

Nr. 33  
19. April 2017

Leopold-Franzens-Universität Innsbruck



Preprint-Series: Department of Mathematics - Applied Mathematics

## Deep Learning for Photoacoustic Tomography from Sparse Data

S. Antholzer, M. Haltmeier and J. Schwab

REVISED VERSION, AUGUST 2018



---

Technikerstraße 13 - 6020 Innsbruck - Austria  
Tel.: +43 512 507 53803 Fax: +43 512 507 53898  
<https://applied-math.uibk.ac.at>

# Deep Learning for Photoacoustic Tomography from Sparse Data

Stephan Antholzer

Markus Haltmeier

Johannes Schwab

Department of Mathematics, University of Innsbruck

Technikerstrasse 13, A-6020 Innsbruck, Austria

Corresponding e-mail: [markus.haltmeier@uibk.ac.at](mailto:markus.haltmeier@uibk.ac.at)

## Abstract

The development of fast and accurate image reconstruction algorithms is a central aspect of computed tomography. In this paper, we investigate this issue for the sparse data problem in photoacoustic tomography (PAT). We develop a direct and highly efficient reconstruction algorithm based on deep learning. In our approach image reconstruction is performed with a deep convolutional neural network (CNN), whose weights are adjusted prior to the actual image reconstruction based on a set of training data. The proposed reconstruction approach can be interpreted as a network that uses the PAT filtered backprojection algorithm for the first layer, followed by the U-net architecture for the remaining layers. Actual image reconstruction with deep learning consists in one evaluation of the trained CNN, which does not require time consuming solution of the forward and adjoint problems. At the same time, our numerical results demonstrate that the proposed deep learning approach reconstructs images with a quality comparable to state of the art iterative approaches for PAT from sparse data.

**keywords** Photoacoustic tomography, sparse data, image reconstruction, deep learning, convolutional neural networks, inverse problems.

## 1 Introduction

Deep learning is a rapidly emerging research area that has significantly improved performance of many pattern recognition and machine learning applications [23, 50]. Deep learning algorithms make use of special artificial neural network designs for representing a nonlinear input to output map together with optimization procedures for adjusting the weights of the network during the training phase. Deep learning techniques are currently the state of the art for visual object recognition, natural language understanding or applications in other domains such as drug discovery or biomedical image analysis (see, for example, [5, 14, 27, 41, 45, 52, 54, 75] and the references therein).

Despite its success in various scientific disciplines, in image reconstruction deep learning research appeared only very recently (see [12, 37, 44, 71, 74, 76, 80]). In this paper, we develop

a deep learning framework for image reconstruction in photoacoustic tomography (PAT). To concentrate on the main ideas we restrict ourselves to the sparse data problem in PAT in a circular measurement geometry. Our approach can be extended to an arbitrary measurement geometry in arbitrary dimension. Clearly, the increased dimensionality comes with an increased computational cost. This is especially the case for the training of the network which, however, is done prior to the actual image reconstruction.

## 1.1 PAT and the sparse sampling problem

PAT is a non-invasive coupled-physics biomedical imaging technology which beneficially combines the high contrast of pure optical imaging with the high spatial resolution of ultrasound imaging [7, 46, 59, 73]. It is based on the generation of acoustic waves by illuminating a semi-transparent biological or medical object with short optical pulses. The induced time dependent acoustic waves are measured outside of the sample with acoustic detectors, and the measured data are used to recover an image of the interior (see Figure 1). High spatial resolution in PAT can be achieved by measuring the acoustic data with high spatial and temporal sampling rate [33, 56]. While temporal samples can be easily collected at or above the Nyquist rate, the spatial sampling density is usually limited [3, 6, 26, 29, 62, 63]. In fact, each spatial measurement requires a separate sensor and high quality detectors are often costly. Moving the detector elements can increase the spatial sampling rate, but is time consuming and also can introduce motion artifacts. Therefore, in actual applications, the number of sensor locations is usually small compared to the desired resolution which yields to a so-called sparse data problem.



Figure 1: BASIC PRINCIPLE OF PAT. An object is illuminated with a short optical pulse (left) that induces an acoustic pressure wave (middle). The pressure signals are recorded outside of the object, and are used to recover an image of the interior (right).

Applying standard algorithms to sparse data yields low-quality images containing severe undersampling artifacts. To some extent, these artifacts can be reduced by using iterative image reconstruction algorithms [4, 8, 34, 40, 43, 66, 72] which allow to include prior knowledge such as smoothness, sparsity or total variation (TV) constraints [1, 15, 20, 25, 60, 64]. These algorithms tend to be time consuming as the forward and adjoint problems have to be solved repeatedly. Further, iterative algorithms have their own limitations. For example, the reconstruction quality strongly depends on the used a-priori model about the objects

to be recovered. For example, TV minimization assumes sparsity of the gradient of the image to be reconstructed. Such assumptions are often not strictly satisfied in real world scenarios which again limits the theoretically achievable reconstruction quality.

To overcome the above limitations, in this paper we develop a new reconstruction approach based on deep learning that comes with the following properties: (i) image reconstruction is efficient and non-iterative; (ii) no explicit a-priori model for the class of objects to be reconstructed is required; (iii) it yields a reconstruction quality comparable to (or even outperforming) existing methods for sparse data. Note that instead of providing an explicit a-priori model, the deep learning approach requires a set of training data and the CNN itself adjusts to the provided training data. By training the network on real word data, it thereby automatically creates a model of the considered PAT images in an implicit and purely data driven manner. While training of the CNN again requires time consuming iterative minimization, we stress that training is performed independent of the particular investigated objects and prior to the actual image reconstruction. Additionally, if the time resources available for training a new network are limited, then one can use weights learned on one data set as good starting value for training the weights in the new network.

## 1.2 Proposed deep learning approach

Our reconstruction approach for the sparse data problem in PAT uses a deep convolutional neural network in combination with any linear reconstruction method as preprocessing step. Essentially, it consists of the following two steps (see Figure 2):

- (D1) In the first step, a linear PAT image reconstruction algorithm is applied, which yields an approximation of the original object including under-sampled artifacts.
- (D2) In the second step, a deep convolutional neural network (CNN) is applied to map the intermediate reconstruction from (D1) to an artifact free final image.

Note that the above two-stage procedure can be viewed as a single deep neuronal network that uses the linear reconstruction algorithm in (D1) for the first layer, and the CNN in (D2) for the remaining layers.

Step (D1) can be implemented by any standard linear reconstruction algorithm including filtered backprojection (FBP) [10, 18, 19, 31, 32, 48, 77], Fourier methods [2, 35, 42, 49, 79], or time reversal [11, 39, 67, 69]. In fact, all these methods can be implemented efficiently using at most  $\mathcal{O}(d^3)$  floating point operations (FLOPS) for reconstructing a high-resolution image on an  $d \times d$  grid. Here  $d$  is number spatial discretization points along one dimension of the reconstructed image. The CNN applied in step (D2) depends on weights that are adjusted using a set of training data to achieve artifact removal. The weights in the CNN are adjusted during the so-called training phase which is performed prior to the actual image reconstruction [23]. In our current implementation, we use the U-net architecture originally designed in [61] for image segmentation. Application of the trained network for image reconstruction is fast. One application of the U-net requires  $\mathcal{O}(F^2 L d^2)$  FLOPS, where  $F$  is the number of channels in the first convolution and  $L$  describes the depth of the network. Typically,  $F^2 L$  will be in the order of  $d$  and the number of FLOPS for evaluating the CNN is comparable to the effort of performing an FBP reconstruction. Moreover,

