This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

Model based learning for accelerated, limited-view 3D photoacoustic tomography

Andreas Hauptmann, Felix Lucka, Marta Betcke, Nam Huynh, Jonas Adler, Ben Cox, Paul Beard, Sebastien Ourselin, and Simon Arridge

***Abstract*—Recent advances in deep learning for tomographic reconstructions have shown great potential to create accurate and high quality images with a considerable speed-up. In this work we present a deep neural network that is specifically designed to provide high resolution 3D images from restricted photoacoustic measurements. The network is designed to represent an iterative scheme and incorporates gradient information of the data fit to compensate for limited view artefacts. Due to the high complexity of the photoacoustic forward operator, we separate training and computation of the gradient information. A suitable prior for the desired image structures is learned as part of the training. The resulting network is trained and tested on a set of segmented vessels from lung CT scans and then applied to *in-vivo* photoacoustic measurement data.**

***Index Terms*—Deep learning, convolutional neural networks, photoacoustic tomography, iterative reconstruction**

I. Introduction

Photoacoustic Tomography (PAT) is an emerging "Imaging from Coupled Physics" technique [1] that can obtain high resolution 3D *in-vivo* images of absorbed optical energy by sensing laser-generated ultrasound (US) [2], [3], [4], [5], [6], [7]. If data is obtained over a complete surface surrounding the domain of interest, and for all times over which the acoustic waves are propagating, then the inverse problem can be solved directly by several analytical or numerical algorithms [8]. The fastest of such methods just require to solve a single wave equation; see Section II-B for details. In many practical applications, restricted spatial and/or temporal sampling of the US signal is either imposed due to geometrical limitations (e.g. limited view) [9], or by the choice to utilise a compressed-sensing (CS) undersampling strategy in order to accelerate data acquisition [10]. In such cases, direct reconstruction methods are not optimal to obtain high quality reconstructions as they give rise to artefacts and/or adverse noise amplification.

Recently, several groups showed that variational image reconstruction methods that iteratively minimise a penalty function involving an explicit model of the US propagation and prior constraints on the image structure can provide significantly better results in these situations [11], [12], [13], [14], [15], [16]. However, a crucial drawback of these methods is their considerably higher computational complexity and the difficulty to handcraft prior constraints that capture the spatial structure of the target accurately enough.

As the strongest contrast in biological soft tissue is given by haemoglobin, a central promise of PAT is to deliver high quality images of blood vessel networks, e.g., for assessing the vascularization of tumors [17], [18]. Consequently we assume in this study that our targets are vessel rich and hence we learn suitable prior constraints from a set of segmented vessels.

A. Deep Learning in Image Reconstruction

The huge recent success of Deep Learning methods in image processing and computer vision has seen an increasing interest in applying similar strategies to tomographic reconstruction problems. Deep Neural Networks (DNN) are especially popular due to the low latency of a forward pass through a network which leads to prospective highly efficient reconstruction algorithms.

In this paper we differentiate between two fundamentally different approaches to involve learning in image reconstruction:

1) Reconstruction followed by learning based post-processing. In this approach image reconstruction is carried out using a simple inversion step, and post-processing is used to remove artefacts and noise.
2) Model based learning and reconstruction. In this approach the forward and adjoint operators of the imaging problem are used directly in the inverse algorithm, with a multi-scale regularisation scheme whose parameters are learned in the training phase.

Many applications of Deep Learning for image reconstruction have been concentrated on the first approach by using a fast and simple direct reconstruction algorithm to obtain low quality and corrupted images and then train a convolutional neural network (CNN) on removing those artefacts, see [19], [20] for an application to CT, [21] for PAT, and MRI [22].

Alternatively following the second approach by including the physical forward model into the network has been studied in [23], [24], [25], [26], [27]. However, these improved results in reconstruction quality typically come at the cost of longer computation times which are effectively limited by the repeated simulation of the physical model.

In this paper we take the second approach. In particular, we utilize our knowledge of the forward operator in the reconstruction process, but we will not invoke handcrafted

Andreas Hauptmann, Felix Lucka, Marta Betcke, Sebastien Ourselin, and Simon Arridge are with the Department of Computer Science, University College London, London, United Kingdom

Nam Huynh, Ben Cox, and Paul Beard are with the Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

Felix Lucka is with the Centrum Wiskunde & Informatica, Amsterdam, Netherlands.

Jonas Adler is with the Department of Mathematics, KTH - Royal Institute of Technology, Stockholm, Sweden and Elekta, Stockholm, Sweden.

Accompanying code for this publication can be found at: https://github.com/asHauptmann/3DPAT_DGD

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

prior constraints on the vessel structures that we are interested in. Instead, we will learn them from the data.

B. Compressed Sensing and Limited View PAT

In several imaging modalities the application of compressed sensing methods has been studied as a means to achieve faster acquisition speeds and/or a reduced dose when using ionising radiation [28], [29], [30]. In PAT this has been studied for example in [13], [31], [32], [33], [34]. Because these methods mandate an appropriate regularisation strategy, the involvement of Deep Learning in compressed sensing is an important topic for study.

As well as data sub-sampling, in this paper we also consider the limited-view problem. Due to geometric restrictions, one can often only access the US field on one side of the tissue. A detailed examination and discussion of sub-sampling combined with the limited view problem for PAT can be found in [13].

C. Overview of this paper

The rest of the paper is organised as follows. In Section II we discuss the physical model of photoacoustic signal generation, as well as direct reconstruction approaches, variational and the corresponding iterative reconstruction approaches, and an outline of the model based learning approach. In Section III we give a detailed description of the architecture and implementation of the model based learning approach as well as a description of its training steps. In Section IV we discuss the measurement details, generation of training data as well as post-processing, i.e. denoising/artifact removal, of direct reconstructions. Results for simulated and 3D *in-vivo* data are shown. Section V provides a detailed evaluation of the results. Finally in Section VI we provide some conclusions and outlook for the future.

II. Photoacoustic Tomography

A. Photoacoustic Signal Generation

To generate the PA signal, a short pulse of near-infrared laser light is sent into biological tissue where the photons will get scattered and absorbed by any chromophores present. Under certain conditions (see [35] for details), part of the absorbed optical energy will be *thermalised*, i.e., converted to heat, and the induced local pressure increase x travels through the tissue as an US wave (*photoacoustic effect*). Spatio-temporal measurements of these waves at the boundary of the tissue constitute the PA signal y. A common way to model the acoustic part of the signal generation is to consider the following initial value problem for the wave equation [8], [12], [35],

$$(\partial_{tt} - c_0^2\Delta)p(r,t) = 0, \quad p(r,t=0) = x, \quad \partial_t p(r,t=0) = 0. \tag{1}$$

The US sensing is then modeled as a linear operator $\mathcal{M}$ acting on the pressure field $p(r,t)$ restricted to the boundary of the computational domain Ω and a finite time window (see [3], [36] for details on measurement systems):

$$y = \mathcal{M}\, p_{|\partial\Omega\times(0,T)}. \tag{2}$$

Equations (1) and (2) define a linear mapping

$$Ax = y, \tag{3}$$

from initial pressure x to measured pressure time series y, which constitutes the *forward problem* in PAT. The corresponding image reconstruction step constitutes the *inverse problem* to (3).

Note that x is a $N_x \times N_y \times N_z$ 3D image of initial pressure and y is a $N_h \times N_v \times N_t$ volume of acquired pressure data as a function of acoustic propagation time. In the examples used in this paper this results in dimension of A of around $7M$ by $4.6M$ which (if fully dense) would require about 123TB of memory in single precision which is intractable for currently available computational resources. Thus image reconstruction methods require either direct, or iterative "matrix-free" implementations as discussed in the next sections.

B. Direct methods for PAT Image Reconstruction

Direct methods are especially attractive in the large scale setting as they only require solution of a single wave equation; i.e., given a computational solver for (1) we can compute an inverse solution with the same computational cost [12]. In particular in this study we choose to compute the *adjoint solution* A^*y, which is close to the inverse solution.

Here, as the wave solver we use a pseudo-spectral time-domain method [37], [38], [39] as implemented in the **k**-`Wave` Matlab Toolbox [40], which allows to run the computations on GPU cards using fast CUDA code.

Whilst direct approaches are computationally efficient they are inadequate for dealing with the sub-sampled limited-view data employed in this paper as we demonstrate next. Figure 1 illustrates the influences of limited-view and sub-sampling on a simple numerical phantom of tubes that should mimic blood vessels. From Figure 1(c), we can see that a reconstruction by A^*y suffers from severe circular artefacts [41] and a systematic loss of contrast with depth. Figure 1(d) shows that these problems are accentuated with sub-sampled data.

