Fig. 1: Example of a two-step approach, where a segmentation is acquired by first applying FBP, after which U-Net is used as a second step. In our approach, we incorporate the physics modelling in the joint learned algorithm, providing both reconstruction and segmentation.

I. Introduction

Photoacoustic tomography (PAT) is a hybrid imaging technique that combines high optical absorption contrast of tissues with high resolution from ultrasound detection. It is being researched for applications in various fields of biomedicine [1], in particular for breast cancer imaging in humans [2]–[4]. One of the main objectives in PAT is to acquire information on the vascular geometry in soft tissue: many diseases, such as breast cancer and rheumatoid arthritis can be characterised with increased blood vessel density and irregular vessel structure. Generally, one first reconstructs the initial pressure image, after which a segmentation algorithm is employed to obtain the segmented vascular geometry. A big disadvantage of such 'two-step' approaches is that reconstruction errors due to inaccurate physics modelling, noise or a lack of data will propagate to the subsequent segmentation step. In this work, we employ a partially learned algorithm to jointly solve the reconstruction and segmentation problem. By utilising the physics model for both problems, we obtain a higher quality segmentation than both learned and non-learned two-step approaches.

Common reconstruction methods for PAT are filtered back-projection (FBP) [5]–[8] and time reversal [9], [10]. These methods apply a variant of the adjoint acoustic operator on the measured data, which has been proven to give the exact result in case of infinite measurements without noise. Empirically, it is still effective if 'enough' measurements with limited noise are available. If this is not the case, these methods fail to give adequate results and one usually switches to iterative approaches employing regularisation [11]–[17]. It is not always clear which regulariser should be selected [15], which limits the practical use of these approaches. Moreover, iterative approaches often need many iterations to converge, which makes them computationally expensive.

Recently there has been a big interest in using deep neural networks for various tomography modalities to give faster and higher quality reconstructions. Very deep networks like U-Net [18] are employed to post-process FBP-results and get a better reconstruction [19]–[22], but can also be used to obtain a segmentation of the tissue under consideration. These networks are not very effective when the quality of the input data is low, as is the case for under-sampled FBP reconstructions. For this reason, other algorithms incorporate the operator that describes the physics in the learning process. DALnet [23] is a non-iterative method that learns the weights of an operator that acts as a pseudo-inverse on the measured data. In [24], the authors first solve some non-learned reconstruction problem, followed by a learned smoothing step; these steps are iteratively executed in the resulting algorithm. Many other learned iterative methods are based on classical iterative methods, but execute far fewer steps in their reconstruction procedure: variational networks [25]–[28] can be seen as a learned variant of proximal gradient methods, where kernel-function pairs of the regularisation term are learned. Learning the regularisation term is also the idea behind [29] and [30]. In [31], a learned variant of gradient-descent is employed, while learned proximal operators are investigated in [32]. The learned primal-dual (L-PD) algorithm [33] can be seen

Y.E. Boink and C.Brune are with the Department of Applied Mathematics, University of Twente, Enschede, 7500 AE, Netherlands. e-mail: y.e.boink@utwente.nl

Y.E. Boink and S. Manohar are with the Biomedical Photonic Imaging Group, University of Twente, Enschede, 7500 AE, Netherlands.

as the learned variant of many very popular and effective primal-dual methods [34]. The key idea in [31], [32] and [33] is that instead of taking a pre-defined gradient or proximal step with respect to the regularisation function, an algorithm learns the steps that have to be taken to obtain a high-quality reconstruction. In [35], L-PD was applied on 3D photoacoustic data. Since their 3D implementation of the acoustic operator is computationally expensive, the authors chose to not learn the whole 'unrolled' algorithm, but learn it iteration by iteration. In case the acoustic operator is expensive, one can also choose to use an approximation of this operator [36], which still gives results superior to non-learned methods. Although empirically strong in their reconstruction capacity, most learned iterative methods lack a mathematical analysis of their convergence and stability. This makes them precarious for implementation in many applications or clinical practice. A convergence proof for a method in which the regulariser was learned is given in [30]. Very recently, Banert *et al.* [37] presented a convergence proof for a class of L-PD algorithms.

Segmentation in biomedical imaging is a well-studied problem; see [38] for an overview. New learned methods for segmentation are developed abundantly, of which U-Net [18] is currently an often employed one. There are not many works on segmentation specifically designed for PAT; a simple segmentation procedure is given in [39] and a complete 'pipeline' from FBP reconstruction to segmentation is presented in [40]. Recently, solving the reconstruction and segmentation problem jointly has become increasingly popular: by treating both problems simultaneously, reconstruction can benefit from information available in the segmentation and vice versa. One such method is given in [41], which also contains an overview on other joint methods.

In this paper, we make use of the partially learned L-PD method and modify it for the goal of joint reconstruction and segmentation in PAT. With this method, accurate information on soft tissue vasculature is directly available, without the necessity of post-processing. By using the acoustic model within the learning of the algorithm, both reconstruction and segmentation are of higher quality than in other learned algorithms. To obtain a robust method, the sensitivity of L-PD to image changes and changes in PAT system settings is investigated and solutions are provided. The method is tested on challenging synthetic and experimental data sets.

To the best of our knowledge, this is the first time the joint problem of photoacoustic reconstruction and segmentation is addressed. Since a major interest in photoacoustic imaging is the visualisation of blood vessels, a segmentation of the vascular geometry is of absolute importance. Our method provides this segmentation, including a reliability estimate, which can be interpreted in an easy way. An analysis of the sensitivity of a learned iterative algorithm is performed, which was not done before.

The remainder of this paper is organised as follows: in section II we give some background on photoacoustic tomography, explain our forward model, and explain the learned primal-dual method. In section III, we first present the neural network architecture that is used, before explaining how its sensitivity is investigated. In the same section, the joint segmentation-reconstruction method is provided, as well as the explanation of the experimental data on which it is tested. Quantitative and visual results of these experiments are given in section IV. In section V we conclude with some remarks and outlook for the future.

II. Theory

A. Photoacoustic tomography

Photoacoustic signals are generated by illuminating tissue with nanosecond laser pulses, which causes the tissue to heat up. Thermoelastic expansion causes the generation of pressure, which is propagated in the form of of ultrasound waves. An array of ultrasound detectors is placed around the tissue to detect this signal. For chosen wavelengths, certain tissue constituents such as haemoglobin have a higher optical absorption coefficient than surrounding tissue, giving the desired high contrast. The generated initial pressure p_0 depends linearly on the combination of optical absorption coefficient μ_a and fluence rate (optical density) Φ:

$$p_0 \propto \mu_a \Phi. \tag{1}$$

Quantitative PAT [42] has the goal to reconstruct μ_a, but has the drawback of being computationally expensive or inaccurate, depending on the light propagation model. The objective of the acoustic PAT inverse problem is to reconstruct p_0. This is computationally less expensive than the quantitative problem and the result already gives insight in the vascular geometry. However, a decaying fluence rate can cause problems for analysing this qualitative image, since the same blood vessel structure can have a lower intensity deeper inside the tissue. Therefore, there is a need for a computationally inexpensive method which does not only provide the reconstructed initial pressure, but also gives a segmentation of the vascular geometry, insensitive to fluence rate reduction in depth.