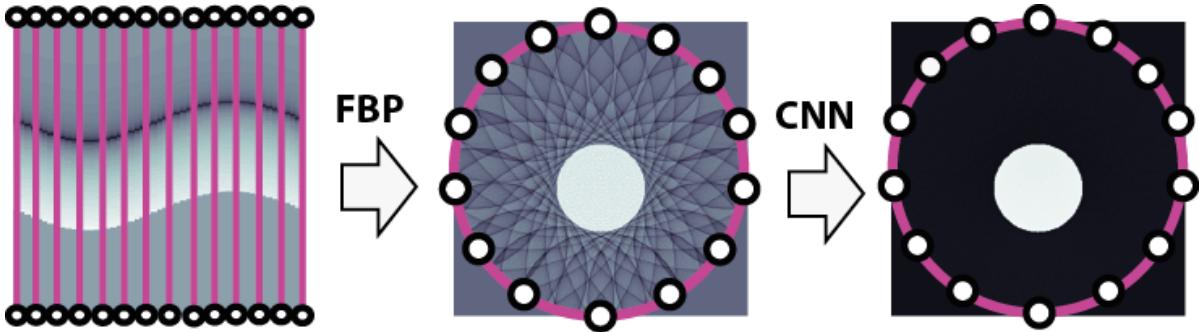


Figure 2: ILLUSTRATION OF THE PROPOSED NETWORK FOR PAT IMAGE RECONSTRUCTION. In the first step, the FBP algorithm (or another standard linear reconstruction method) is applied to the sparse data. In a second step, a deep convolutional neural network is applied to the intermediate reconstruction which outputs an almost artifact free image. This may be interpreted as a deep network with the FBP in the first layer and the CNN in the remaining layers.

evaluation of the CNN can easily be parallelized, which further increases numerical performance. On the other hand, iterative reconstruction algorithms tend to be slower as they require repeated application of the PAT forward operator and its adjoint.

To the best of our knowledge, this is the first paper using deep learning or neural networks for PAT. Related approaches applying CNNs for different medical imaging technologies including computed tomography (CT) and magnetic resonance imaging (MRI) appeared recently in [12, 37, 44, 71, 74, 76, 80]. The author of [71] shares his opinions on deep learning for image reconstruction. In [44], deep learning is applied to imaging problems where the normal operator is shift invariant; PAT does not belong to this class. A different learning approach for addressing the limited view problem in PAT is proposed in [17]. The above references show that a significant amount of research has been done on deep learning for CT and MRI image reconstruction (based on inverse Radon and inverse Fourier transforms). Opposed to that, PAT requires inverting the wave equation, and our work is the first paper that used deep learning and CNNs for PAT reconstruction and inversion of the wave equation.

### 1.3 Outline

The rest of this paper is organized as follows. In Section 2 we review PAT and discuss the sparse sampling problem. In Section 3 we describe the proposed deep learning approach. For that purpose, we discuss neural networks and present CNNs and the U-net actually implemented in our approach. Details on the numerical implementation and numerical results are presented in Section 4. The paper concludes with a short summary and outlook given in Section 5.

## 2 Photoacoustic tomography

As illustrated in Figure 1, PAT is based on generating an acoustic wave inside some investigated object using short optical pulses. Let us denote by  $h: \mathbb{R}^d \rightarrow \mathbb{R}$  the initial pressure distribution which provides diagnostic information about the patient and which is the quantity of interest in PAT. Of practical relevance are the cases  $d = 2, 3$  (see [47, 59, 78]). For keeping the presentation simple and focusing on the main ideas we only consider the case of  $d = 2$ . Among others, the two-dimensional case arises in PAT with so called integrating line detectors [10, 59]. Extensions to three spatial dimensions are possible by using the FBP algorithm for 3D PAT [19] in combination with the 3D U-net designed in [13]). Further, we restrict ourselves to the case of a circular measurement geometry, where the acoustic measurements are made on a circle surrounding the investigated object. In general geometry, one can use the so-called universal backprojection formula [32, 77] that is exact for general geometry up to an additive smoothing term [32]. In this case, the CNN can be used to account for the under-sampling issue as well as to account the additive smooth term. Such investigations, however, are beyond the scope of this paper.

### 2.1 PAT in circular measurement geometry

In two spatial dimensions, the induced pressure in PAT satisfies the following initial value problem for the 2D wave equation

$$\begin{cases} \partial_t^2 p(x, t) - \Delta p(x, t) = 0 & \text{for } (x, t) \in \mathbb{R}^2 \times (0, \infty) \\ p(x, 0) = h(x) & \text{for } x \in \mathbb{R}^2 \\ \partial_t p(x, 0) = 0 & \text{for } x \in \mathbb{R}^2, \end{cases} \quad (2.1)$$

where we assume a constant sound-speed that is rescaled to one. In the circular measurement geometry, the initial pressure  $h$  is assumed to vanish outside the disc  $B_R := \{x \in \mathbb{R}^2 \mid \|x\| < R\}$ . Note that the solution of used forward wave equation (2.1) is, for positive times, equal to the causal solution of the wave equation with source term  $\delta'(t)h(x)$ ; see [36]. Both models (either with source term or with initial condition) are frequently used in PAT. The goal of PAT image reconstruction is to recover  $h$  from measurements of the acoustic pressure  $p$  made on the boundary  $\partial B_R$ .

In a complete data situation, PAT in a circular measurement geometry consist in recovering the function  $h$  from data

$$(\mathbf{P}h)(z, t) := p(z, t) \quad \text{for } (z, t) \in \partial B_R \times [0, T], \quad (2.2)$$

where  $p$  denotes the solution of (2.1) with initial data  $h$  and  $T$  is the final measurement time. Here complete data refers to data prior to sampling that are known on the full boundary  $\partial B_R$  and up to times  $T \geq 2R$ . In such a case, exact and stable PAT image reconstruction is theoretically possible; see [34, 68]. Several efficient methods for recovering  $h$  from complete data  $\mathbf{P}h$  are well investigated (see, for example, [2, 10, 11, 18, 19, 31, 32, 35, 39, 42, 48, 49, 67, 69, 77, 79]). As an example, we mention the FBP formula derived in [18],

$$h(r) = -\frac{1}{\pi R} \int_{\partial B_R} \int_{|r-z|}^{\infty} \frac{(\partial_t t \mathbf{P}h)(z, t)}{\sqrt{t^2 - |r - z|^2}} dt dS(z). \quad (2.3)$$

Note that (2.3) requires data for all  $t > 0$ ; see [18, Theorem 1.4] for a related FBP formula that only uses data for  $t < 2R$ . For the numerical results in this paper we truncate (2.3) at  $t = 2R$ , in which situation all singularities of the initial pressure are contained in the reconstructed image and the truncation error is small.

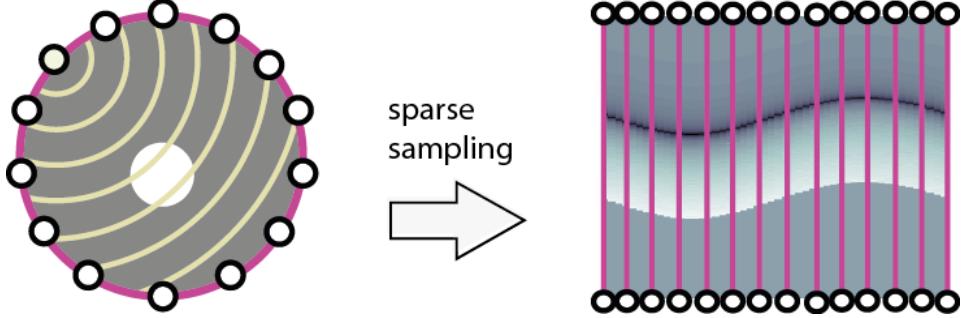


Figure 3: SPARSE SAMPLING PROBLEM IN PAT IN CIRCULAR GEOMETRY. The induced acoustic pressure is measured at  $M$  detector locations on the boundary of the disc  $B_R$  indicated by white dots in the left image. Every detector at location  $z_m$  measures a time dependent pressure signal  $\mathbf{p}[m, \cdot]$ , corresponding to a column in the right image.