C. Variational approach to PAT image reconstruction

Variational methods aim to recover the PA image x in (3) from the measurement y as a minimiser of a penalty function,

$$x \in \operatorname*{argmin}_{x'} \{J(x')\} = \operatorname*{argmin}_{x'} \{d(y, Ax') + \lambda R(x')\}, \tag{4}$$

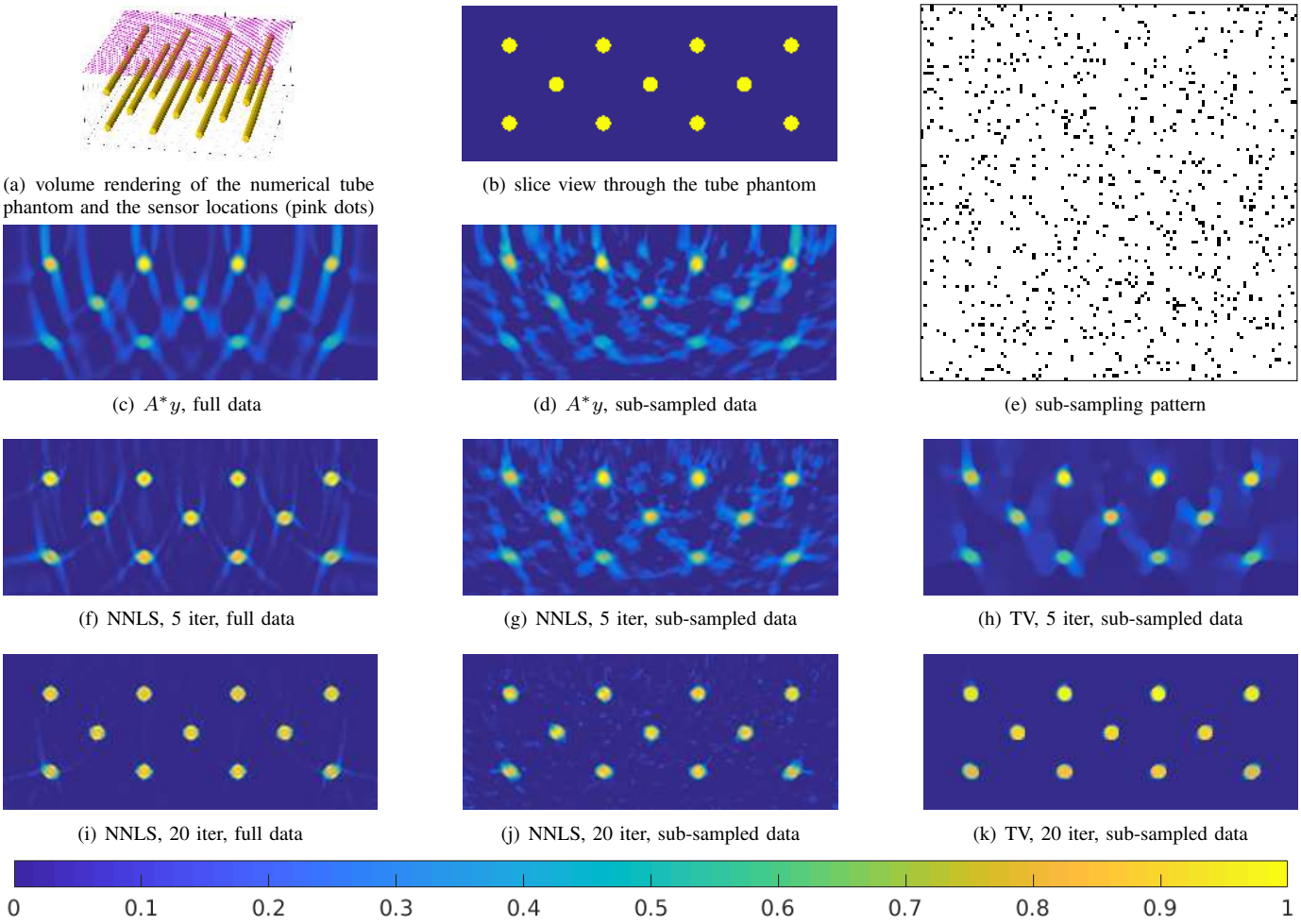
where the fidelity term $d(y, Ax')$ measures the data fit and a regularising term $R(x)$ encodes prior knowledge about the structures in the image. Often, $R(x)$ is convex but not differentiable. A simple approach to find a solution to (4) is given by a *proximal-gradient-descent* scheme:

$$x_{k+1} = \text{prox}_{R,(\lambda\gamma_{k+1})}\left(x_k - \gamma_{k+1}\nabla d(y, Ax_k)\right), \tag{5}$$

with step length $\gamma > 0$ and where the *proximal operator* solves an image denoising problem:

$$\text{prox}_{R,\alpha}(y) = \operatorname*{argmin}_x \left\{ R(x) + \frac{1}{2\alpha}\|x - y\|_2^2 \right\}. \tag{6}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging



(a) volume rendering of the numerical tube phantom and the sensor locations (pink dots)

(b) slice view through the tube phantom

(c) A^*y, full data

(d) A^*y, sub-sampled data

(e) sub-sampling pattern

(f) NNLS, 5 iter, full data

(g) NNLS, 5 iter, sub-sampled data

(h) TV, 5 iter, sub-sampled data

(i) NNLS, 20 iter, full data

(j) NNLS, 20 iter, sub-sampled data

(k) TV, 20 iter, sub-sampled data

Fig. 1. Illustration of the properties and errors of different image reconstruction methods using a simple numerical phantom consisting of tubes. (a)&(b): Visualizations of the numerical phantoms. (e): Illustration of the sub-sampling pattern. Every pixel corresponds to one of the 118×118 scanning locations shown as pink dots in (a). We sub-sample by a factor of 16, i.e., of all locations, a fraction of $1/16$ is chosen at random and visualized by a black pixel. (c)-(d)&(f)-(j): Slice views through the reconstructions of the tube phantom by different methods and for full or sub-sampled data.

The drawback of the above procedure is the difficulty to choose a suitable regularisation term $R(x)$, a regularisation parameter $\lambda > 0$, that balances data fit and the regularisation properties, and the potentially large number of iterations it takes to converge.

As shown in, e.g., [11], [12], [13], iterative image reconstruction methods of the form (5) that solve variational regularisation problems [42] like (4) can improve upon the direct image reconstruction methods. For instance, we can incorporate the physical constraint that the initial pressure increase x is always positive by choosing $R(x)$ to be 0 if $x \geqslant 0$ and ∞ otherwise. For this, $\text{prox}_{R,\alpha}(y) = \max(y, 0)$. With the canonical choice $d(y, Ax') = \frac{1}{2}\|Ax' - y\|_2^2$, (4) simply becomes a non negative least squares (NNLS) solution. Figures 1(f), 1(i) demonstrate that with increasing number of iterations, both limited-view artefacts and the systematic loss of contrast disappear. However, they also show that the convergence in deeper, non-central parts of the image is considerably slower and the limited-view will still manifest in blurry edges. For the sub-sampled data case shown in Figures 1(g), 1(j) we see similar effects although in addition, noise-like artefacts remain. As examined in [13], using noise-reducing, edge-preserving regularisation like the (isotropic) total variation (TV) functional $R(x) = \|\nabla x\|_1$ can further improve such results as can be seen in Figures 1(h), 1(k). The main problem of such iterative approaches is in terms of computation times, compared to the linear backprojection by A^*y which requires the solution of one wave equation, computing 20 iterations of NNLS or TV requires in total 40 additional solutions of a wave equation.

D. Model Based Learning

Regularisation functionals like TV are popular because they often allow for a mathematical analysis of the minimisers of (4) and have been designed to perfectly recover certain aspects of x, e.g., its singularities [43]. As such, they often yield spectacular results for simple numerical or experimental phantoms like the ones shown in Figure 1. In many applications however, typical images x have a more involved structure and the prior information expressed by simple regularisers like TV does not lead to optimal results. One example is given by sub-sampled PAT measurements of vessel networks [13]. If we have a set of typical PA images of vessel networks, we could try to learn more suitable prior information and how to best incorporate it

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

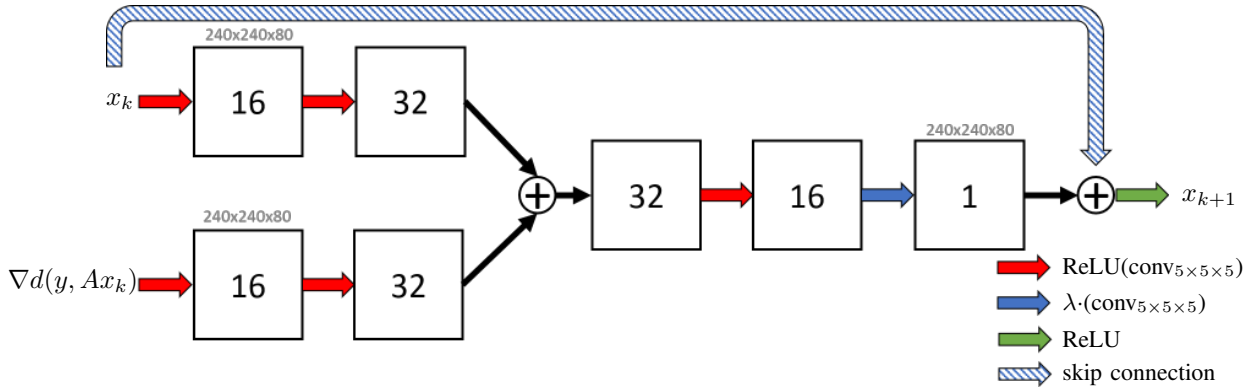


Fig. 2. Diagram of one convolutional neural network, denoted as G_{θ_k}, representing one iteration of the deep gradient descent. Image size for input and output is indicated in gray. The red arrows denote a convolutional layer with $5 \times 5 \times 5$ kernels followed by a ReLU, the resulting channels in each layer are indicated in the squares. The blue arrow denotes a convolutional layer followed by a scalar multiplication. The residual update (by the skip connection) is then projected to the positive numbers by the last ReLU.

in an iterative image reconstruction approach that also utilizes measurement information over the gradient of the data fit,

$$\nabla \tfrac{1}{2}\|Ax_k - y\|_2^2 = A^*(Ax_k - y), \tag{7}$$

at every step k.