B. Forward model

In this section we mathematically describe the acoustic PAT forward problem. From this point onwards we write for the initial pressure u instead of p_0 to improve readability of formulas and algorithms. We make use of a projection model with calibration as described in [43]:

$$p(\mathbf{x},t) = |\mathbf{x}-\mathbf{x}_p| \left(\frac{1}{t} \iint_{|\mathbf{x}-\tilde{\mathbf{x}}|=ct} u(\tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}} \right) *_t p_{\text{cal}}\left(\mathbf{x}, t'\right). \tag{2}$$

Here $p(\mathbf{x},t)$ is the measured pressure at location $\mathbf{x}$, which is the result of a spherical mean transform [5] applied to the initial pressure u convolved with a calibration measurement $p_{\text{cal}}\left(\mathbf{x}, t'\right)$. The last is the measurement of a point source located at $\mathbf{x}_p$, which is experimentally approximated. We assume the sound speed c to be known. After pre-processing, we arrive at the simplified equation

$$f = Ku := \iint_{|\mathbf{x}-\tilde{\mathbf{x}}|=ct} u(\tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}}, \tag{3}$$

where K is the acoustic forward operator mapping the initial pressure to the preprocessed measurements. For a more detailed explanation of the calibration measurement and preprocessing, we refer to [7, Chapter 2] and [15].

C. From non-learned to learned primal-dual

Many inverse problems, including the PAT reconstruction problem (3), can be written as

$$f = Ku + \varepsilon, \tag{4}$$

where data f is known, an approximate forward model K is available and the noise level of the additive noise ε is estimated. In case a direct reconstruction method does not give an adequate result, variational methods are employed to solve the minimisation problem

$$\min_u F_f(Au) + G(u). \tag{5}$$

Here F_f is an operator that acts on the dual domain and handles the data, and G is an operator that acts on the primal domain. For instance, writing down a classical $L^2 - TV$ model [45] in discrete form, it reads

$$\min_u \frac{1}{2} \|Ku - f\|_2^2 + \alpha \|\nabla u\|_1 . \tag{6}$$

This can be put in the form of (5) by choosing $A := [K, \nabla]$ and for F_f choose a combination of the 2-norm and 1-norm acting on Au, while $G(u) := 0$. The reason of writing it down in such a form is that any primal-dual algorithm can be applied to solve this minimisation problem [34], of which primal-dual hybrid gradient (PDHG) [46] is a popular choice for tomography problems [15], [47]. A key element in primal-dual algorithms is updating via the proximal operator, which for a general scaled functional $\gamma H(u)$ is defined as

$$\text{prox}_{H(u)} = \underset{\tilde{u}}{\text{argmin}} \left\{ \frac{1}{2} \|u - \tilde{u}\|_2^2 + \gamma H(\tilde{u}) \right\}. \tag{7}$$

This proximal operator takes a descent step in the functional $H(\tilde{u})$, while staying close to a previous iterate u.

for $n \leftarrow 1$ to $\tilde{N}$ **do**
$q^{n+1} = \text{prox}_{\sigma F_f^*}\Big(q^n + \sigma A \left[(1+\theta)u^n - \theta u^{n-1}\right]\Big),$
$u^{n+1} = \text{prox}_{\tau G}\Big(u^n - \tau A^* q^{n+1}\Big).$
end for

Algorithm 1: PDHG (non-learned).

In Algorithm 1, PDHG is stated, where the θ-acceleration step is applied in the first update. The first update takes a proximal step with respect to the convex conjugate of F and acts in the dual domain; its inputs are the previous dual update q^n, data f and the acoustic operator A applied on the primal updates u^{n-1} and u^n. The second update takes a proximal step with respect to G and acts on the previous primal update u^n and the adjoint operator A^* applied on the dual update q^{n+1}.

for $n \leftarrow 1$ to N **do**
$q^{n+1}_{\{1,\ldots,k\}} = q^n_{\{1,\ldots,k\}} + \Gamma_{\Theta_n}\left(q^n_{\{1,\ldots,k\}}, Au^n_1, f\right),$
$u^{n+1}_{\{1,\ldots,k\}} = u^n_{\{1,\ldots,k\}} + \Lambda_{\Theta_n}\left(u^n_{\{1,\ldots,k\}}, A^* q^{n+1}_1\right).$
end for

Algorithm 2: L-PD (learned).

In this work, we make use of the learned primal-dual (L-PD) algorithm [33], which can be derived from PDHG. The idea for the L-PD approach (Algorithm 2) is to not choose the functionals F and G explicitly, but learn the best update steps for each iteration. This is achieved by a CNN, here represented by the nonlinear functions Γ_{Θ_n} and Λ_{Θ_n}, where Θ_n describes the learned weights for iteration n. The key idea is that the inputs of the learned steps of Algorithm 2 are still the same as in Algorithm 1, such that the primal-dual structure is preserved. Note that weights of the network can be different for every iteration $n \in \{1, \ldots, N\}$. Moreover, instead of updating one channel of the primal and dual, we allow the network to use k channels. This could encode some kind of 'history', which is similar to the acceleration of PDHG, where not only u^n, but also u^{n-1} is an input in the first proximal step. However, it can be more general, such that different channels learn different structures in the primal and dual domain. In section IV it will be shown that the L-PD algorithm provides higher quality reconstructions than PDHG applied to a TV regularised model (6). An additional big advantage is that L-PD can be enforced to only use a small amount of iterations, whereas in PDHG, one has to wait till convergence, which generally takes many more iterations.

III. Methodology

A major interest in photoacoustic tomography is the visualisation of blood vessels for breast imaging. For this reason, we made use of retinal blood vessel images from the openly available DRIVE-dataset [48]. A total of 768 training images and 192 test images were obtained by preprocessing patches with a size of 192×192 pixels from this dataset. These training and test images have been further processed to contribute to our two goals: obtaining robustness to variations in images and system settings, and obtaining an additional segmentation in a realistic photoacoustic scenario. This is explained in detail in sections III-A and III-B respectively.

For the L-PD algorithm, both a small network and a larger network were trained. In the first one, the number of iterations was chosen to be $N = 10$ and the number of primal- and dual channels $k = 5$ (cf. Algorithm 2). In the second one, $N = 5$ and $k = 2$ were chosen. In both networks, there are 32 channels in 2 hidden layers. The filter size for the convolutions is 3×3 and ReLu's are chosen as activation functions. The Adam optimiser [49] with an MSE-loss on the difference between ground truth and reconstruction is used:

$$L = \min_{\Theta_{\{1,\ldots,N\}}} \left\| u_{GT} - u_1^N \right\|_2^2 . \tag{8}$$

For stability in the optimisation, the batch size increases from 2 to 16 in three steps [50] during 200 Epochs.