## 2.2 Discretization and sparse sampling

In practical applications, the acoustic pressure  $\mathbf{P}h$  can only be measured with a finite number of acoustic detectors. The standard sampling scheme for PAT in circular geometry assumes uniformly sampled values

$$\mathbf{p}[m, \cdot] := \mathbf{P}h(z_m, \cdot), \quad \text{for } m = 1, \dots, M, \quad (2.4)$$

$$\text{with } z_m := \begin{bmatrix} R \cos(2\pi(m-1)/M) \\ R \sin(2\pi(m-1)/M) \end{bmatrix}. \quad (2.5)$$

Here  $\mathbf{p}[m, \cdot]: [0, T] \rightarrow \mathbb{R}$  is the signal corresponding to the  $m$ th detector, and  $M$  is the total number of detector locations. Of course, in practice also the signals  $\mathbf{p}[m, \cdot]$  have to be represented by discrete samples. However, temporal samples can easily be collected at a high sampling rate compared to the spatial sampling, where each sample requires a separate sensor.

In the case that a sufficiently large number of detectors is used, according to Shannon's sampling theory, implementations of full data methods yield almost artifact free reconstructions (for a detailed analysis of sampling in PAT see [33]). As the fabrication of an array of detectors is demanding, experiments using integrating line detectors are often carried out using a single line detector, scanned on circular paths using scanning stages [28, 57], which is very time consuming. Recently, systems using arrays of 64 parallel line detectors have been demonstrated [6, 26]. To keep production costs low and to allow fast imaging the number  $M$  will typically be kept much smaller than advised by Shannon's sampling theory and one deals with highly under-sampled data.

Due to the high frequency information contained in time, there is still hope to recover high resolution images from spatially under-sampled data. For example, iterative algorithms, using TV minimization yield good reconstruction results from undersampled data

(see [3, 29, 55, 64]). However, such algorithms are quite time consuming as they require evaluating the forward and adjoint problem repeatedly (for TV typically at least several hundreds of times). Moreover, the reconstruction quality depends on certain a-priori assumptions on the class of objects to be reconstructed such as sparsity of the gradient. Image reconstruction with a trained CNN is direct and requires a smaller numerical effort compared to iterative methods. Further, it does not require an explicit model for the prior knowledge about the objects to be recovered. Instead, such a model is implicitly learned in a data-driven manner based on the training data by adjusting the weights of the CNN to the provided training data during the training phase.

### 3 Deep learning for PAT image reconstruction

Suppose that sparsely sampled data of the form (2.4), (2.5) are at our disposal. As illustrated in Figure 2 in our deep learning approach we first apply a linear reconstruction procedure to the sparsely sampled data  $(\mathbf{p}[m, \cdot])_{m=1}^M$  which outputs a discrete image  $\mathbf{X} \in \mathbb{R}^{d \times d}$ . According to Shannon's sampling theory an aliasing free reconstruction requires  $M \geq \pi d$  detector positions [33]. However, in practical applications we will have  $M \ll d$ , in which case severe undersampling artifacts appear in the reconstructed image. To reduce these artifacts, we apply a CNN to the intermediate reconstruction which outputs an almost artifact free reconstruction  $\mathbf{Y} \in \mathbb{R}^{d \times d}$ . How to implement such an approach is described in the following.

#### 3.1 Image reconstruction by neural networks

The task of high resolution image reconstruction can be formulated as supervised machine learning problem. In that context, the aim is finding a restoration function  $\Phi: \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$  that maps the input image  $\mathbf{X} \in \mathbb{R}^{d \times d}$  (containing undersampling artifacts) to the output image  $\mathbf{Y} \in \mathbb{R}^{d \times d}$  which should be almost artifact free. For constructing such a function  $\Phi$ , one assumes that a family of training data  $\mathcal{T} := (\mathbf{X}_n, \mathbf{Y}_n)_{n=1}^N$  are given. Any training example  $(\mathbf{X}_n, \mathbf{Y}_n)$  consist of an input image  $\mathbf{X}_n$  and a corresponding artifact-free output image  $\mathbf{Y}_n$ . The restoration function is constructed in such a way that the training error

$$E(\mathcal{T}; \Phi) := \sum_{n=1}^N d(\Phi(\mathbf{X}_n), \mathbf{Y}_n) \quad (3.1)$$

is minimized, where  $d: \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d} \rightarrow \mathbb{R}$  measures the error made by the function  $\Phi$  on the training examples.

Particular powerful supervised machine learning methods are based on neural networks (NNs). In such a situation, the restoration function is taken in the form

$$\Phi_{\mathcal{W}} = (\sigma_L \circ W_L) \circ \cdots \circ (\sigma_1 \circ W_1), \quad (3.2)$$

where any factor  $\sigma_\ell \circ W_\ell$  is the composition of a linear transformation (or matrix)  $W_\ell \in \mathbb{R}^{D_{\ell+1} \times D_\ell}$  and a nonlinearity  $\sigma_\ell: \mathbb{R} \rightarrow \mathbb{R}$  that is applied component-wise. Here  $L$  denotes the number of processing layers,  $\sigma_\ell$  are so called activation functions and  $\mathcal{W} := (W_1, \dots, W_L)$

is the weight vector. Neural networks can be interpreted to consist of several layers, where the factor  $\sigma_\ell \circ W_\ell$  maps the variables in layer  $\ell$  to the variables in layer  $\ell+1$ . The variables in the first layer are the entries of the input vector  $\mathbf{X}$  and the variables in the last layer are the entries of the output vector  $\mathbf{Y}$ . Note that in our situation we have an equal number of variables  $D_1 = D_{L+1} = d^2$  in the input and the output layer. Approximation properties of NNs have been analyzed, for example, in [21, 38].

The entries of the weight vector  $\mathcal{W}$  are called weights and are the variable parameters in the NN. They are adjusted during the training phase prior to the actual image reconstruction process. This is commonly implemented using gradient descent methods to minimize the training set error [9, 23]

$$E(\mathcal{T}, \mathcal{W}) := E(\mathcal{T}, \Phi_{\mathcal{W}}) = \sum_{n=1}^N d(\Phi_{\mathcal{W}}(\mathbf{X}_n), \mathbf{Y}_n) \quad (3.3)$$

The standard gradient method uses the update rule  $\mathcal{W}^{(k+1)} = \mathcal{W}^{(k)} - \eta \nabla E(\mathcal{T}, \mathcal{W}^{(k)})$ , where  $\nabla E$  denotes the gradient of the error function in the second component and  $\mathcal{W}^{(k)}$  the weight vector in the  $k$ th iteration. In the context of neural networks the update term is also known as error backpropagation. If the number of training examples is large, then the gradient method becomes slow. In such a situation, a popular acceleration is the stochastic gradient descent algorithm [9, 23]. Here for each iteration a small subset  $\mathcal{T}^{(k)}$  of the whole training set is chosen randomly at any iteration and the weights are adjusted using the modified update formula  $\mathcal{W}^{(k+1)} = \mathcal{W}^{(k)} - \eta \nabla E(\mathcal{T}^{(k)}, \mathcal{W}^{(k)})$  for the  $k$ th iteration. In the context of image reconstruction similar acceleration strategies are known as ART or Kaczmarz type reconstruction methods [16, 24, 56]. The number of elements in  $\mathcal{T}^{(k)}$  is called batch size and  $\eta$  is referred to as the learning rate. To stabilize the iterative process, it is common to add a so-called momentum term  $\beta (\mathcal{W}^{(k)} - \mathcal{W}^{(k-1)})$  with some nonnegative parameter  $\beta$  in the update of the  $k$ th iteration.

### 3.2 CNNs and the U-net

In our application, the inputs and outputs are high dimensional vectors. Such large-scale problems require special network designs, where the weight matrices are not taken as arbitrary matrices but take a special form reducing its effective dimensionality. When the input is an image, convolutional neural networks (CNNs) use such special network designs that are widely and successfully used in various applications [9, 51]. A main property of CNNs is the invariance with respect to certain transformations of the input. In CNNs, the weight matrices are block diagonal, where each block corresponds to a convolution with a filter of small support and the number of blocks corresponds to the number of different filters (or channels) used in each layer. Each block is therefore a sparse band matrix, where the non-zero entries of the band matrices determine the filters of the convolution. CNNs are currently extensively used in image processing and image classification, since they outperform most comparable algorithms [23]. They are also the method of choice for the present paper.