Inspired by [44], [45] we take the structure of (5) as a starting point: Each iteration consists of updating x_k by combining measurement information delivered through the gradient $\nabla d(y, Ax_k)$ with an image processing step. Instead of deriving the concrete form of this combination from a fixed penalty function (4), we propose to learn instead an update function for each iteration

$$x_{k+1} = G_{\theta_k}(\nabla d(y, Ax_k), x_k). \tag{8}$$

This implies that the effect of the regularising term is now learned from the data during training. The functions G_{θ_k} correspond to CNNs with different, learned parameters θ_k but with the same architecture. The network structure is kept simple and should mimic a proximal gradient update (5). Due to the representation of each update by a CNN applied to the current x_k and the gradient $\nabla d(y, Ax_k)$, we call the whole algorithm *deep gradient descent* (DGD).

In contrast to [23], [26], [45] we train the DGD layer by layer (layer corresponding here to one iterate), i.e. we learn the parameters θ_k for each iteration separately. In this way we can exclude the photoacoustic operator from the training procedure. This is necessary to make the training feasible. Note, that the photoacoustic operator has complexity $\mathcal{O}(N^4 \log(N))$ [12], for a volume of size $n = N \times N \times N$, compared to CT and MRI where A has complexity $\mathcal{O}(N^3 \log(N))$ for a volume of size $n = N \times N \times N$. Therefore we think that such layer by layer training scheme is the only feasible approach for iterative high-resolution 3D PAT imaging at the present stage.

III. Implementation

In a CNN, each layer is of the following form: Given the input g and output h with channel index sets $\mathcal{I}$, $\mathcal{J}$ respectively, then

$$h_i = \varphi\left(b_i + \sum_{j \in \mathcal{J}} \omega_{i,j} * g_j\right), \ i \in \mathcal{I},$$

with a componentwise nonlinear function φ and convolution $*$. The whole parameter set θ of the network is therefore given by the biases $b_i \in \mathbb{R}$ and convolutional filters $\omega_{i,j} \in \mathbb{R}^{s^n}$ (with kernel size s and spatial dimension n) of each network layer.

The specific architecture we have chosen for the CNNs performing the update in equation (8) is illustrated in Figure 2. In each iteration we input x_k and $\nabla d(y, Ax_k)$ to a similar pipeline, where both are spread to 16 and then 32 channels by a convolutional layer with kernel size $s = 5$ and dimension $n = 3$, equipped with a rectified linear unit as nonlinearity, that is defined as

$$\mathrm{ReLU}(x) = \max(x, 0).$$

The output of both pipelines is added together and first reduced to 16 channels, equipped with a ReLU, and then to 1 channel without a nonlinearity, but a simple scalar multiplication. The result is added to the current iterate and projected to the positive numbers by a ReLU, similar to the proximal for NNLS discussed in Section II-C.

The architecture in this study is motivated by a typical network structure consisting of an analysis/encoding and a reassembling/decoding part. In this analysis part, the number of channels is increased between layers to refine the analysis of the features extracted in the layer before. In the reassembling part, these features must be merged/thresholded to produce an output image, so the number of channels is decreased. Since the main contribution of this work is not the specific neural network architecture, we use a simple architecture following this convention. In particular, the network structure is kept rather small with the motivation in mind that each G_{θ_k} primarily learns how to combine current iterate and gradient as well as a data specific filters, in contrast to a large post-processing network. Furthermore, a compact structure is necessary to minimise the needed memory on the GPU.

A. Training of the deep gradient descent

Given a training set $\{y^i, x^i_{true}\}_i$, we have two options to train the parameters θ_k. The first is to pre-define a maximum of iterations k_{max} and train all θ_k, for $k = 0, \ldots, k_{max} - 1$,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

together to minimise the difference between x^i_{true} and the result of the last iteration $x^i_{k_{max}}$; that is we seek to find

$$\mathcal{E}_{k_{max}} = \min_{\theta_0,\dots,\theta_{k_{max}-1}} \sum_i \|x^i_{k_{max}} - x^i_{true}\|, \tag{9}$$

for some suitable norm. The second approach is to train the parameters sequentially: θ_0 is trained to minimise the difference between x^i_{true} and x^i_1 given data y^i, for all indices i. After that, θ_1 is trained to minimise the difference between x^i_{true} and x^i_2, given the optimal x^i_1 found in the training of the first CNN G_{θ_0}. That means the minimisation of (9) is split into k_{max} independent optimisation problems w.r.t. disjoint subsets of parameters $\theta_k,\ k = 0,\dots,k_{max}-1$, given by

$$\min_{\theta_k} \sum_i \|x^i_{k+1} - x^i_{true}\|, \quad x^i_{k+1} = G_{\theta_k}(\nabla d(y, Ax^i_k), x^i_k). \tag{10}$$

The first approach has the advantage that the network is more flexible to achieve the best possible result after k_{max} iterations, but during the training, the operators A and A^* need to be evaluated many times, since for each training step all x_k and their corresponding gradients have to be computed to evaluate (9). While the second approach is not optimal in the sense that it does not lead to minimal training error, it has two important advantages. Firstly, the computation of the gradient $A^*(Ax - y)$ and training decouple, which is important in view of the cost of application of A and A^* in PAT. Secondly, it provides an upper bound on the training error (9). In fact, (10) can be viewed as a greedy approach which seeks to obtain a minimum in each layer k given x_{k-1} from the previous training step. We note that this property can be used to determine the number of layers k_{max} of the DGD in training by controlling the training error from layer to layer in contrast to choosing it a priori. Therefore, the second approach could also be used as a pre-training stage to initialize the weights for the first approach.

As the computational complexity of simulating acoustic wave propagation in 3D prohibits computing the gradient during any training scheme, we need to follow the second approach here. The whole training procedure we use is summarized in Algorithm 1 for a given number of maximum iterations k_{max} and the reference solution x_{true}.

Algorithm 1 Training Procedure

1: $x_0 \leftarrow A^*y$
2: **function** TRAININGCYCLE
3: $k \leftarrow 0$
4: **while** $k < k_{max}$ **do**
5: Compute $\nabla d(y, Ax_k) = A^*(Ax - y)$
6: **function** TRAINITERATE($\nabla d(y, Ax_k), x_k, x_{true}$)
7: Train for given accuracy
8: **end function**(**return** θ_k)
9: $x_{k+1} \leftarrow G_{\theta_k}(\nabla d(y, Ax_k), x_k)$
10: $k \leftarrow k + 1$
11: **end while**
12: **end function**

B. Evaluation of the deep gradient descent

After training the parameter sets $\{\theta_k\}_{k=0}^{k_{max}-1}$, the learned iterative reconstruction scheme can be evaluated as follows: The new iterate x_{k+1} is computed by applying the network G_{θ_k} to the current iterate x_k and the gradient of the data fit, in particular this means that the gradient has to be computed in every iteration. This procedure is equivalent to Algorithm 1 without calling TRAINITERATE in line 6-8.

IV. Experiments

In this study we are interested in reconstructing human *in-vivo* data and hence we do not have a true target available for the training of measured data. This lack of a ground truth is one of the main challenges in supervised learning. Nevertheless, we chose to train the DGD with supervised learning using simulated data and hence a meaningful data set is crucial for a successful training, for that purpose we use segmented human vessel structures from CT scans as discussed in the next section. The training and evaluation of each network G_{θ_k} has been implemented with TensorFlow [46] in Python. All computations are done on a Titan Xp GPU with 12GB memory.