A. Training for robustness

In this experiment we investigate robustness of the L-PD network to changes in the image and system settings, and give solutions to improve this robustness. Digital phantoms are created by setting the background of the vascular images to 0 and setting the maximum to 1. Synthetic data is generated

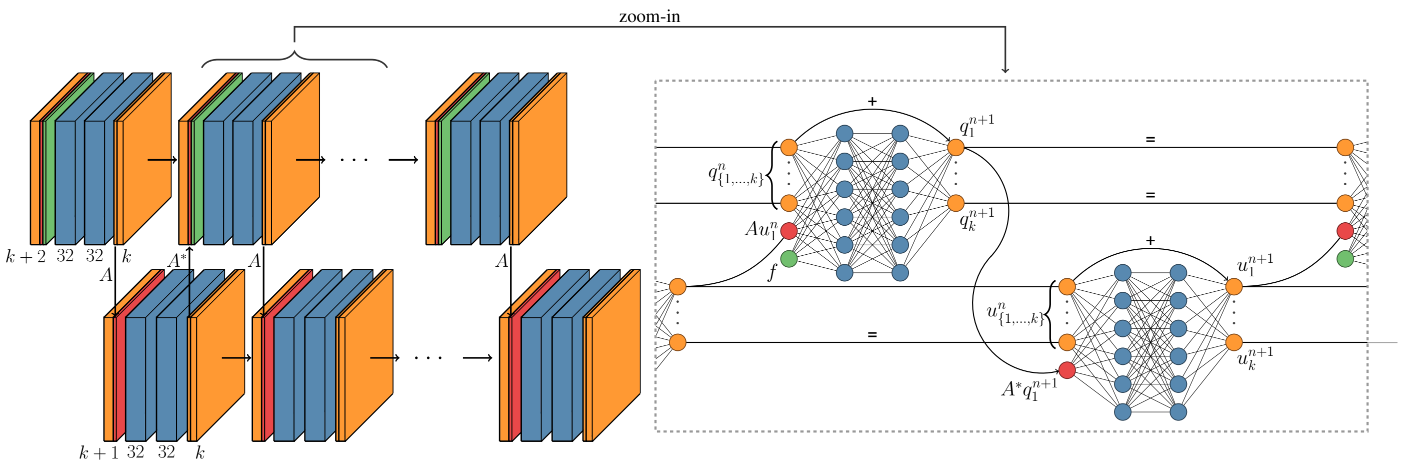

Fig. 2: Visualisation of learned primal-dual architecture as described in Algorithm 2. Number of channels is indicated below the layers on the left. On the right, every node corresponds with a 3×3 convolution followed by a ReLU-activation function.

by applying the forward model as explained in section II-B to these phantoms. For the standard data set, a setting of 32 detectors is used, uniformly placed around the phantom. The L-PD algorithm is trained with the architecture as specified above.

1) Image uncertainty: In Table I, 8 classes of images are defined, in which one or multiple image properties have been changed. For a visual impression, one instance of every class has been shown in Fig. 3. To analyse the sensitivity of L-PD, the algorithm was trained on class 0 and applied to the test set of class i, with $i = 1, 2, \ldots, 7$. Then the network was retrained on a training set of class i and again applied to the test set. We investigate how this affects the reconstruction quality.

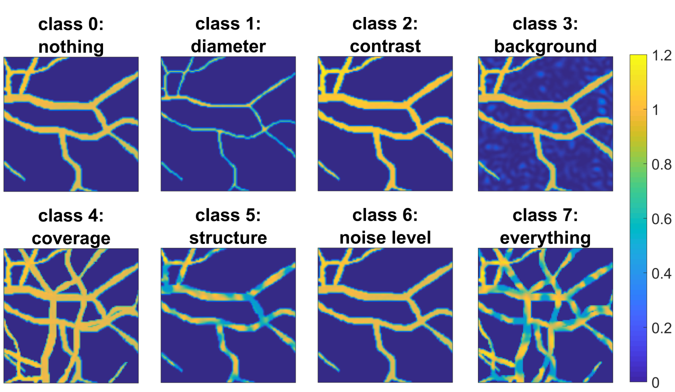

Fig. 3: Example image where one (class 1-6) or more (class 7) features are changed. See Table I for detailed information.

2) System uncertainty: To investigate the sensitivity of L-PD to changes in system settings, the algorithm is first trained on data from 32 uniformly placed detectors. The trained algorithm is applied to test data with 16 or 64 uniformly placed detectors and in a limited view scenario. Next, the network is retrained on a training set where the detectors were randomly chosen from a possibility of 128 detectors. We investigate how retraining on this more diverse training set affects the reconstruction quality.

B. Joint reconstruction and segmentation

The experiments in section III-A are meant to analyse the robustness of the L-PD algorithm to changes in image or in system settings. In these experiments, a homogeneous internal fluence rate is assumed, so that the initial pressure within the blood vessels is uniform.

This is not a realistic scenario for breast imaging, since the fluence rate drops rapidly with depth from the illuminated surface due to scattering and absorption in tissue. To get more realistic digital phantoms, we model light propagation with the diffusion approximation (DA) of the radiative transfer equation (RTE) [51]. This model is applied to the digital phantoms of class 0 in III-A, which serve as the optical absorption coefficient μ_a. The DA has been implemented in the FEniCS framework [52] in Python. In the test set, 48 images have homogeneous internal fluence rate and 144 images have a heterogeneous internel fluence rate modelled with the DA with illumination from either 1, 2 or 4 sides.

For the test set, values for absorption μ_a and reduced scattering μ_s' were selected from literature [53] to represent the values of whole blood for the vessels and glandular tissue or lipid for the background. For the training set, vessel absorption and background scattering were chosen in a wider range to improve reconstruction robustness, as can be seen in section IV-A. The exact values are stated in Table II.

For these more realistic phantoms with spatially varying fluence reate it is difficult to detect the vascular geometry, since blood vessels deep in the tissue will give a lower initial pressure than shallow ones. Moreover, low intensity parts of the reconstruction can be invisible due to artefacts originating from higher intensity parts. To accurately reveal the vascular geometry in all tissue depth, we propose a method to get the segmentation jointly with the photoacoustic reconstruction: instead of solely training for the best reconstruction, the neural network architecture is also trained for a segmentation output. This is done by extending the reconstruction-based loss function (8) with a binary cross-entropy loss on a segmentation

TABLE I: Explanation of image classes where a specific feature (class 1-6) or multiple features (class 7) are changed. See Fig. 3 for a visual impression.

class (c_i)	feature change	explanation
0	nothing	*For an example image see bottom left image in Fig. 3.*
1	diameter	Increase in diameter randomly chosen from the set $\{-2, -1, 0, 1, 2\}$ pixels.
2	contrast	Vessel intensity randomly chosen from the set $\{0.5, 0.7, 1, 1.4, 2\}$.
3	background	With probability $p = \frac{1}{2}$ a uniform increase in background intensity, with $p = \frac{1}{2}$ a speckled background.
4	coverage	With probability $p = \frac{1}{3}$ a removal of all the smallest vessels, with $p = \frac{1}{3}$ nothing, with $p = \frac{1}{3}$ a doubling of vessels.
5	structure	Inhomogeneous vessel intensity scaled between $[m, 1]$, where m is randomly chosen from $\{0.2, 0.4, 0.6, 0.8, 1\}$.
6	noise level	Gaussian noise level randomly chosen from $[0, 3\sigma]$, where σ is the standard noise level
7	All of the above	all of the above changes have all been applied with given probabilities.