There are various CNN designs that can differ in the number of layers, the form of the activation functions and the particular form of the weight matrices  $W_\ell$ . In this paper, we use a particular CNN based on the so-called U-net introduced in [61]. It has been

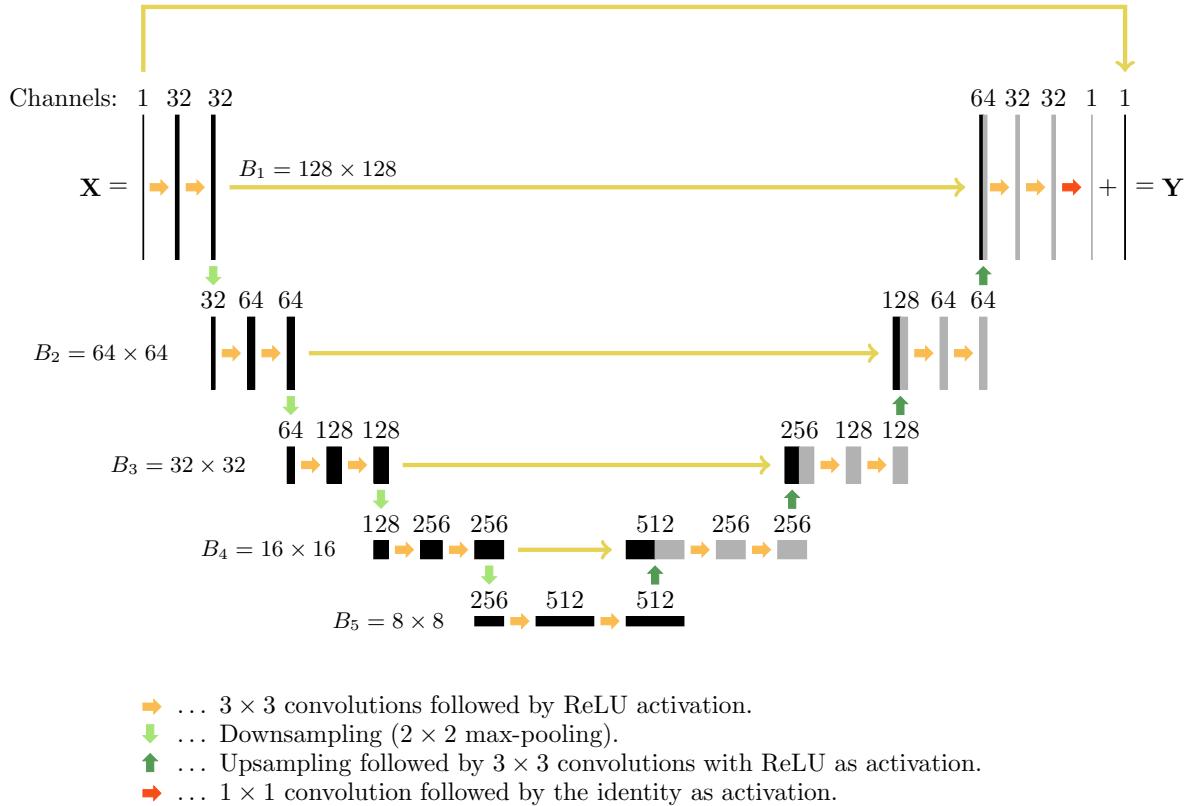


Figure 4: ARCHITECTURE OF THE USED CNN. The number written above each layer denotes the number of convolution kernels (channels), which is equal to number of images in each layer. The numbers  $B_1, \dots, B_5$  denote the dimension of the images (the block sizes in the weight matrices), which stays constant in every row. The long yellow arrows indicate direct connections with subsequent concatenation or summation for the upmost arrow.

originally designed for biomedical image segmentation and recently been used for low dose CT in [37, 44]. The U-net is based on the so-called fully convolutional network used in reference [53]. Such network architectures employ multichannel filtering which means that the weight matrix in every layer consists of a family of multiple convolution filters followed by the rectified linear unit (ReLU) as activation function. The rectified linear unit is defined by  $\text{ReLU}(x) := \max\{x, 0\}$ . As shown in [37], the residual images  $\mathbf{X} - \mathbf{Y}$  often have a simpler structure and are more accessible to the U-net than the original outputs. Therefore, learning the residuals and subtracting them from the inputs after the last layer is more effective than directly training for  $\mathbf{Y}$ . Such an approach is followed in our implementation. The resulting deep neural network architecture is shown in Figure 4.

### 3.3 PAT using FBP combined with the U-net

We are now ready to present the proposed deep learning approach for PAT image reconstruction from sparse data, that uses the FBP algorithm as linear preprocessing step followed by the U-net for removing undersampling artifacts. Recall that we have given sparsely sampled data  $(\mathbf{p}[m, \cdot])_{m=1}^M$  of the form (2.4), (2.5). A discrete high resolution approximation  $\mathbf{Y} \in \mathbb{R}^{d \times d}$  with  $d \gg M$  of the original object is then reconstructed as follows.

- (S1) Apply the FBP algorithm to  $\mathbf{p}$  which yields an reconstruction  $\mathbf{X} \in \mathbb{R}^{d \times d}$  containing undersampling artifacts.
- (S2) Apply the U-net shown in Figure 4 to  $\mathbf{X}$  which yields an image  $\mathbf{Y} \in \mathbb{R}^{d \times d}$  with significantly reduced undersampling artifacts.

The above two steps can also be combined to a single network with the FBP in the first layer and the U-net for the remaining layers. Note that the first step could also be replaced by another linear reconstruction methods such as time reversal and the second step by a different CNN. Such alternative implementations will be investigated in future studies. In this work, we use the FBP algorithm described in [18] for solving step (S1). It is based on discretizing the inversion formula (2.3) by replacing the inner and the outer integration by numerical quadrature and uses an interpolation procedure to reduce the numerical complexity. For details on the implementation we refer to [18, 30].

A crucial ingredient in the above deep learning method is the adjustment of the actual weights in the U-net, which have to be trained on an appropriate training data set. For that purpose we construct training data  $\mathcal{T} = (\mathbf{X}_n, \mathbf{Y}_n)_{n=1}^N$  by first creating certain phantoms  $\mathbf{Y}_n$ . We then simulate sparse data by numerically implementing the well-known solution formula for the wave equation and subsequently construct  $\mathbf{X}_n$  by applying the FBP algorithm of [18] to the sparse data. For training the network we apply the stochastic gradient algorithm for minimizing the training set error (3.1), where we take the error measure  $d$  corresponding to the  $\ell^1$ -norm  $\|\mathbf{Y}\|_1 = \sum_{i_1, i_2=1}^d |\mathbf{Y}[i_1, i_2]|$ .

## 4 Numerical realization

In this section, we give details on the numerical implementation of the deep learning approach and present reconstruction results under various scenarios.

### 4.1 Data generation and network training

For all numerical results presented below we use  $d = 128$  for the image size and take  $R = 1$  for the radius of the measurement curve. For the sparse data in (2.4) we use  $M = 30$  detector locations and discretize the pressure signals  $\mathbf{p}[m, \cdot]$  with 300 uniform samples in the time interval  $[0, 2]$ . In our initial studies, we generate simple phantoms consisting of indicator functions of ellipses with support in the unit cube  $[-1, 1]^2 \subseteq \mathbb{R}^2$ . For that purpose, we randomly generate solid ellipses  $E$  by sampling independent uniformly distributed random variables. The centers are selected uniformly in  $(-0.5, 0.5)$  and the minor and major axes uniformly in  $(0.1, 0.2)$ .