A. Training on segmented lung vessels

The training data needs to be as realistic as possible to provide a meaningful basis for the algorithm. To achieve this we have used the publicly available data from *ELCAP Public Lung Image Database*[1]. The data set consists of 50 whole-lung CT scans, from which we have segmented about 1200 volumes of vessel structures with a Frangi vesselness filter [47], [48]. The segmented volumes were of size $40 \times 120 \times 120$, and were then scaled up by a factor of 2 to the final target size of $80 \times 240 \times 240$. Out of these volumes we chose 1024 as ground truth x_{true} for the training and simulated limited-view, sub-sampled data using the same measurement setup used in the in-vivo data: We assume that each voxel has the isotropic length $dx = 84.75\mu$m and that the full data is recorded at locations on a grid with grid size $2dx$ on one of the two 240×240 sized outer planes of the volume (i.e., the scanning geometry is similar to Figure 1(a)). In time, $N_t = 486$ pressure samples are recorded with $dt = 16.6$ns. The full data is then sub-sampled as illustrated in Figure 1(e) but by a sub-sampling factor of 4. We have added normally distributed noise to the measured data, such that the resulting SNR was approximately 15 for all measurements and we assumed a sound speed of $c_0 = 1580$m/s. In a nutshell, we obtained the data $y = Ax_{true} + \varepsilon$, where ε denotes the added noise.

The training set for one CNN G_{θ_k} then consists of current iterate x_k, the gradient of the data fit $\nabla d(y, Ax_k) = A^*(Ax_k - y)$, and the ground truth x_{true}. We initialize the iteration with

$$x_0 = A^*y.$$

Precomputing the gradient information for each CNN takes about 10 hours.

[1] http://www.via.cornell.edu/databases/lungdb.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

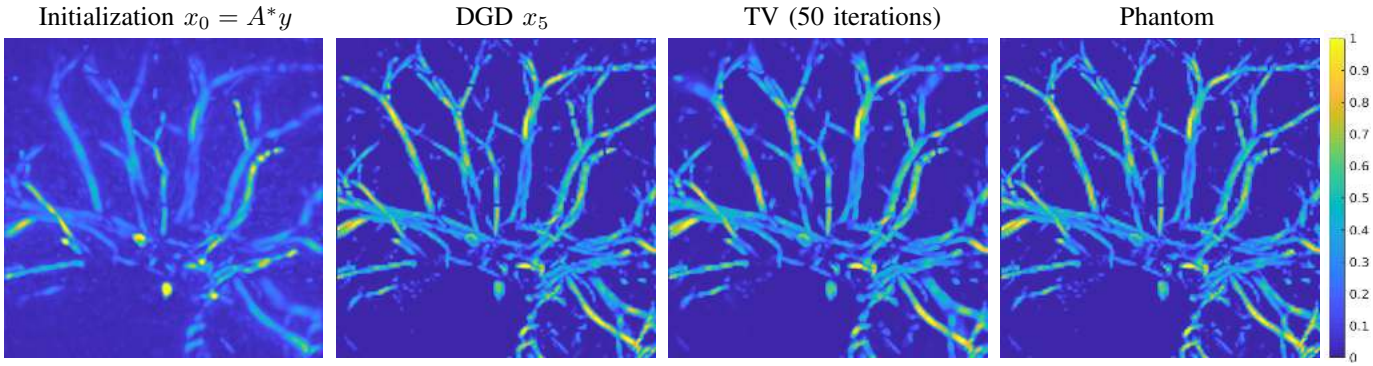


Fig. 3. Reconstruction results for a test image from the segmented CT data (not included in the training), presented images are top-down maximum intensity projections. From left to right: Backprojection of the data and initialization of the network, result of DGD with 5 iterations, TV reconstruction with 50 iterations, phantom used to produce the data.

We trained the CNNs using TensorFlow's implementation of Adam [49]. For the training we used batches of size 2, since this already fills up the memory (12GB) of the GPU completely. We trained each G_{θ_k} for 25000 iterations (i.e. approximately 50 epochs) with an initial step size of $5 \cdot 10^{-5}$ (learning rate), The minimised loss function, i.e. the norm in (10), is chosen as the ℓ^2-distance of new iterate to the true solution x_{true},

$$\text{loss}(x) = \|x - x_{true}\|_2^2.$$

For training the first CNN G_{θ_0} we added an additional constraint to avoid the local minima of zero solutions by penalizing a small norm

$$\text{loss}_{add}(x) = -\alpha \min(\|x\|_2 - \beta, 0),$$

with small $\alpha, \beta > 0$. The training of each CNN G_{θ_k} took about 1 day on the GPU. We have trained 5 iterates, i.e. $k_{max} = 5$, for the deep gradient descent. In total the whole training took 7 days. We note, that this could be speed up by initialisation of θ_k with θ_{k-1} or by more advanced optimisation strategies, see for instance the review [50]. At this point we would like to note, that had we included the operator A and A^* in the training and trained all 5 iterates together, then the time needed for 25000 iterations would be in the order of 70 days, and used at least 5 times more memory. The result of the DGD for simulated data is shown in Figure 3 for an example that was not included in the training set.

B. Post-processing by Deep Learning

To complement this study, we have also implemented the first approach of learning in image reconstruction, see Section I-A, *viz.* taking an initial direct reconstruction and train a network to remove artefacts and noise. Especially popular for improving these initial reconstructions is a CNN introduced as U-Net [51]. We refer to the original paper for the architecture, but roughly summarized its strength lies in a series of skip connections in a multilevel decomposition. For our application, we have followed the modified U-Net architecture proposed by [20] for post-processing of 2D X-ray tomography, that learns to compute an update to the initial reconstruction. We made the necessary modifications for a three-dimensional setting and implemented training and evaluation with TensorFlow.

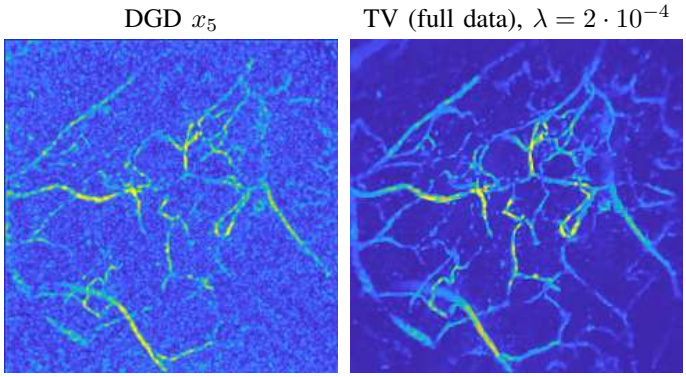


Fig. 4. Reconstruction from real measurement data of a human palm, without adjustments of the training data. The images shown are top-down maximum intensity projections. Left: Result of the DGD trained on images without added background. Right: TV reconstruction as reference from fully sampled data.

To be consistent with the previous section our direct reconstruction, which we seek to improve upon, is obtained by the application of the adjoint $x_0 = A^*y$. The modified U-Net is then trained on the set of pairs $\{x_0^i, x_{true}^i\}_i$. Due to memory restrictions we were only able to train one pair at a time. The loss function is chosen as the combination $\text{loss}(x) + \text{loss}_{add}(x)$, see previous section. The training is then performed with Adam for 75 epochs and a learning rate of 10^{-4}; this took 3 days. The results for simulated data will be discussed in Section V-B.

C. Application to in-vivo data

We now apply our method to in-vivo data of a human palm. The details of the measurement set-up and procedure are described in [14], [52]. All other features like spatial dimensions of reconstruction volume, temporal sampling or the sub-sampling pattern are exactly the same as for the simulated data (cf. Section IV-A).

In-vivo data has different characteristics that are not perfectly represented by the training on synthetic data and hence a direct application of the trained network does not lead to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