TABLE II: Optical properties in the test set.

data set	part of phantom	value
Test set	vessel	$\mu_a = 0.40$ mm^{-1} $\mu_s' = 0.45$ mm^{-1}
	background	$\mu_a = 0.004$ mm^{-1} $\mu_s' = 0.97$ mm^{-1}
Training set	vessel	$\mu_a \in \left[0.20 \text{ mm}^{-1}, 0.60 \text{ mm}^{-1}\right]$ $\mu_s' = 0.45$ mm^{-1}
	background	$\mu_a = 0.004$ mm^{-1} $\mu_s' \in \left[0.50 \text{ mm}^{-1}, 2.00 \text{ mm}^{-1}\right]$

output compared to the ground truth segmentation u_S.

$$L = \min_{\Theta_{\{1,\dots,N\}}} \Big\{ \left\| u_{GT} - u_1^N \right\|_2^2 \qquad (9)$$
$$-\beta \Big(u_S \log(u_2^N) + (1 - u_S) \log(1 - u_2^N) \Big) \Big\}.$$

Here β is a parameter that determines the weighting between reconstruction quality and segmentation quality. For all our experiments, $\beta = 0.5$ was chosen, but empirically it is seen that solutions are not very sensitive to this parameter.

Because the neural network needs to be differentiable, the segmentation output u_2^N is not binary, but in the range $[0, 1]$. This can be interpreted as some kind of reliability estimate: when a pixel value is very close to 1, it is almost surely part of a vessel, while it is almost surely not when its value is very close to 0. To obtain a truly binary segmentation, the non-binary output will be thresholded at the value that gives the highest segmentation accuracy and highest Dice-coefficient [54]. The interpretation of the output as a reliability estimate implies that this threshold should be similar for all outputs, which is verified by experiments. For this reason, we choose one threshold which gives the best binary segmentation, averaged over the whole training set.

C. Application to experimental data

1) Experimental setup: In our experimental setup, we make use of a tomographic photoacoustic imager as specified in [44]. A laser delivers 5 ns pulses of optical energy with a wavelength of 532 nm. For the recording of pressure waves a 1D piezoelectric detector array with 64 elements in a half-circle is used. The detector array has a central frequency of 7.5 MHz with a fractional bandwidth of 85%. Both the detectors and fibre bundles are simultaneously rotatable over 360 degrees. They can also be translated in order to image multiple slices. The detector array has a narrow focus (0.6 mm slice thickness) in one dimension, making it suitable for 2D slice based imaging. A schematic overview of the experimental setup is shown in Figure 4. For a more extensive explanation of the setup, we refer to [44].

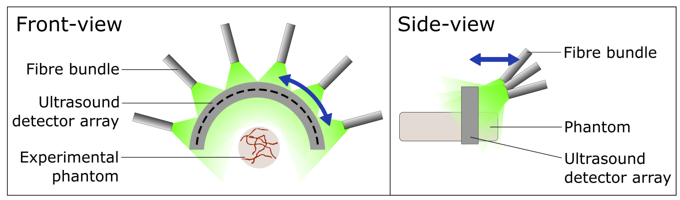

Fig. 4: Schematic overview of the experimental setup, where a phantom is measured.

2) Vascular phantom: For experimental data, a phantom with vessel-shaped absorbers in an optically scattering medium was created. Absorbing filaments were gently placed on stiff agar gel, after which a second layer of agar solution was poured on the whole and allowed to harden. A photograph of the filaments before pouring the second layer of agar is shown in Fig. 5. For details on the phantom creation, we refer to [15].

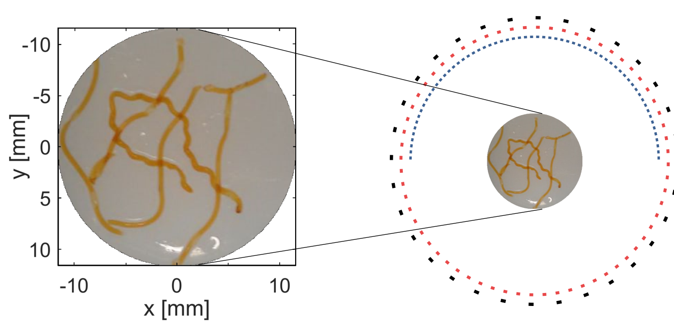

Fig. 5: **Left:** experimental phantom consisting of absorbing filaments in a scattering medium. **Right:** Schematic of detector settings used in experiments. Black: 32 detectors uniform. Red: 64 detectors uniform. Blue: 64 detectors limited-view.

Data is obtained in a limited-view and a uniform detector setting: in the former, fibre bundles and detector array are placed on the top of the phantom; in the latter, fibre bundles and detector array are rotated over an angle of 180 degrees, which means that the detector array covers the whole circle. We point out that by this rotation not only acquisition is uniform, but also light is more homogeneously spread over the surface of the phantom. A schematic drawing of the detector

TABLE III: Performance comparison between basic training and training with more variety in image properties.

PSNR values for two network sizes using different training classes (c)	class 0: nothing	class 1: diameter	class 2: contrast	class 3: backgr.	class 4: coverage	class 5: structure	class 6: noise level	class 7: everything
large network (10 iterations) trained on class i	**41.00**	39.81	**39.53**	**38.27**	**39.18**	**39.09**	**37.72**	**31.86**
large network (10 iterations) trained on class 0	**41.00**	**40.78**	28.77	31.86	38.00	35.97	36.64	21.64
small network (5 iterations) trained on class i	39.32	39.40	37.33	36.22	37.20	37.46	36.38	30.21
small network (5 iterations) trained on class 0	39.32	39.09	25.50	31.04	36.61	34.72	35.48	18.88

setting is given in Fig. 5, with two uniform settings (32 and 64 detectors) and one limited-view setting (64 detectors).

IV. Results

In Fig. 6 reconstructions of a digital phantom from class 0 have been shown, where data was obtained by simulating for a uniform placement of 32 detectors. In line with the results in [33] it can be seen that the non-learned methods FBP and TV show artefacts in the background due to limited measurements and noise, while L-PD gives an almost perfect reconstruction with clear contrast and zero background.

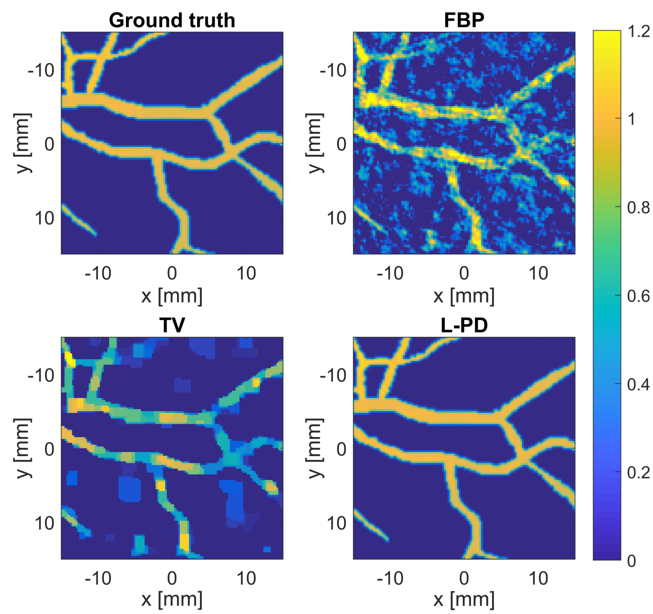

Fig. 6: Reconstructions of a vascular phantom of class 0 using a uniform placement of 32 detectors.