For the training of the network on the ellipse phantoms we generate two different data sets, each consisting of  $N = 1000$  training pairs  $(\mathbf{X}_n, \mathbf{Y}_n)_{n=1}^N$ . One set of training data corresponds to pressure data without noise and for the second data set we added random noise to the simulated pressure data. The outputs  $\mathbf{Y}_n$  consist of the sum of indicator functions of ellipses generated randomly as described above that are sampled on the  $128 \times 128$  imaging grid. The number of ellipses in each training example is also taken randomly according to the uniform distribution on  $\{1, \dots, 5\}$ . The input images are generated numerically by first computing the sparse pressure data using the solution formula for the wave equation and then applying the FBP algorithm to obtain  $\mathbf{X}_n$ .

For actual training, we use the stochastic gradient descent algorithm with a batch size of one for minimizing (3.3). We train for 60 epochs which means we make 60 sweeps through the whole training dataset. We take  $\eta = 10^{-3}$  for the learning rate, include a momentum parameter  $\beta = 0.99$ , and use the mean absolute error for the distance measure in (3.3). The weights in the  $j$ th layer are initialized by sampling the uniform distribution on  $[-H_\ell, H_\ell]$  where  $H_\ell := \sqrt{6}/\sqrt{D_\ell + D_{\ell+1}}$  and  $D_\ell$  is the size of the input in layer  $\ell$ . This initializer is due to Glorot [22]. We use  $F = 32$  channels for the first convolution and the total number of layers is  $L = 19$ .

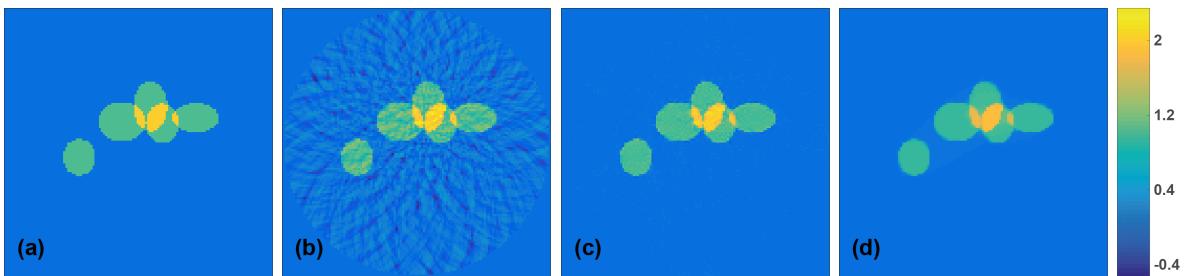


Figure 5: RESULTS FOR SIMULATED DATA (ALL IMAGES ARE DISPLAYED USING THE SAME COLORMAP). (a) Superposition of 5 ellipses as test phantom; (b) FBP reconstruction; (c) Reconstruction using the proposed CNN; (d) TV reconstruction.

## 4.2 Numerical results

We first test the network trained above on a test set of 50 pairs  $(\mathbf{X}, \mathbf{Y})$  that are generated according to the random model for the training data described above. For such random ellipse phantoms, the trained network is in all tested case able to almost completely eliminate the sparse data artifacts in the test images. Figure 5 illustrates such results for one of the test phantoms. Figure 5(a) shows the phantom, Figure 5(b) the result of the FBP algorithm which contains severe undersampling artifacts and Figure 5(c) the result of applying the CNN (right) which is almost artifact free. The actual relative  $\ell^2$ -reconstruction error  $\|\mathbf{Y}_{\text{CNN}} - \mathbf{Y}\|_2/\|\mathbf{Y}\|_2$  of the CNN reconstruction is 0.0087 which is much smaller than the relative error of FBP reconstruction which is 0.1811.

We also compared our trained network to penalized TV minimization [1, 64]

$$\frac{1}{2} \|\mathbf{p} - \mathcal{P}(\mathbf{Y})\|_2^2 + \lambda \|\mathbf{Y}\|_{\text{TV}} \rightarrow \min_{\mathbf{Y}} . \quad (4.1)$$

Here  $\mathcal{P}$  is a discretization of the PAT forward operator using  $M$  detector locations and  $d$  spatial discretization points and  $\|\cdot\|_{\text{TV}}$  is the discrete total variation. For the presented results, we take  $\lambda = 0.002$  and used the lagged diffusivity algorithm [70] with 20 outer and 20 inner iterations for numerically minimizing (4.1). TV minimization exploits the sparsity of the gradient as prior information and therefore is especially well suited for reconstructing sums of indicator functions and can be seen as state of the art approach for reconstructing such type of objects. As can be seen from the results in Figure 5(d), TV minimization in fact gives very accurate results. Nevertheless, the deep learning approach yields comparable results in both cases. In terms of the relative  $\ell^2$ -reconstruction error, the CNN reconstruction even outperforms the TV reconstruction (compare with Table 1).

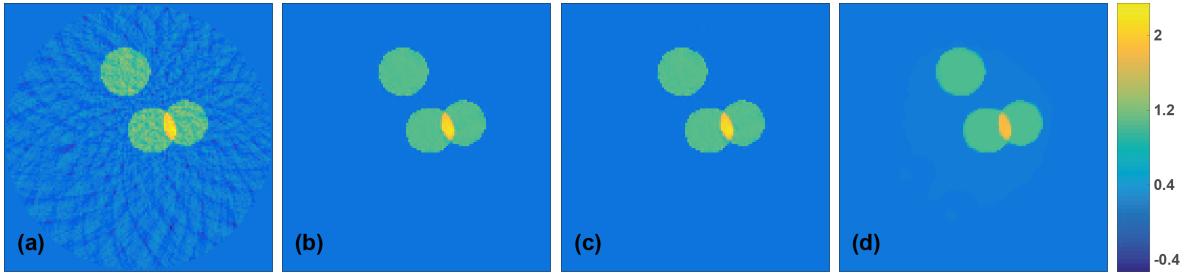


Figure 6: RESULTS FOR NOISY TEST DATA WITH 2% GAUSSIAN NOISE ADDED (ALL IMAGES ARE DISPLAYED USING THE SAME COLORMAP). (a) Reconstruction using the FBP algorithm; (b) Reconstruction using the CNN trained without noise; (c) Reconstruction using the CNN trained on noisy images; (d) TV reconstruction.

In order to test the stability with respect to noise we also test the network on reconstructions coming from noisy data. For that purpose, we added Gaussian noise with a standard deviation equal to 2% of the maximal value to simulated pressure data. Reconstruction results are shown in Figure 6. There we show reconstruction results with two differently trained networks. For the results shown in Figure 6(b) the CNN has been trained on the exact data, and for the results shown in Figure 6(c) it has been trained on noisy data. The reconstructions using each of the networks are again almost artifact free. The reconstruction from the same data with TV minimization is shown in Figure 6(d). The relative  $\ell^2$ -reconstruction errors for all reconstructions are given in Table 1.