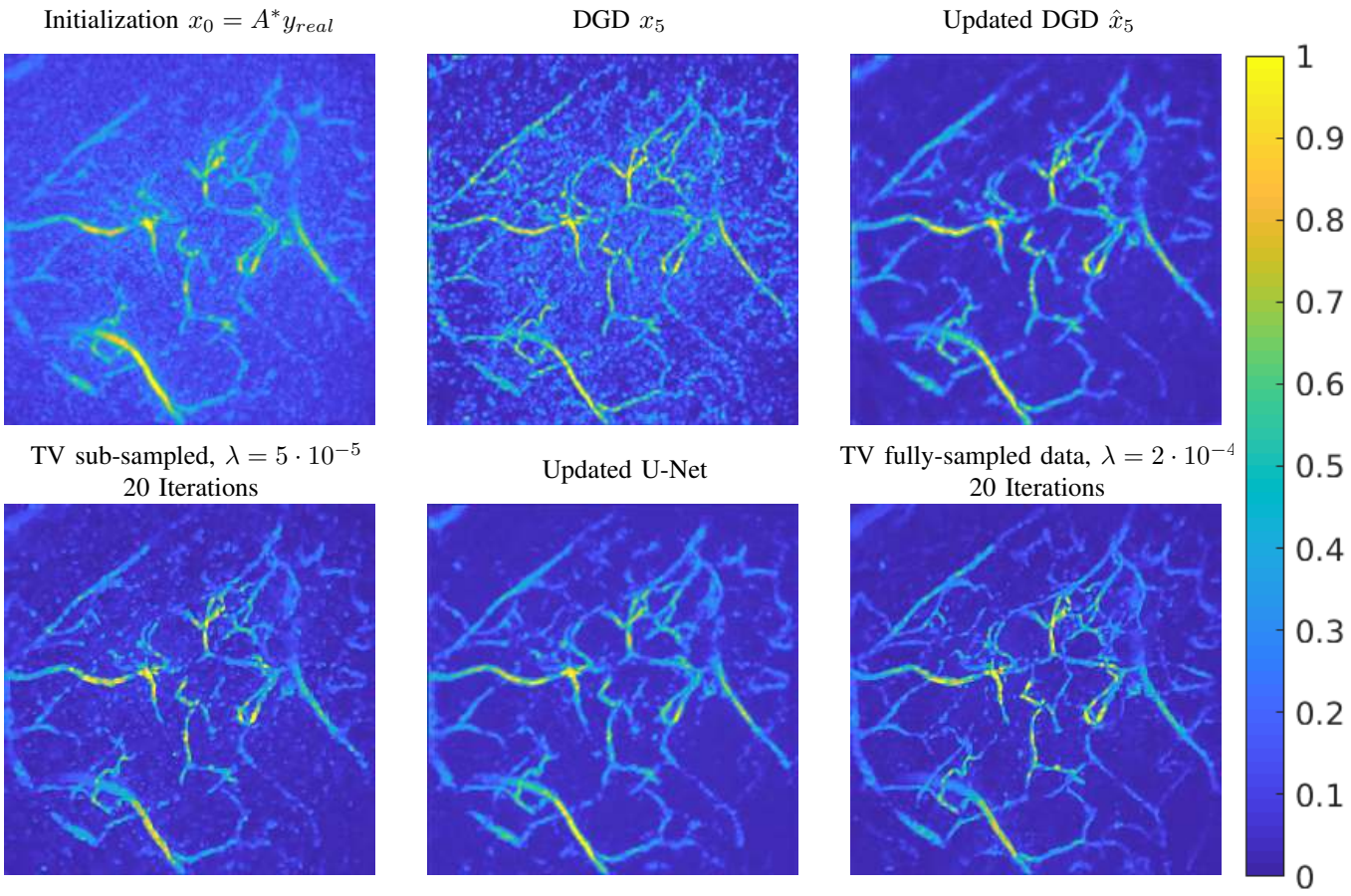


Fig. 5. Example for real measurement data of a human palm. The images shown are top-down maximum intensity projections. First row: from left to right, the initialization from sub-sampled data, the output of DGD trained on background added data after 5 iterations, and updated DGD $G_{\hat{\theta}_k}$ after 5 iterations. Second row: from left to right, TV reconstruction of sub-sampled data with a emphasis on the data fit, updated U-Net reconstruction, reference TV reconstruction of fully-sampled limited-view data. All TV reconstructions have been computed with 20 iterations.

satisfactory results, as illustrated by comparing it to a TV reconstruction in Figure 4. In particular, we see that the network has not learned to effectively threshold the noise-like artefacts in the low absorption regions i.e. regions with low concentration of chromophores. To train our approach to remove these features we simulated the effect of the low absorbing background as a Gaussian random field with short spatial correlation length, clipped the negative parts, scaled it to maximal value 0.1 and added it to each segmented volume x_{true} where ever the intensity of x_{true} did not exceed 0.1 (i.e., the maximum intensity of x_{true} stays 1). The synthetic CT volumes with the added background were then used for the data generation, i.e. $y^i_{back} = Ax^i_{back} + \varepsilon$, whereas the clean volumes x_{true} are used as reference for the training. Here ε is again chosen (see Section IV-A) such that the resulting measurement y^i_{back} had a SNR of approximately 15. We expect the network trained on the modified pairs $\{y^i_{back}, x^i_{true}\}_i$ to be capable of effectively removing the background.

Furthermore, since the expected contrast in the images is crucial for the trained reconstruction procedure, we scaled the measurement as follows. First we computed the standard deviation of the measurement data for all simulated targets. Then we rescaled the sub-sampled real measured data to have a similar standard deviation. This rescaled data is then used for reconstructing with the DGD. The result after 5 iterations is shown in Figure 5.

The results can be further improved performing a transfer training of the previously trained networks G_{θ_k}. This however requires some reference reconstructions from the same or a similar system. We were able to perform such a transfer training with a set of 20 (fully sampled) measurements of a human finger, wrist, and palm from the same system. We then sub-sampled the data (fourfold) to obtain the training data y_{real}. As reference we took weakly regularised TV reconstructions from the fully sampled data, x_{TV}. To update the DGD we have performed an additional 10 epochs of training on the pairs $\{y_{real}, x_{TV}\}$, with a reduced learning rate of 10^{-5}. Such transfer training takes only 90 minutes to update the entire DGD. We denote the updated CNNs by $G_{\hat{\theta}_k}$ and the resulting outputs by $\hat{x}_k$. The effect of the updated DGD is shown in Figure 5.

Additionally, for a full comparison we have performed an update training of the U-Net with the same parameters as above, i.e. 10 epochs and a reduced learning rate of 10^{-5}. The update training of the network took only 20 minutes and the result is shown in Figure 5.

V. Discussion of Results

The results shown in Figure 3 and Figure 5 suggest that the formulation of a gradient descent scheme as a CNN in each iteration does produce competitive results with a considerable reduction in iterations needed, as we will discuss in this

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

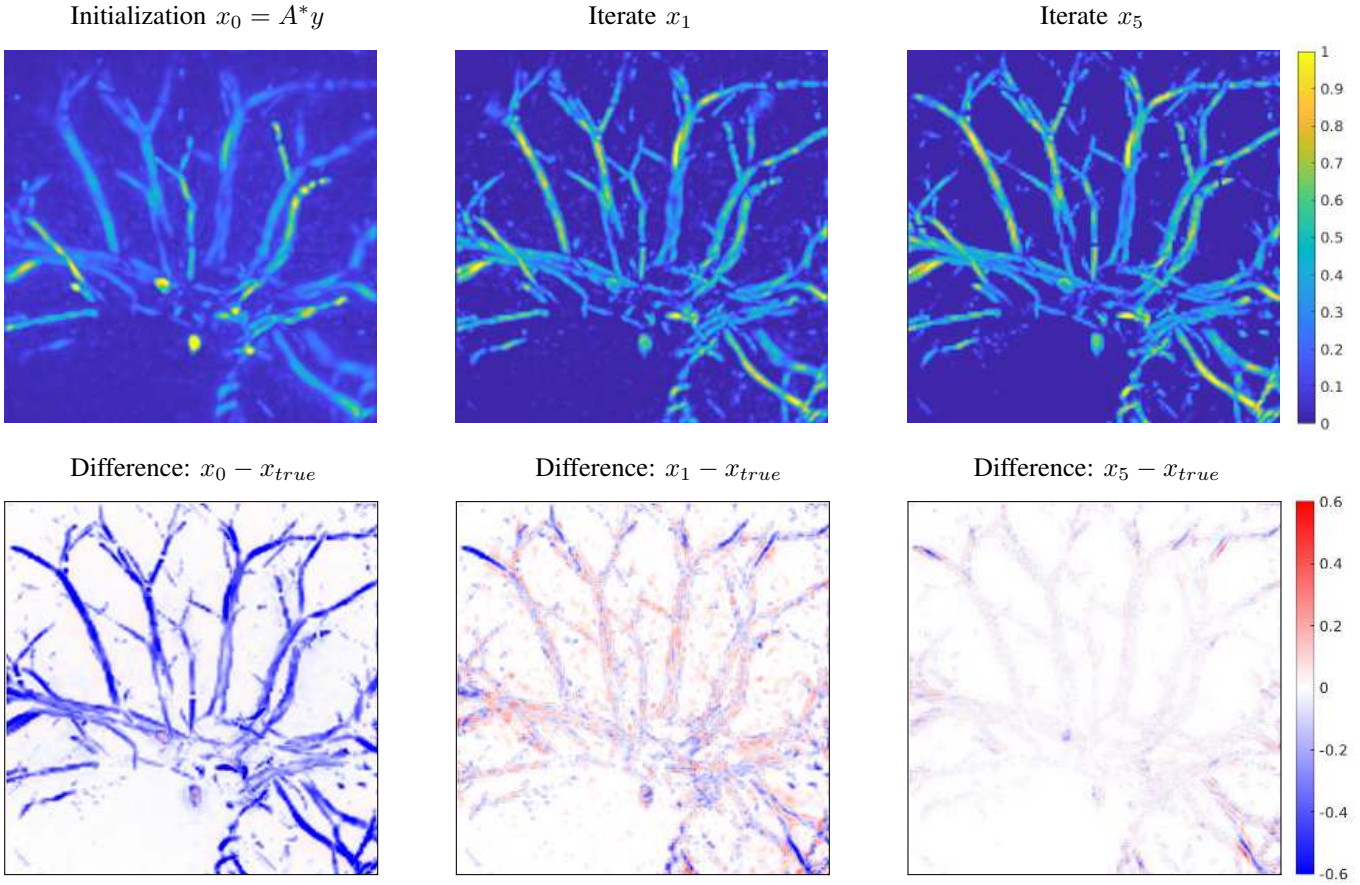


Fig. 6. Progress of iterations in the DGD for a test image from the segmented CT data. Images shown are top-down maximum intensity projections. The top row shows reconstructions and the bottom row shows difference images to the true solution. Difference images are on the same scale, with blue for a negative difference and red for positive. Left: initialization and input to the DGD, maximal value of difference is 0.8492. Middle: output after one iteration with DGD, maximal value of difference is 0.6171. Right: result after 5 iterations of DGD, maximal value of difference is 0.4124.

section. Furthermore, it is robust in the transition to real measurement data, which is one of the most important aspects in inverse problems and image reconstruction.