A. *robustness against uncertainty*

Both the large and the small L-PD network were trained on all the image classes as defined in Table I and shown in Fig. 3. We compared reconstructions of L-PD trained on class 0 to L-PD trained on the same class as the test class.

One visual comparison is provided in Fig. 7. Here it can be seen that the L-PD network trained on class 0 incorrectly removes most of the background when applied on test data from class 3: only two spots of high background intensity remain. One might argue that this is a positive property, since out-of-focus elements or low-absorbing structures without diagnostic value can be removed. However, spurious high-intensity background elements can be misclassified as a part of the vascular geometry, which is undesired. After training on

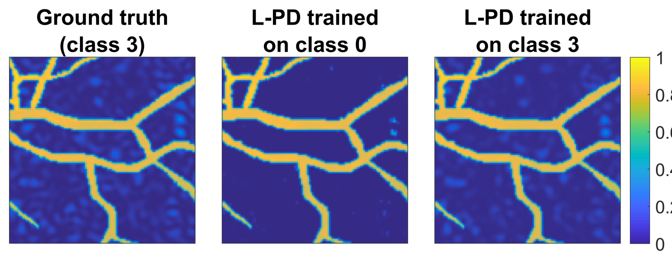

Fig. 7: Example reconstructions of L-PD trained on different classes, both applied to the same test data (class 3).

class 3, the complete background is correctly identified, which makes it less likely that background elements are misclassified as vessels.

We compare peak signal-to-noise ratios (PSNR) of the small and large network for the whole test set for all classes. In Table III it can be seen that for almost all image classes, it helps to train on the specific image class instead of the class 0. This is not very surprising, since the specific class is tailored to the test set on which the trained network is applied. It is however interesting to see that the degree of improvement was very different for the different classes: geometrical changes such as diameter (class 1) and coverage (class 4) did not show any significant improvement, while intensity changes such as contrast (class 2) and background (class 3) showed that it was absolutely necessary to retrain the network for this specific class. Furthermore it was surprising that retraining for different amounts of noise (class 6) was not really necessary, indicating that the denoising capacity of L-PD seems to be quite robust. It can be seen as well that taking a larger network gives a minor improvement over the small network, but no extreme changes are seen. Finally, it is also apparent that the class that contains all the variety of all the other classes benefits from retraining on the same class, but due to this variety, the PSNR value is lower than for other classes. It is unclear to which extent this could be solved by simply taking a bigger training set.

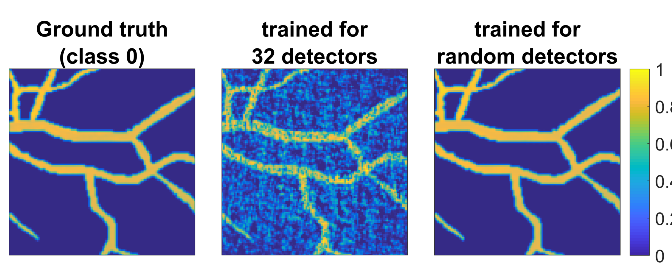

Fig. 8: Reconstructions of L-PD trained on 32 or a random number of detectors, applied to test data from 64 detectors.

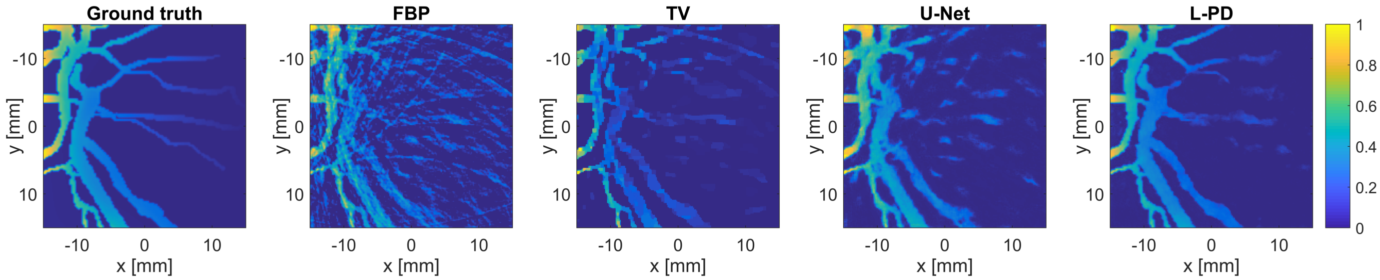

Fig. 9: Reconstructions of a realistic vascular phantom using a uniform placement of 32 detectors.

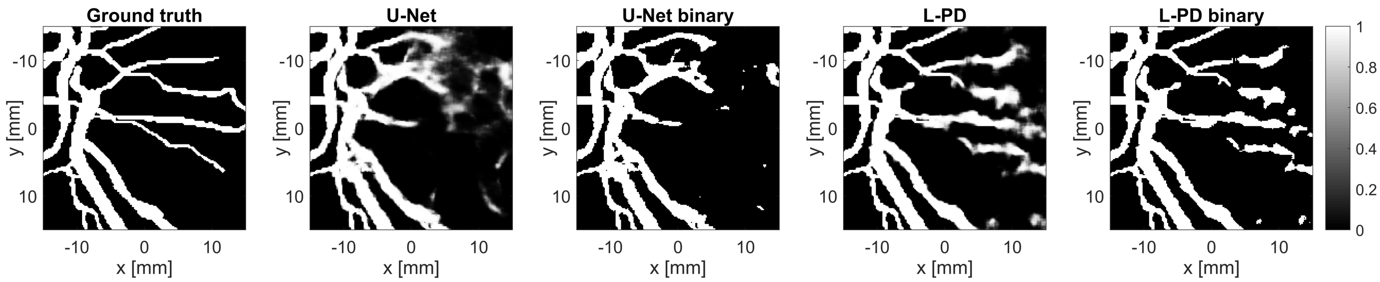

Fig. 10: Segmentations of a realistic vascular phantom using a uniform placement of 32 detectors.

Next we compare reconstructions of L-PD trained on a fixed setting of 32 detectors to L-PD trained on a random number of detectors, as described in section III-A2. In both cases, the algorithm was applied to data from 64 detectors. In Fig. 8, it can be seen that without training for invariance in detector placement, there are a lot of undesired local intensity changes in the reconstruction. When training is done with data from a random detector placement, we obtain the desired high-quality reconstruction. This means that L-PD can be applied to data from different number of detectors, without having to perform a time-consuming retraining of the algorithm.

B. Joint reconstruction and segmentation

In this section, the joint reconstruction-segmentation algorithm as explained in section III-B is compared to four different well known methods. For reconstruction, comparisons with the direct FBP method, the iterative TV-regularised method (6) and the U-Net approach of [19] are made, where a neural network was used to post-process the FBP-result for a higher quality reconstruction. For segmentation, a comparison with the original U-Net approach [18] applied on FBP-reconstructed images is made.