### 4.3 Results for Shepp-Logan type phantom

In order to investigate the limitations of the proposed deep learning approach, we additionally applied the CNN (trained on the ellipse phantom class) to test phantoms where the training data are not appropriate (Shepp-Logan type phantoms). Reconstruction results for such a Shepp-Logan type phantom from exact data are shown in Figure 7, which compares results using FBP (Figure 7(a)), TV minimization (Figure 7(b)) and CNN improved versions using the ellipse phantom class without noise (Figure 7(c)) and with noise (Figure 7(d)) as training data. Figure 8 shows similar results for noisy measurement data with added Gaussian noise with a standard deviation equal to 2% of the maximal pressure value. As expected, this time the network does not completely remove all artifacts. How-

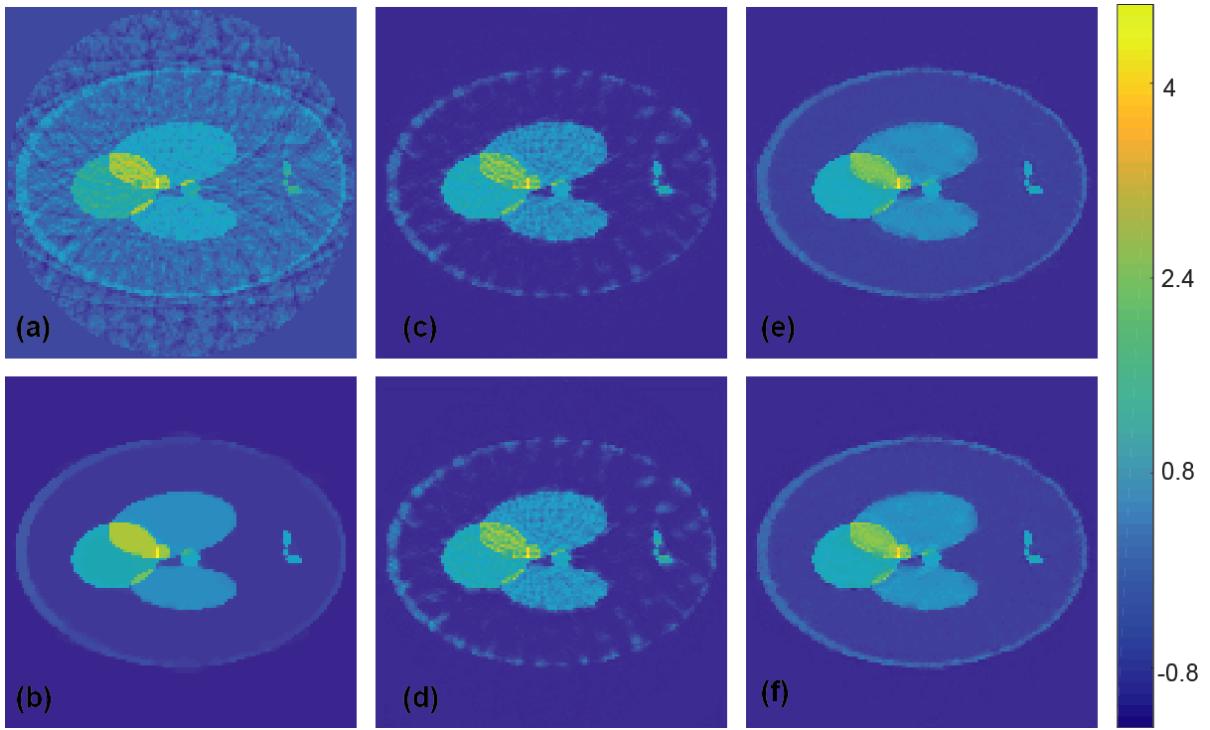


Figure 7: RECONSTRUCTION RESULTS FOR A SHEPP-LOGAN TYPE PHANTOM FROM DATA WITH 2% GAUSSIAN NOISE ADDED (ALL IMAGES ARE DISPLAYED USING THE SAME COLORMAP). (a) FBP reconstruction; (b) Reconstruction using TV minimization. (c) Proposed CNN using wrong training data without noise added; (d) Proposed CNN using wrong training data with noise added; (e) Proposed CNN using appropriate training data without noise added; (f) Proposed CNN using appropriate training data with noise added

ever, despite the Shepp-Logan type test object has features not appearing in the training data, still many artifacts are removed by the network trained on the ellipse phantom class.

We point out, that the less good performance of CNN in Figure 7(a)-(d) and 8(a)-(d) is due to the non-appropriate training data and not due to the type of phantoms or the CNN approach itself. To support this claim, we trained additional CNNs on the union of 1000 randomly generated ellipse phantoms and 1000 randomly generated Shepp-Logan type phantoms. The Shepp-Logan type phantoms have position, angle, shape and intensity of every ellipse chosen uniformly at random under the side constraints that the support of every ellipse lies inside the unit disc. The results of the CNN trained on the new training data and are shown in Figures 7 (e), (f) for exact measurement data and in Figure 8 (e), (f) for noisy measurement data. For both results we applied a CNN trained using training data without (e) and with noise (f). And indeed, when using appropriate training data including Shepp-Logan typ phantoms, the CNN is again comparable to TV minimization. We see these results quite encouraging; future work will be done to extensively test the framework using a variety of training and test data sets, including real world data.

The relative  $\ell^2$ -reconstruction errors for all presented numerical results are summarized in Table 1.

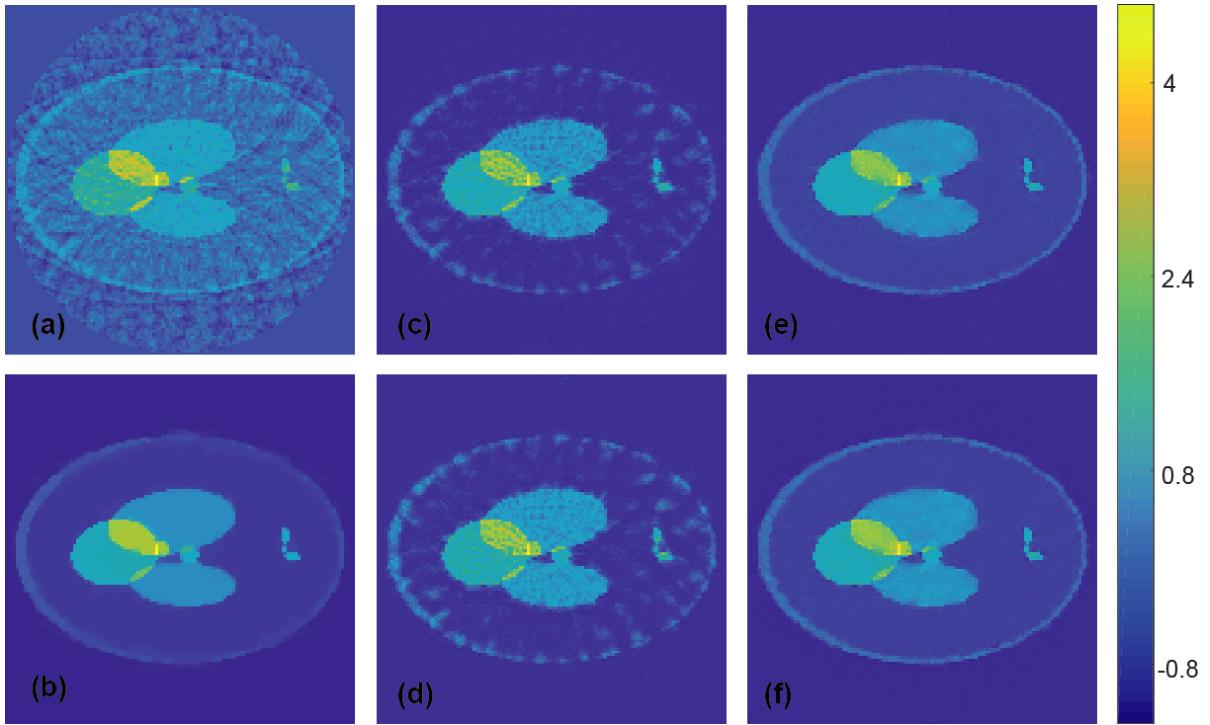


Figure 8: RECONSTRUCTION RESULTS FOR A SHEPP-LOGAN TYPE PHANTOM USING SIMULATED DATA (ALL IMAGES ARE DISPLAYED USING THE SAME COLORMAP). (a) FBP reconstruction; (b) Reconstruction using TV minimization. (c) Proposed CNN using wrong training data without noise added; (d) Proposed CNN using wrong training data with noise added; (e) Proposed CNN using appropriate training data without noise added; (f) Proposed CNN using appropriate training data with noise added.

phantom	FBP	TV	ELL	ELLn	SL	SLn
5 ellipses (exact)	0.1811	0.0144	0.0087	-	-	-
3 ellipses (noisy)	0.1952	0.0110	0.0051	0.0038	-	-
Shepp-Logan (exact)	0.3986	0.0139	0.1017	0.1013	0.0168	0.0186
Shepp-Logan (noisy)	0.3889	0.0154	0.1054	0.1027	0.0198	0.0206

Table 1: **Relative  $\ell^2$ -reconstruction errors for the four different test cases.** Compared are FBP, TV, and the proposed CNN reconstruction trained on the class of ellipse phantoms without noise (ELL) and with noise (ELLn), as well as trained on a class containing Shepp-Logan type phantom without noise (SL) and with noise (SLn).