During the reconstruction procedure, a major improvement is achieved in the first step, as shown in Figure 6. After one iteration of the DGD the background is cleared and the contrast is mostly restored, but there are still a few noisy patches around the vessels visible. The difference image also indicates that there are still parts insufficiently recovered on the outer area close to the boundary; these are typical limited view artefacts. After the 5th iteration these artefacts are considerably reduced and the error inside the domain is mostly uniform.

In the following, we discuss some particular aspects in more detail.

A. *Quantitative analysis of simulated data*

For a quantitative evaluation of the performance we have computed the relative ℓ^2-error for the simulated example shown in this study, see e.g. Figure 4. More precisely the reconstruction quality is evaluated using a scaled and unbiased relative error defined by

$$\mathrm{err}(x) = \min_{a,b} \frac{\|ax - x_{true} - b\|_2}{\|x_{true}\|_2}, \tag{11}$$

as suggested in [20]. This unbiased error is used to not disadvantage TV and NNLS reconstructions in the comparison. While the networks know the absolute contrast from the training data, classical iterative methods often either need many iterations to recover it from the data or suffer from systematic contrast errors. Consequently, the optimal parameters for the reconstructions of DGD and U-Net are in most cases $a = 1$ and $b = 0$ and hence err reduces to the standard relative ℓ^2-error. For a full comparison, we have computed the mean error for 16 test samples that were not included in the training set. We compare the two networks, U-net and DGD, with TV and NNLS reconstructions, as described in Section II, with the regularisation parameter for TV chosen such that $\mathrm{err}(x)$ is minimized. The resulting errors are plotted in Figure 7. After one iteration U-Net achieves clearly the best result, but already with 2 iterations DGD achieves a smaller error down to a substantially smaller error after 5 iterations. TV and NNLS converge considerably slower, but achieve the U-Net quality after 50 iterations and will likely go lower.

The computational time is dominated by the application of A and its adjoint A^* . Computing either takes about 12 seconds on the Titan Xp GPU, see Table I for the complete computation times for each reconstruction approach. Note however, that as our implementations involve communication overhead between Matlab and Python, theses timings give an indication for the methods' efficiency rather than an absolute comparison. Consequently, a reduction in iterations has a considerable impact on the total computation time. In this respect, the U-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

Net structure is clearly the cheapest with just one application to compute $x_0 = A^*y$. Iterative algorithms require additionally two applications for each iterate to compute the gradient $\nabla d(y, Ax_k) = A^*(Ax_k - y)$. Thus, having similar results after 2 iterations with DGD and 50 iterations of TV, see Figure 7, leads to a prospective speed-up by 20 (including the initial reconstruction $x_0 = A^*y$). We note that the computation time for U-Net can be considerably reduced by using a k-space method [53] for the initial reconstruction.

TABLE I
EVALUATION TIMES: INCLUDING INITIALIZATION $x_0 = A^*y$ AND COMMUNICATION OVERHEAD FOR DGD AND U-NET

	DGD 5 iterations	TV 5 iterations	NNLS 5 iterations	U-NET
TIME IN SEC.	184.19	165.72	147.7	19.75

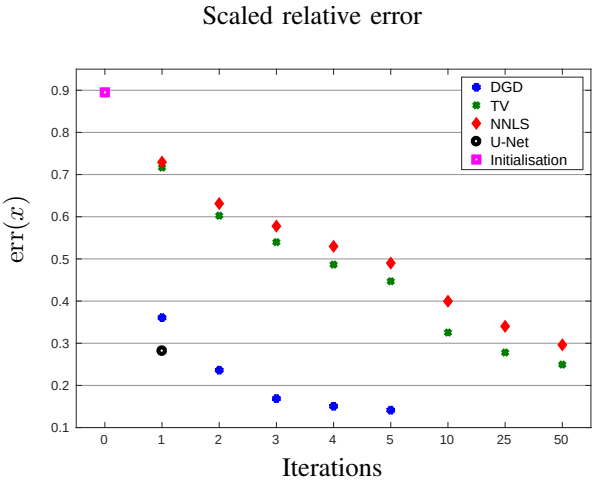


Fig. 7. Convergence plot of mean error for 16 samples from simulated test data. The x-axis shows number of iterations (note the nonlinear scale). The y-axis denotes the unbiased relative ℓ^2-errors (11). The parameter for TV has been chosen such that the best error is achieved for the given iterations.

B. Comparison to post-processing by Deep Learning

First using a direct reconstruction and then applying a DNN to remove artifacts is a valid approach in many applications, especially if one is interested in fast and prospectively real-time reconstructions. This approach only needs an initial direct reconstruction and one application of the trained network. Especially for full-view data, this is a promising approach, but even in our limited-view case this approach proves to be quite powerful. A comparison of DGD and U-Net for simulated data is shown in Figure 8 (top row). The resulting image is cleaned up and many vessels are properly reconstructed. Some smaller details are missing and can not be recovered from the initial reconstruction. The difference to the true target is also shown in Figure 8 (bottom row). The differences are most pronounced in the outer parts of the domain as a consequence of the limited view geometry. In comparison the reconstruction by DGD has a much smaller overall error, but this is especially true in the center of the domain. The maximal error of the U-net reconstruction is 0.6012 (on the scale of $[0, 1]$) and of the DGD reconstruction 0.4081. In conclusion we can say that the U-net architecture performs very well and is even capable of removing some limited-view artefacts, but is ultimately limited by the information contained in the initial reconstruction.

C. In-vivo data

Even though the results for simulated data are very impressive, applying the DGD trained on images with a clean background is not sufficient for real data as shown in Figure 4. The reason is that the algorithm interprets all structures in the data as important and enhances them equally. Adding a background to the training data set in order to teach the DGD thresholding those structures immensely improves the results and even fine details that were not visible before are now recovered after 5 iterations, as seen in Figure 5. Nevertheless, just an adjustment of the simulated data is not sufficient as can be seen from the quantitative measures in Table II, computed with respect to the reference reconstruction from fully-sampled limited-view data. Thus, further improvement can be achieved by an update of the DGD if one has a set of similar measurements from fully sampled data available. This update training has a considerable impact on the reconstruction quality as can been seen in Table II. Both learned methods show excellent reconstruction quality after transfer training and are able to successfully remove the undesired background structures. In comparison to the iterative reconstruction with TV both learned methods achieve a higher PSNR and SSIM to the reference reconstruction from fully-sampled data. Noteworthy, the lowest unbiased relative ℓ^2-error (err), see (11), is achieved by the classical TV minimisation with an emphasis on the data term, this is likely due to the fact that the reference is a TV reconstruction from fully-sampled data.

TABLE II
QUANTITATIVE MEASURES FOR IN-VIVO EXPERIMENT: IN COMPARISON TO REFERENCE TV RECONSTRUCTION FROM FULLY-SAMPLED LIMITED-VIEW DATA.

	PSNR	SSIM	ERR	REL. ℓ^2-ERROR
DGD x_5	32.93	0.723	0.76	1.54
UPDATED DGD $\hat{x}_5$	**41.40**	**0.945**	0.56	**0.58**
U-NET	40.81	0.933	0.62	0.62
TV SUB-SAMP., $\lambda = 5 \cdot 10^{-5}$	38.05	0.912	**0.52**	0.86
TV SUB-SAMP., $\lambda = 10^{-4}$	37.68	0.902	0.58	0.89

D. Generalisation and robustness

Deep Learning approaches are especially powerful in a fixed measurement protocol and consistent targets, as illustrated for the simulated test data. The big question is how robust these networks are with respect to perturbations of measurement procedures or targets. First experiments indicate that the iterative network allows for small perturbations in the forward operator such as varying sub-sampling patterns (of same sub-sampling rate) or deviations in sound speed, as well as slightly varying noise level in the data. However, each variation will lead to slight deterioration of reconstruction quality. In contrast, the one step approach by U-Net was found much more sensitive to variations. In particular, we have found that a change in sampling pattern leads to a mean (for 16