In Fig. 9 a comparison between the different reconstruction methods is shown for synthetic data using a uniform placement of 32 detectors. In this specific phantom, light was modelled to come from the left hand side of the image, resulting in a decaying fluence from left to right. It can be seen that TV and U-Net remove part of the background and partly smooth the vascular structure, but only L-PD is able to correctly identify the entire background while retaining most of the vascular structure.

The same synthetic data and was used for Fig. 10, where a comparison between the two learned segmentation methods is shown. Both the non-binary and binary output of both algorithms indicate the superior performance of the L-PD method. The L-PD method provides the correct vascular geometry, while U-Net fails in the region where the initial pressure is lower. The somewhat fuzzy segmentation of L-PD on the right indicates that on a pixel-level the segmentation is less reliable than on the left. However, it is still clear that four blood vessels are migrating to the right; this possibly important diagnostic information is completely absent in the U-Net segmentation.

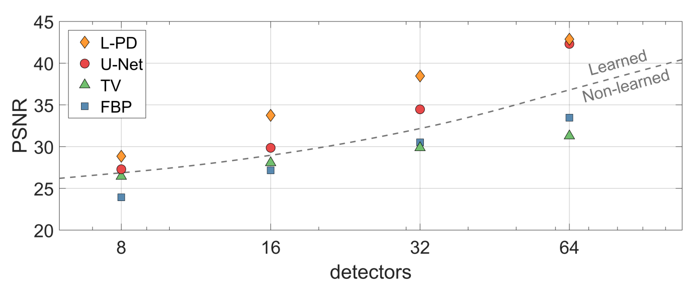

Fig. 11: Reconstruction quality for four reconstruction methods under uniform compressive sampling: L-PD outperforms other methods in all cases.

Compressive sampling is a useful way to limit measurement time, thereby minimising influences due to patient movement. We quantitatively show the superior performance of L-PD in both reconstruction and segmentation under uniform compressive samplings. In Fig. 11 it can be seen that both learned reconstruction methods (U-Net and L-PD) give higher quality reconstructions than non-learned methods (FBP and TV). Moreover, it can be seen that L-PD is significantly better than U-Net, especially in case of substantial compressive sampling, meaning a smaller number of detectors. This can be explained by the fact that the FBP-reconstruction with less than 32 detectors sometimes misses important features in the vascular geometry, which cannot be regained with the U-Net approach. With 64 detectors, the FBP-reconstruction does not really miss any important features, so U-Net can be adequately employed as an artefact removal tool.

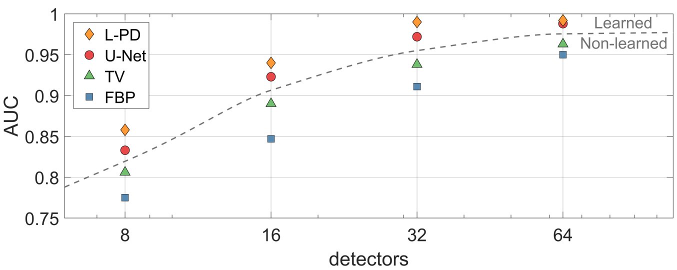

Fig. 12: Segmentation quality for four segmentation methods under uniform compressive sampling: L-PD outperforms other methods in all cases.

In Fig. 12 the segmentation quality is assessed with the area under ROC-curve. The ROC-cuve is obtained by comparing the false positive rate to the true positive rate of the thresholded reconstruction for various threshold values. The area under the curve (AUC) is then used as a single value reflecting the segmentation accuracy. Because the input image for computing the AUC is a non-binary image, our learned segmentations, which hold a value between 0 and 1, can also be compared to the FBP- and TV-reconstructions to assess their segmentation capacity. In Fig. 12 it can be seen that the segmentation quality is very much in line with the reconstruction quality that was shown before: learned methods give more accurate segmentations than non-learned methods. Again, the L-PD method is particularly interesting in case of highly limited number of detectors: it helps to add the acoustic operator in the training, since otherwise previously created artefacts can be difficult to remove.

C. *Experimental results*

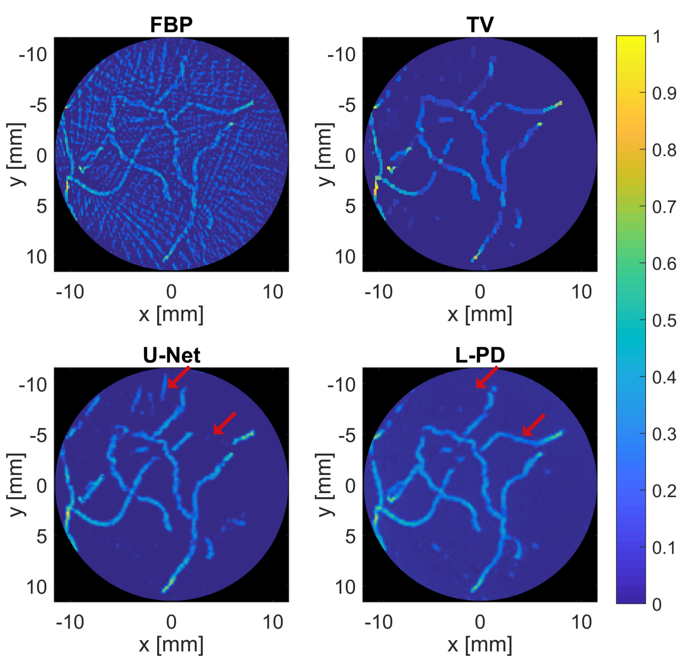

Fig. 13: Reconstructions of an experimental phantom using a uniform placement of 64 detectors.

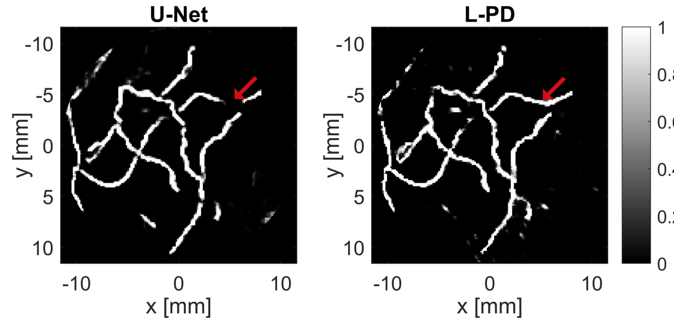

Fig. 14: Segmentations of an experimental phantom using a uniform placement of 64 detectors.

All methods have been tested on experimental data obtained by measuring the experimental phantom explained in section III-C2. Reconstructions and segmentations of the experimental phantom for a uniform sampling of 64 detectors are shown in Fig. 13 and Fig. 14 respectively. Although the sampling of 64 detectors allows for a reasonable reconstruction with either TV or U-Net, the reconstruction with L-PD shows a clearer vascular structure with less artefacts in the background. This is also seen in the segmentation, where a larger part of the vascular geometry is identified.