#### 4.4 Discussion

The above results demonstrate that deep learning based methods are a promising tool to improve PAT image reconstruction. The presented results show that appropriately trained CNNs can significantly reduce under sampling artifacts and increase reconstruction quality. To further support this claim, in Table 2 we show the averaged relative  $\ell^2$  reconstruction

error for 100 Shepp-Logan type phantoms (similar to the ones in Figure 7. We see that even in the case where we train the network for the different class of ellipse shape phantoms, the error decreases significantly compared to FBP.

phantom	FBP	TV	ELL	ELLn	SL	SLn
Exact	0.2188	0.0140	0.1218	0.1078	0.0215	0.0213
Noisy	0.2374	0.0150	0.1290	0.1141	0.0256	0.0243

Table 2: **Average error analysis.** Relative  $\ell^2$ -reconstruction error averaged of over 100 Shepp-Logan type phantoms. Compared are FBP, TV, proposed CNN reconstruction trained on a class of ellipse phantoms without noise (ELL) and with noise (ELLn), as well as trained on a class of containing Shepp-Logan type phantom without SL and with noise (SLn).

Any reconstruction method for solving an ill-posed underdetermined problem, either implicitly or explicitly requires a-priori knowledge about the objects to be reconstructed. In classical variational regularization, a-priori knowledge is incorporated by selecting an element which has minimal (or at least small value) of a regularization functional among all objects consistent with the data. In the case of TV, this means that phantoms with small total variation ( $\ell^1$  norm of gradient) are reconstructed. On the other hand, in deep learning based reconstruction methods a-priori knowledge is not explicitly prescribed in advance. Instead, the a-priori knowledge is encoded in the given training class and CNNs are trained to automatically learn the structure of desirable outputs. In the above results, the training class consists of piecewise constant phantoms which have small total variation. Consequently, TV regularization is expected to perform well. It is therefore not surprising, that in this case TV minimization outperforms CNN based methods. However, the CNN based methods are more flexible in the sense that by changing the training set they can be adjusted to very different type of phantoms. For example, one can train the CNN for classes of experimental PAT data, where it may be difficult to find an appropriate convex regularizer.

## 4.5 Computational efforts

Application of the trained CNN for image reconstruction is non-iterative and efficient. In fact, one application of the used CNN requires  $\mathcal{O}(F^2 L d^2)$  FLOPS, where  $F$  is the number of channels for the first convolution and  $L$  describes the depth of the network. Moreover, CNNs are easily accessible to parallelization. For high resolution images,  $F^2 L$  will be in the order of  $d$  and therefore the effort for one evaluation of the CNN is comparable to effort of one evaluation of the PAT forward operator and its adjoint, which both require  $\mathcal{O}(d^3)$  FLOPS. However, for computing the minimizer of (4.1) we have to repeatedly evaluate the PAT forward operator and its adjoint. In the examples presented above, for TV minimization we evaluated both operations 400 times, and therefore the deep learning image reconstruction approach is expected to be faster than TV minimization or related iterative image reconstruction approaches.

For training and evaluation of the U-net we use the Keras software (see <https://keras.io/>), which is a high-level application programming interface written in Python. Keras runs on top of TensorFlow (see <https://www.tensorflow.org/>), the open-source software library for machine intelligence. These software packages allow an efficient and simple implementation of the modified U-net according to Figure 4. The filtered backprojection and the TV-minimization have been implemented in MATLAB. We perform our computations using an Intel Core i7-6850K CPU and a Nvidia Geforce 1080 Ti GPU. The training time for the CNN using the training set of 1000 ellipse phantoms has been 16 seconds per epoch, yields 16 minutes for the overall training time (using 60 epochs). For the larger mixed training data set (consisting of 1000 ellipse phantoms and 1000 Shepp-Logan type phantoms) one epoch requires 25 seconds. Recovering a single image requires 15 milliseconds for the FBP algorithm and 5 milliseconds for applying the CNN. The reconstruction time for the TV-minimization (with 20 outer and 20 inner iterations) algorithm has been 25 seconds. In summary, the total reconstruction time using the two-stage deep learning approach is 20 milliseconds, which is over 1000 times faster than the time required for the TV minimization algorithm. Of course, the reconstruction times strongly depend on the implementation of TV-minimization algorithm and the implementation of the CNN approach. However, any step in the iterative TV-minimization has to be evaluated in a sequential manner, which is a conceptual limitation of iterative methods. Evaluation of the CNN, on the other hand, is non-iterative and inherently parallel, which allows efficient parallel GPU computations.

## 5 Conclusion

In this paper, we developed a deep learning approach for PAT from sparse data. In our approach, we first apply a linear reconstruction algorithm to the sparsely sampled data and subsequently apply a CNN with weights adjusted to a set of training data. Evaluation of the CNN is non-iterative and has a similar numerical effort as the standard FBP algorithm for PAT. The proposed deep learning image reconstruction approach has been shown to offer a reconstruction quality similar to state of the art iterative algorithms and at the same time requires a computational effort similar to direct algorithms such as FBP. The presented numerical results can be seen as a proof of principle, that deep learning is feasible and highly promising for image reconstruction in PAT.

As demonstrated in Section 4 the proposed deep learning framework already offers a reconstruction quality comparable to state of the art iterative algorithms for the sparse data problem in PAT. However, as illustrated by Figures 7 and Figures 8 this requires the PAT image to share similarities with the training data used to adjust the weights of the CNN. In future work, we will therefore investigate and test our approach under various real-world scenarios including realistic phantom classes for training and testing, different measurement geometries, and increased discretization sizes. In particular, we will also train and evaluate the CNNs on real world data.

Note that the results in the present paper assume an ideal impulse response of the acoustic measurement system. For example, this is appropriate for PAT using integrating optical line detectors, which have broad detection bandwidth; see [59]. In the case that piezoelectric sensors are used for acoustic detection, the limited bandwidth is an impor-

tant issue that must be taken into account in the PAT forward model and the PAT inverse problem [58]. In particular, in this case, the CNN must also be trained to learn a deconvolution process addressing the impulse response function. Such investigations, as well as the application to real data, are beyond the scope of this paper and have been addressed in our very recent work [65]. (Note that the presented paper has already been submitted much earlier, initially in April 14, 2017 and to IPSE in July 1, 2017.)

It is an interesting line of future research using other CNNs that may outperform the currently implemented U-net. We further work on the theoretical analysis of our proposal providing insight why it works that well, and how to steer the network design for further improving its performance and flexibility.

## References

- [1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Probl.*, 10(6):1217–1229, 1994.
- [2] M. Agranovsky and P. Kuchment. Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography with variable sound speed. *Inverse Probl.*, 23(5):2089–2102, 2007.
- [3] S. Arridge, P. Beard, M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, and E. Zhang. Accelerated high-resolution photoacoustic tomography via compressed sensing. *Phys. Med. Biol.*, 61(24):8908, 2016.
- [4] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby. On the adjoint operator in photoacoustic tomography. *Inverse Probl.*, 32(11):115012 (19pp), 2016.
- [5] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate, 2014. arXiv:1409.0473.
- [6] J. Bauer-Marschallinger, K. Felbermayer, K.-D. Bouchal, I. A. Veres, H. Grün, P. Burgholzer, and T. Berer. Photoacoustic projection imaging using a 64-channel fiber optic detector array. In *Proc. SPIE*, volume 9323, 2015.
- [7] P. Beard. Biomedical photoacoustic imaging. *Interface focus*, 1(4):602–631, 2011.
- [8] Z. Belhachmi, T. Glatz, and O. Scherzer. A direct method for photoacoustic tomography with inhomogeneous sound speed. *Inverse Probl.*, 32(4):045005, 2016.
- [9] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [10] P. Burgholzer, J. Bauer-Marschallinger, H. Grün, M. Haltmeier, and G. Paltauf. Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors. *Inverse Probl.*, 23(6):S65–S80, 2007.
- [11] P. Burgholzer, G. J. Matt, M. Haltmeier, and G. Paltauf. Exact and approximate imaging methods for photoacoustic tomography using an arbitrary detection surface. *Phys. Rev. E*, 75(4):046706, 2007.