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

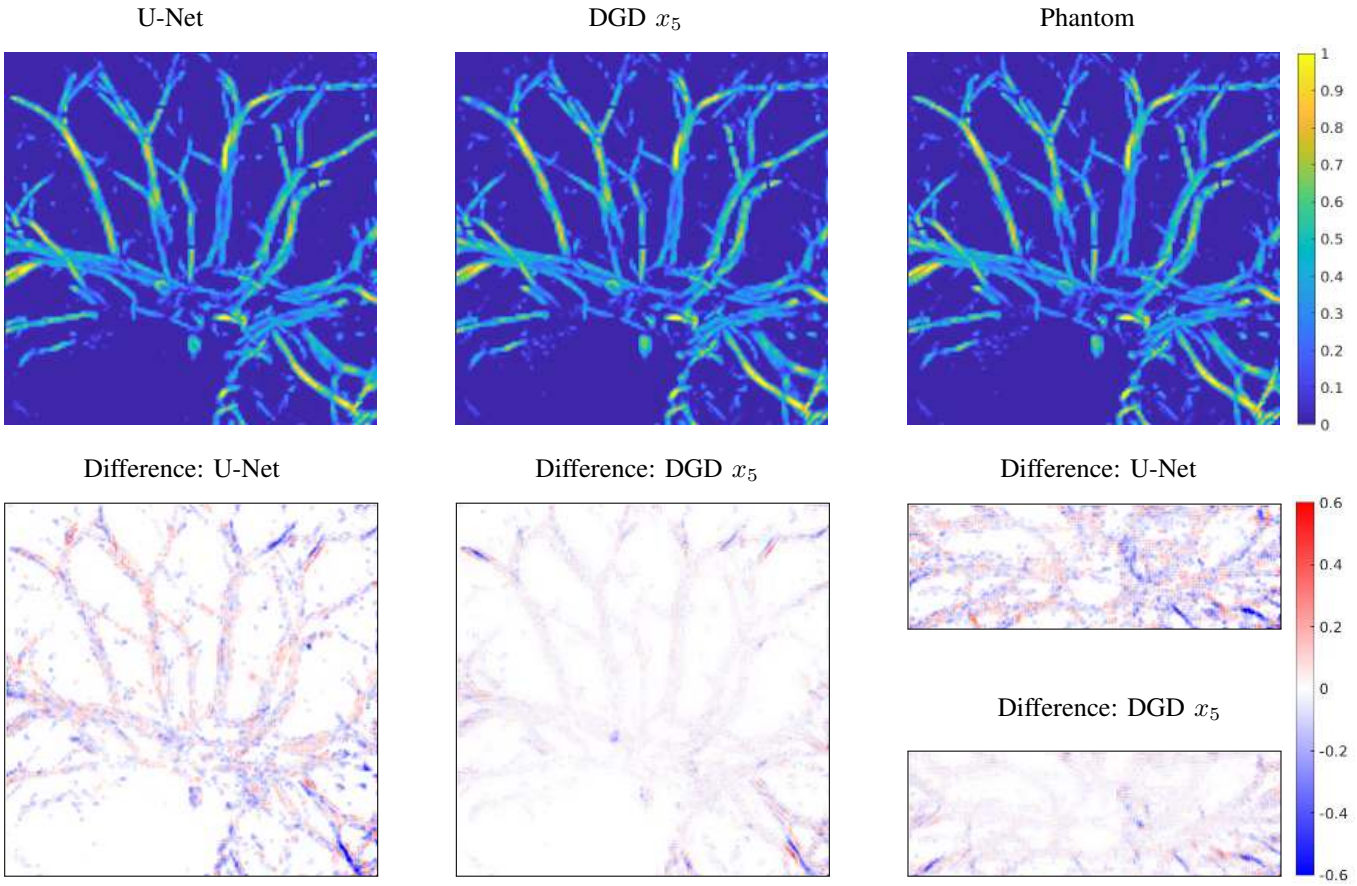


Fig. 8. Comparison of reconstructions for a test image from the segmented CT data. Images are top-down maximum intensity projections and the difference images are on the same scale, with blue for a negative difference and red for positive.. Left: top and bottom shows the result by applying U-Net to the initialization x_0 and the difference to the phantom, maximal value of difference is 0.6012. Middle: shows the result of the DGD after 5 iterations and the difference to the phantom, maximal value of difference is 0.4081. Right bottom: difference images as side projections for the results of DGD and U-Net.

samples) deterioration in err by 0.5% for DGD and U-Net by 5% for the simulated test data. We think that this is due to the fact, that the gradient in each iteration encodes the model variations and as such small perturbations are corrected in the iterative network. If larger changes in the measurement protocol are expected, it is recommended to either retrain the network or perform an update training, as has been done for the in-vivo data.

Furthermore, the iterative method seems to be more robust with respect to structural differences between the target and the training set. This is illustrated in Figure 9, where we have tested the networks trained on the clean segmented vessels on a tumor phantom [13]. With 5 iterations we achieve a similar err as TV after 20 iterations. As it can be seen, the network does reproduce vessels with similar characteristic as in the training set, this might be due to the learned prior-like filters. Whereas the U-Net reconstruction does not perform well with the new image structures.

VI. Conclusions

In limited-view, sub-sampled photoacoustic tomography it is essential to incorporate the physical model into the reconstruction procedure to reduce artefacts with an appropriate regularisation strategy. Here we considered three possible strategies: i) iterative total variation, ii) backprojection followed by a learned denoiser, iii) learned iterative reconstruction. In terms of image quality and robustness to perturbations in the model i) and iii) were superior to ii). Method ii) was fastest at the cost of inferior image quality and flexibility. Method iii) was considerably faster than i). Thus, we believe that learned iterative reconstructions are a realistic technique for 3D PAT. The choice between learned post-processing versus learned iterative reconstruction is a matter of speed versus quality.

This study is particularly focused on method iii) and we have shown that incorporating the physical model as the gradient of the data fit and learning an iterative algorithm consisting of several convolutional neural networks leads to a superior reconstruction quality with a considerable speed-up compared to classical, and well established, iterative reconstruction schemes. With minor modifications we were able to apply the learned algorithm to experimental in-vivo data of a human wrist and obtained far more detailed reconstructions from sub-sampled data than by TV minimisation of the same data.

Additionally, we have investigated method ii) that consists of post-processing a fast and basic direct reconstruction with a CNN, in particular we implemented an architecture introduced as U-Net that has been proven to work well on medical images. In our study this approach shows promise to produce a fast and good initial reconstruction, but since many features are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2820382, IEEE Transactions on Medical Imaging

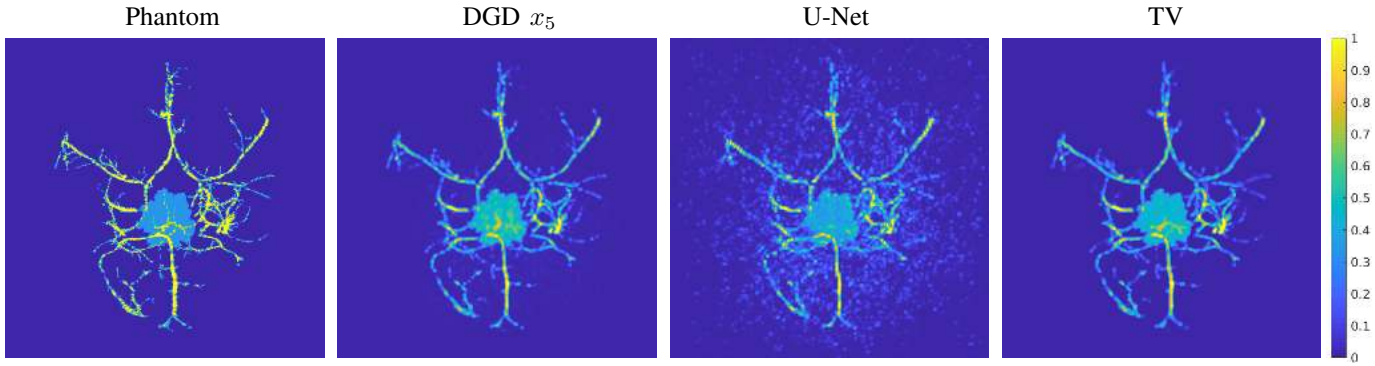


Fig. 9. Reconstruction of a tumor phantom with features that are not included in the training data. DGD and U-Net reconstructions are done with the networks trained on the segmented vessel phantoms. The TV reconstruction is computed with 20 iterations and a regularisation parameter $\lambda = 10^{-4}$. Reconstruction errors with the unbiased err are: DGD 0.4925, U-Net 0.6584, TV 0.4749.

not present in simple direct reconstructions, for limited-view, sub-sampled data, this approach is limited by the quality of the initial reconstruction. Even though certain features can not be recovered, post-processing with Deep Learning is promising for applications where low latency is more important than a best quality reconstruction, such as navigational tasks during surgery. Furthermore, our study suggests that iterative networks are more robust with respect to changes in the measurement setup or imaged target.