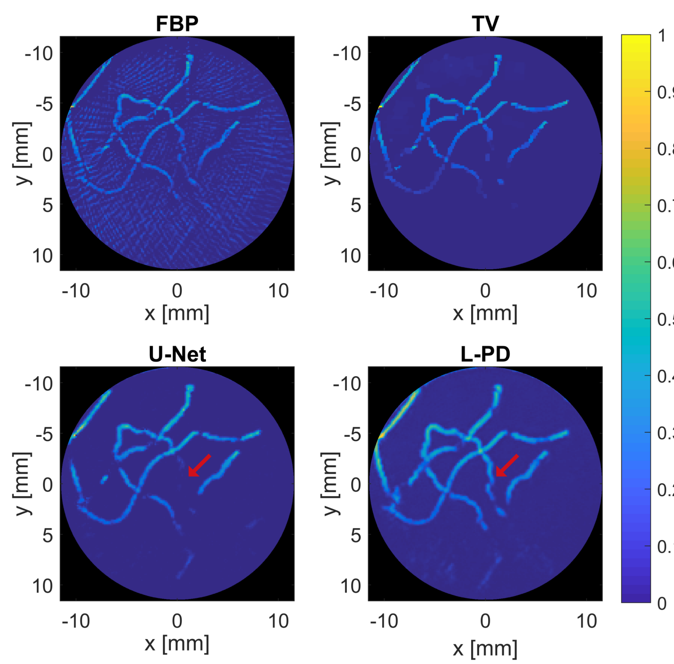

Fig. 15: Reconstructions of experimental phantom using a limited view placement of 64 detectors at the top of the image.

In Fig. 15 and Fig. 16, reconstructions and segmentations for the limited-view sampling of 64 detectors are shown. Recall that in this experimental setup, light deposition and ultrasound detection is at the top half of the phantom. Also here it can be seen that the L-PD approach gives superior reconstructions to the other methods. The L-PD segmentation does not create the spurious parts that are visible in the U-Net segmentation.

V. Conclusion and Outlook

In this paper we developed a partially learned joint reconstruction and segmentation method for PAT. The sensitivity of

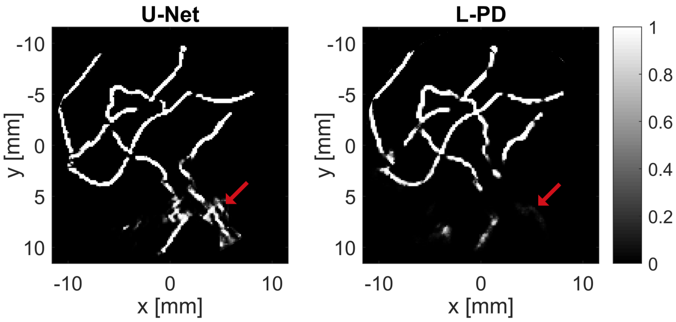

Fig. 16: Segmentations of experimental phantom using a limited view placement of 64 detectors at the top of the image.

the algorithm for reconstruction was investigated and solutions to improve robustness were given. Compared to other learned and non-learned methods, our method gives higher quality reconstruction and segmentation and has a lower computational cost than non-learned iterative methods. Our approach can be applied to other reconstruction problems with different underlying physics or can be used to perform other higher-level tasks different from segmentation.

In our experiments only 2D reconstruction and segmentation were considered, while both 2D [30] and 3D systems [3] are in use. Our method can be applied to 3D without changing the structure of the algorithm. If the 3D photoacoustic operator is computationally too expensive, one might have to change the learning of the algorithm to a step-by-step approach, as was done in [35], but this does not change the structure of the CNN.

One approach to receive the desired robust algorithm is to train on a big variety of input images and system settings. However, when this variety is large, training will take a considerate amount of time. Moreover, it is never possible to train for all possible settings. Therefore, future research could focus on learning a representation of the manifold in which the reconstructions and segmentations should lie. This could be done by combining L-PD with a generative adversarial network [55] or variational autoencoder [56] which learns a low-dimensional representation of the ground truth images. These networks could then learn from a finite number of settings and interpolate on the learned manifold to cover all reasonable inputs and system settings.

Acknowledgment

The authors thank Maura Dantuma for the acquisition of the experimental data. Yoeri Boink, Srirang Manohar and Christoph Brune acknowledge support by the SACAMIR project of Top Sector Life Science & Health. Srirang Manohar and Christoph Brune acknowledge support by the 4TU programme Precision Medicine. Srirang Manohar acknowledges the European Union's Horizon 2020 research and innovation programme H2020 ICT 2016-2017 under grant agreement No 732411, which is an initiative of the Photonics Public Private Partnership.

References

[1] Y. Zhou, J. Yao, and L. V. Wang, "Tutorial on photoacoustic tomography," *J. Biomed. Opt.*, vol. 21, no. 6, p. 061007, 2016.

[2] M. Heijblom, D. Piras, M. Brinkhuis, J. C. G. van Hespen, F. M. van den Engh, M. van der Schaaf, J. M. Klaase, T. G. van Leeuwen, W. Steenbergen, and S. Manohar, "Photoacoustic image patterns of breast carcinoma and comparisons with Magnetic Resonance Imaging and vascular stained histopathology," *Sci. Rep.*, vol. 5, no. May, p. 11778, 2015.

[3] M. Toi, Y. Asao, Y. Matsumoto, H. Sekiguchi, A. Yoshikawa, M. Takada, M. Kataoka, T. Endo, N. Kawaguchi-Sakita, M. Kawashima, E. Fakhrejahani, S. Kanao, I. Yamaga, Y. Nakayama, M. Tokiwa, M. Torii, T. Yagi, T. Sakurai, K. Togashi, and T. Shiina, "Visualization of tumor-related blood vessels in human breast by photoacoustic imaging system with a hemispherical detector array," *Sci. Rep.*, vol. 7, no. December 2016, p. 41970, 2017.

[4] 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," *Nat. Commun.*, vol. 9, no. 1, 2018.

[5] R. Kruger, P. Liu, Y. Fang, and A. Robert, "Photoacoustic ultrasound (PAUS)—Reconstruction tomography," *Med. Phys.*, vol. 22, no. 10, p. 1605, 1995.

[6] D. Finch and S. K. Patch, "Determining a Function from Its Mean Values Over a Family of Spheres," *SIAM J. Math. Anal.*, vol. 35, no. 5, pp. 1213–1240, 2004.

[7] G. R. Willemink, *Image reconstruction in a passive element enriched photoacoustic tomography setup*. 2010.

[8] M. Haltmeier, "Universal Inversion Formulas for Recovering a Function from Spherical Means," *SIAM J. Math. Anal.*, vol. 46, no. 1, pp. 214–232, 2014.

[9] P. Burgholzer, G. J. Matt, M. Haltmeier, and G. Paltauf, "Exact and approximative imaging methods for photoacoustic tomography using an arbitrary detection surface," *Phys. Rev. E*, vol. 75, no. 4, p. 046706, 2007.

[10] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields.," *J. Biomed. Opt.*, vol. 15, no. 2, p. 021314, 2010.

[11] X. L. Dean-Ben, A. Buehler, V. Ntziachristos, and D. Razansky, "Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography," *IEEE Trans. Med. Imaging*, vol. 31, no. 10, pp. 1922–1928, 2012.

[12] C. Huang, K. Wang, L. Nie, L. V. Wang, and M. A. Anastasio, "Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1097–1110, 2013.

[13] S. R. Arridge, P. Beard, M. M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, and E. Zhang, "Accelerated high-resolution photoacoustic tomography via compressed sensing," *Phys. Med. Biol.*, vol. 61, pp. 8908–8940, dec 2016.