As inherent in all learning approaches, the limitation of the proposed method is dictated by the quality of the training data and the possibility to perform an update training. In future research we will consider combing the U-Net architecture with a model based approach. For instance by replacing the CNNs representing one iteration in our *deep gradient descent* with a U-Net like structure. For high resolution 3D imaging this would need computational resources exceeding a local workstation. Consequently, if the computational resources are available including the forward operator in the training will likely improve results even further.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. This study was supported in part by funding from the European Union's Horizon 2020 research and innovation programme H2020 ICT 2016-2017 under grant agreement No 732411 and is an initiative of the Photonics Public Private Partnership.

AH acknowledges support from the Wellcome-EPSRC project NS/A000027/1, "Image-Guided Intrauterine Minimally Invasive Fetal Diagnosis and Therapy" (GiftSurg). FL acknowledges financial support from EPSRC project EP/K009745/1 "Dynamic High Resolution Photoacoustic Tomography System" and of the Netherlands Organization for Scientific Research (NWO), project nr. 613.009.106/2383. JA acknowledges Swedish Foundation of Strategic Research grants AM13-0049 and ID14-0055 as well as support from Elekta.

REFERENCES

[1] S R Arridge and O Scherzer. Imaging from coupled physics. *Inverse Problems*, 28(8):080201, 2012.

[2] L. V. Wang. Multiscale photoacoustic microscopy and computed tomography. *Nature Photonics*, 3(9):503–509, Sep 2009.

[3] P. Beard. Biomedical photoacoustic imaging. *Interface Focus*, 2011.

[4] Liming Nie and Xiaoyuan Chen. Structural and functional photoacoustic molecular tomography aided by emerging contrast agents. *Chemical Society reviews*, 43(20):7132–70, 2014.

[5] Keerthi S Valluru, Katheryne E Wilson, and Jürgen K Willmann. Photoacoustic imaging in oncology: translational preclinical and early clinical experience. *Radiology*, 280(2):332–349, 2016.

[6] Yong Zhou, Junjie Yao, and Lihong V Wang. Tutorial on photoacoustic tomography. *Journal of biomedical optics*, 21(6):061007, 2016.

[7] J. Xia and L. V. Wang. Small-animal whole-body photoacoustic tomography: A review. *IEEE Transactions on Biomedical Engineering*, 61(5):1380–1389, May 2014.

[8] P Kuchment and L Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In O Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 817–865. Springer New York, 2011.

[9] LV Wang Y Xu, G Ambartsoumian, and P Kuchment. Reconstructions in limited-view thermoacoustic tomography. *Med Phys*, 31:724, 2004.

[10] Markus Haltmeier and Linh V. Nguyen. Analysis of iterative methods in photoacoustic tomography with variable sound speed. *SIAM Journal on Imaging Sciences*, 10(2):751–781, 2017.

[11] Chao Huang, Kun Wang, Liming Nie, L.V. Wang, and M.A. Anastasio. Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media. *Medical Imaging, IEEE Transactions on*, 32(6):1097–1110, June 2013.

[12] S.R. Arridge, M.M. Betcke, B.T. Cox, F. Lucka, and B.E. Treeby. On the adjoint operator in photoacoustic tomography. *Inverse Problems*, 32(11):115012, 2016.

[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. *Physics in Medicine and Biology*, 61(24):8908, 2016.

[14] N. Huynh, F. Lucka, E.Z. Zhang, M.M. Betcke, S.R. Arridge, P. Beard, and B.T. Cox. Sub-sampled Fabry-Perot photoacoustic scanner for fast 3D imaging. *Proc. SPIE*, 10064:100641Y–5, 2017.

[15] Y. E. Boink, M. J. Lagerwerf, W. Steenbergen, S. A. van Gils, S. Manohar, and C. Brune. A Reconstruction Framework for Total Generalised Variation in Photoacoustic Tomography. *ArXiv e-prints*, July 2017.

[16] Johannes Schwab, Sergiy Pereverzyev Jr, and Markus Haltmeier. A galerkin least squares approach for photoacoustic tomography. *SIAM Journal on Numerical Analysis*, 56(1):160–184, 2018.

[17] S. Zackrisson, S. M W Y van de Ven, and S. S. Gambhir. Light In and Sound Out: Emerging Translational Strategies for Photoacoustic Imaging. *Cancer Research*, 74(4):979–1004, 2014.

[18] Amit P. Jathoul, Jan Laufer, Olumide Ogunlade, Bradley Treeby, Ben Cox, Edward Zhang, Peter Johnson, Arnold R. Pizzey, Brian Philip, Teresa Marafioti, Mark F. Lythgoe, R. Barbara Pedley, Martin A. Pule, and Paul Beard. Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter. *Nature Photon*, 9(4):239–246, 2015.

[19] Eunhee Kang, Junhong Min, and Jong Chul Ye. A deep convolutional

neural network using directional wavelets for low-dose x-ray ct reconstruction. *Medical physics*, 44(10), 2017.

[20] K H Jin, M T McCann, E Froustey, and M Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[21] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. Deep learning for photoacoustic tomography from sparse data. *arXiv preprint arXiv:1704.04587*, 2017.

[22] C M Sandino, N Dixit, J Y Cheng, and S S Vasanawala. Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging. *preprint*.

[23] Jonas Adler and Ozan Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007, 2017.

[24] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 2018.

[25] Hu Chen, Yi Zhang, Weihua Zhang, Huaiqiaing Sun, Peixi Liao, Kun He, Jiliu Zhou, and Ge Wang. Learned Experts' Assessment-based Reconstruction Network ("LEARN") for Sparse-data CT. *arXiv preprint arXiv:1707.09636*, 2017.

[26] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic Resonance in Medicine*, 2017.

[27] Brendan Kelly, Thomas P Matthews, and Mark A Anastasio. Deep learning-guided image reconstruction from incomplete data. *arXiv preprint arXiv:1709.00584*, 2017.

[28] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, Apr. 2006.

[29] Emmanuel J. Candés, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[30] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, Mar. 2008.

[31] J. Provost and F. Lesage. The application of compressed sensing for photo-acoustic tomography. *IEEE Transactions on Medical Imaging*, 28(4):585–594, April 2009.

[32] Zijian Guo, Changhui Li, Liang Song, and Lihong V Wang. Compressed sensing in photoacoustic tomography in vivo. *J. Biomed. Opt.*, 15(2):021311, 2011.

[33] Markus Haltmeier, Thomas Berer, Sunghwan Moon, and Peter Burgholzer. Compressed sensing and sparsity in photoacoustic tomography. *Journal of Optics*, 18(11):114004, 2016.

[34] M. Betcke, B. Cox, N. Huynh, E. Zhang, P. Beard, and S. R. Arridge. Acoustic wave field reconstruction from compressed measurements with application in photoacoustic tomography. *IEEE Transactions on Computational Imaging*, PP(99):1–1, 2017.

[35] Kun Wang and MarkA. Anastasio. Photoacoustic and thermoacoustic tomography: Image formation principles. In Otmar Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 781–815. Springer New York, 2011.

[36] Christian Lutzweiler and Daniel Razansky. Optoacoustic imaging and tomography: Reconstruction approaches and outstanding challenges in image performance and quantification. *Sensors*, 13(6):7345, 2013.

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

[38] B. T. Cox, S. Kara, S. R. Arridge, and P. C. Beard. k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics. *The Journal of the Acoustical Society of America*, 121(6), 2007.

[39] Bradley E Treeby, Edward Z Zhang, and B T Cox. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Problems*, 26(11):115003, 2010.

[40] Bradley E. Treeby and B. T. Cox. k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of Biomedical Optics*, 15(2):021314–021314–12, 2010.

[41] Jrgen Frikel and Eric Todd Quinto. Artifacts in incomplete data tomography with applications to photoacoustic tomography and sonar. *SIAM Journal on Applied Mathematics*, 75(2):703–725, 2015.

[42] O Scherzer, M Grasmair, H Grossauer, M Haltmeier, and F Lenzen. *Variational methods in imaging*, volume 320. Springer, 2009.

[43] Martin Burger and Stanley Osher. A Guide to the TV Zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging*, Lecture Notes in Mathematics, pages 1–70. Springer International Publishing, 2013.

[44] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*, pages 3981–3989, 2016.

[45] Patrick Putzky and Max Welling. Recurrent inference machines for solving inverse problems. *arXiv preprint arXiv:1706.04008*, 2017.

[46] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[47] R Manniesing and W Niessen. Multiscale vessel enhancing diffusion in ct angiography noise filtering. In *Biennial Inter. Conf. on Information Proc. in Med. Imaging*, pages 138–149. Springer, 2005.

[48] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.

[49] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[50] Rene Vidal, Joan Bruna, Raja Giryes, and Stefano Soatto. Mathematics of deep learning. *arXiv preprint arXiv:1712.04741*, 2017.

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[52] Edward Zhang, Jan Laufer, and Paul 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*, 47(4):561–577, 2008.

[53] Kornel P Köstli, Martin Frenz, Hans Bebie, and Heinz P Weber. Temporal backward projection of optoacoustic pressure transients using fourier transform methods. *Physics in medicine and biology*, 46(7):1863, 2001.