[14] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby, "On the Adjoint Operator in Photoacoustic Tomography," *Inverse Probl.*, vol. 32, p. 115012, nov 2016.

[15] Y. E. Boink, M. J. Lagerwerf, W. Steenbergen, S. A. Van Gils, S. Manohar, and C. Brune, "A framework for directional and higher-order reconstruction in photoacoustic tomography," *Phys. Med. Biol.*, vol. 63, no. 4, 2018.

[16] L. V. Nguyen and M. Haltmeier, "Reconstruction algorithms for photoacoustic tomography in heterogenous damping media," no. 47, 2018.

[17] J. Frikel and M. Haltmeier, "Efficient regularization with wavelet sparsity constraints in photoacoustic tomography," *Inverse Probl.*, vol. 34, no. 2, 2018.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI 2015, Med. Image Comput. Comput. Interv.*, pp. 234–241, 2015.

[19] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[20] S. Antholzer, M. Haltmeier, and J. Schwab, "Deep Learning for Photoacoustic Tomography from Sparse Data," pp. 1–13, 2017.

[21] A. Kofler, M. Haltmeier, C. Kolbitsch, M. Kachelriess, and M. Dewey, *A U-Nets Cascade for Sparse View Computed Tomography*, vol. 11074. Springer International Publishing, 2018.

[22] D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, and L. Maier-Hein, "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," in *Photons Plus Ultrasound Imaging Sens. 2018*, no. February, p. 98, 2018.

[23] J. Schwab, S. Antholzer, R. Nuster, and M. Haltmeier, "DALnet: High-resolution photoacoustic projection imaging using deep learning," pp. 1–18, 2018.

[24] B. Kelly, T. P. Matthews, and M. A. Anastasio, “Deep Learning-Guided Image Reconstruction from Incomplete Data,” no. Nips, 2017.
[25] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock, “Variational Networks: Connecting Variational Methods and Deep Learning,” vol. 10496, pp. 244–253, 2017.
[26] K. Hammernik, T. Würfl, T. Pock, and A. Maier, “A deep learning architecture for limited-angle computed tomography reconstruction,” *Bild. für die Medizin 2017*, pp. 92–97, 2017.
[27] E. Kobler, M. Muckley, B. Chen, F. Knoll, K. Hammernik, T. Pock, D. Sodickson, and R. Otazo, “Variational Deep Learning for Low-Dose Computed Tomography,” in *IEEE Int. Conf. Acoust. Speech Signal Process.*, no. 640156, pp. 6687–6691, 2018.
[28] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated MRI data,” *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.
[29] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, “LEARN: Learned Experts’ Assessment-Based Reconstruction Network for Sparse-Data CT,” *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1333–1347, 2018.
[30] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, “NETT: Solving Inverse Problems with Deep Neural Networks,” pp. 1–21, 2018.
[31] J. Adler and O. Öktem, “Solving ill-posed inverse problems using iterative deep neural networks,” no. 1, pp. 1–24, 2017.
[32] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, “Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 1799–1808, 2017.
[33] J. Adler and O. Oktem, “Learned Primal-dual Reconstruction,” *IEEE Trans. Med. Imaging*, vol. 0062, no. 3, pp. 1–11, 2018.
[34] E. Esser, X. Zhang, and T. F. Chan, “A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science,” *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.
[35] 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 Trans. Med. Imaging*, vol. 37, no. 6, pp. 1382–1393, 2018.
[36] A. Hauptmann, B. Cox, F. Lucka, N. Huynh, M. Betcke, P. Beard, and S. Arridge, *Approximate k-Space Models and Deep Learning for Fast Photoacoustic Reconstruction*, vol. 11074. Springer International Publishing, 2018.
[37] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, “Data-driven nonsmooth optimization,” pp. 1–33, 2018.
[38] S. T. Acton and N. Ray, “Biomedical Image Analysis: Segmentation,” *Synth. Lect. Image, Video, Multimed. Process.*, vol. 4, no. 1, pp. 1–108, 2009.
[39] B. Soetikno, S. Hu, E. Gonzales, Q. Zhong, K. Maslov, J.-M. Lee, and L. V. Wang, “Vessel segmentation analysis of ischemic stroke images acquired with photoacoustic microscopy,” vol. 822345, no. February 2012, p. 822345, 2012.
[40] P. Raumonen and T. Tarvainen, “Segmentation of vessel structures from photoacoustic images with reliability assessment,” *Biomed. Opt. Express*, vol. 9, no. 7, pp. 2887–2904, 2018.
[41] V. Corona, M. Benning, M. J. Ehrhardt, L. F. Gladden, R. Mair, A. Reci, A. J. Sederman, S. Reichelt, and C.-B. Schoenlieb, “Enhancing joint reconstruction and segmentation with non-convex Bregman iteration,” pp. 1–31, 2018.
[42] B. Cox, J. G. Laufer, S. R. Arridge, and P. C. Beard, “Quantitative spectroscopic photoacoustic imaging: a review,” *J. Biomed. Opt.*, vol. 17, no. 6, 2012.
[43] Y. Wang, D. Xing, Y. Zeng, and Q. Chen, “Photoacoustic imaging with deconvolution algorithm.,” *Phys. Med. Biol.*, vol. 49, no. 14, pp. 3117–3124, 2004.
[44] P. Van Es, R. Vlieg, S. Biswas, E. Hondebrink, J. Van Hespen, H. Moens, W. Steenbergen, and S. Manohar, “Coregistered photoacoustic and ultrasound tomography of healthy and inflamed human interphalangeal joints,” *Prog. Biomed. Opt. Imaging - Proc. SPIE*, vol. 9539, pp. 14–16, 2015.
[45] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D Nonlinear Phenom.*, vol. 60, no. 1-4, pp. 259–268, 1992.
[46] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
[47] E. Y. Sidky, J. H. Jørgensen, and X. Pan, “Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm.,” *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–91, 2012.
[48] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
[49] D. P. Kingma and J. L. Ba, “ADAM: A Method for Stochastic Optimization,” *ICLR 2015*, pp. 1–15, 2015.
[50] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, “Don’t Decay the Learning Rate, Increase the Batch Size,” no. 2017, pp. 1–11, 2017.
[51] S. R. Arridge, M. Schweiger, M. Hiraoka, and D. T. Delpy, “A finite element approach for modeling photon transport in tissue,” *Med. Phys.*, vol. 20, no. 2, pp. 299–309, 1993.
[52] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, “Unified Form Language: A domain-specific language for weak formulations of partial differential equations,” *ACM Trans. Math. Softw.*, vol. 40, no. 2, pp. 1–37, 2014.
[53] S. L. Jacques, “Optical Properties of Biological Tissues: A Review,” *Phys. Med. Biol.*, vol. 58, no. 11, pp. R37–61, 2013.
[54] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, “Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation,” *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 716–724, 1994.
[55] I. J. Goodfellow, J. Pouget-abadie, M. Mirza, B. Xu, and D. Warde-farley, “Generative-Adversarial-Nets,” *NIPS*, pp. 1–9, 2014.
[56] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” no. Ml, pp. 1–14, 2013.