- [12] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang. Low-dose CT via convolutional neural network. *Biomed. Opt. Express*, 8(2):679–694, 2017.
- [13] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation, 2016. arXiv:1606.06650.
- [14] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537, 2011.
- [15] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.
- [16] A. De Cezaro, M. Haltmeier, A. Leitão, and O. Scherzer. On steepest-descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations. *Appl. Math. Comp.*, 202(2):596–607, 2008.
- [17] F. Dreier, M. Haltmeier, and S. Pereverzyev, Jr. Operator learning approach for the limited view problem in photoacoustic tomography. arXiv:1705.02698, 2017.
- [18] D. Finch, M. Haltmeier, and Rakesh. Inversion of spherical means and the wave equation in even dimensions. *SIAM J. Appl. Math.*, 68(2):392–412, 2007.
- [19] D. Finch, S. K. Patch, and Rakesh. Determining a function from its mean values over a family of spheres. *SIAM J. Math. Anal.*, 35(5):1213–1240, 2004.
- [20] J. Frikel and M. Haltmeier. Efficient regularization with wavelet sparsity constraints in PAT. arXiv:1703.08240, 2017.
- [21] K. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Netw.*, 2(3):183–192, 1989.
- [22] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9, pages 249–256, 2010.
- [23] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
- [24] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *J. Theor. Biol.*, 29(3):471–481, 1970.
- [25] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparsity in inverse geophysical problems. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, pages 763–784. Springer Berlin Heidelberg, 2010.
- [26] S. Gratt, R. Nuster, G. Wurzinger, M. Bugl, and G. Paltauf. 64-line-sensor array: fast imaging system for photoacoustic tomography. *Proc. SPIE*, 8943:894365, 2014.

- [27] H. Greenspan, B. van Ginneken, and R. M. Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging*, 35(5):1153–1159, 2016.
- [28] H. Grün, T. Berer, P. Burgholzer, R. Nuster, and G. Paltauf. Three-dimensional photoacoustic imaging using fiber-based line detectors. *J. Biomed. Optics*, 15(2):021306–021306–8, 2010.
- [29] Z. Guo, C. Li, L. Song, and L. V. Wang. Compressed sensing in photoacoustic tomography in vivo. *J. Biomed. Opt.*, 15(2):021311–021311, 2010.
- [30] M. Haltmeier. A mollification approach for inverting the spherical mean Radon transform. *SIAM J. Appl. Math.*, 71(5):1637–1652, 2011.
- [31] M. Haltmeier. Inversion of circular means and the wave equation on convex planar domains. *Comput. Math. Appl.*, 65(7):1025–1036, 2013.
- [32] M. Haltmeier. Universal inversion formulas for recovering a function from spherical means. *SIAM J. Math. Anal.*, 46(1):214–232, 2014.
- [33] M. Haltmeier. Sampling conditions for the circular radon transform. *IEEE Trans. Image Process.*, 25(6):2910–2919, 2016.
- [34] M. Haltmeier and L. V. Nguyen. Analysis of iterative methods in photoacoustic tomography with variable sound speed. *SIAM J. Imaging Sci.*, 10(2):751–781, 2017.
- [35] M. Haltmeier, O. Scherzer, and G. Zangerl. A reconstruction algorithm for photoacoustic imaging based on the nonuniform FFT. *IEEE Trans. Med. Imag.*, 28(11):1727–1735, November 2009.
- [36] Mark. us Haltmeier, M. Sandbichler, T. Berer, J. Bauer-Marschallinger, P. Burgholzer, and L. Nguyen. A new sparsification and reconstruction strategy for compressed sensing photoacoustic tomography. *arXiv:1801.00117, to appear in J. Acoust. Soc. Am.*, 2017.
- [37] Y. Han, J. J. Yoo, and J. C. Ye. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis, 2016. <http://arxiv.org/abs/1611.06391>.
- [38] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Netw.*, 4(2):251–257, March 1991.
- [39] Y. Hristova, P. Kuchment, and L. Nguyen. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Probl.*, 24(5):055006 (25pp), 2008.
- [40] C. Huang, K. Wang, L. Nie, and M. A. Wang, L. V. and Anastasio. Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media. *IEEE Trans. Med. Imag.*, 32(6):1097–1110, 2013.

- [41] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015. arXiv:1502.03167.
- [42] M. Jaeger, S. Schüpbach, A. Gertsch, M. Kitz, and M. Frenz. Fourier reconstruction in optoacoustic imaging using truncated regularized inverse k-space interpolation. *Inverse Probl.*, 23:S51–S63, 2007.
- [43] A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. *IEEE Trans. Med. Imag.*, 36(3):696–706, 2017.
- [44] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. arXiv:1611.03679, 2016.
- [45] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [46] R.A Kruger, P. Lui, Y.R. Fang, and R.C. Appledorn. Photoacoustic ultrasound (paus) – reconstruction tomography. *Med. Phys.*, 22(10):1605–1609, 1995.
- [47] P. Kuchment and L. Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In *Handbook of Mathematical Methods in Imaging*, pages 817–865. Springer, 2011.
- [48] L. A. Kunyansky. Explicit inversion formulae for the spherical mean Radon transform. *Inverse Probl.*, 23(1):373–383, 2007.
- [49] L. A. Kunyansky. A series solution and a fast algorithm for the inversion of the spherical mean Radon transform. *Inverse Probl.*, 23(6):S11–S20, 2007.
- [50] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [51] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, 1989.
- [52] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis, 2017.
- [53] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.
- [54] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik. Deep neural nets as a method for quantitative structure–activity relationships. *Journal of chemical information and modeling*, 55(2):263–274, 2015.

- [55] J. Meng, L. V. Wang, D. Liang, and L. Song. In vivo optical-resolution photoacoustic computed tomography with compressed sensing. *Optics letters*, 37(22):4573–4575, 2012.
- [56] F. Natterer. *The Mathematics of Computerized Tomography*, volume 32 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 2001.
- [57] R. Nuster, M. Holotta, C. Kremser, H. Grossauer, P. Burgholzer, and G. Paltauf. Photoacoustic microtomography using optical interferometric detection. *J. Biomed. Optics*, 15(2):021307–021307–6, 2010.
- [58] G. Paltauf, P. Hartmair, G. Kovachev, and R. Nuster. Piezoelectric line detector array for photoacoustic tomography. *Photoacoustics*, 8:28–36, 2017.
- [59] G. Paltauf, R. Nuster, M. Haltmeier, and P. Burgholzer. Photoacoustic tomography using a Mach-Zehnder interferometer as an acoustic line detector. *Appl. Opt.*, 46(16):3352–3358, 2007.
- [60] J. Provost and F. Lesage. The application of compressed sensing for photo-acoustic tomography. *IEEE Trans. Med. Imag.*, 28(4):585–594, 2009.
- [61] O. Ronneberge, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, 2015.
- [62] A. Rosenthal, V. Ntziachristos, and D. Razansky. Acoustic inversion in optoacoustic tomography: A review. *Current medical imaging reviews*, 9(4):318–336, 2013.
- [63] M. Sandbichler, F. Krahmer, T. Berer, P. Burgholzer, and M. Haltmeier. A novel compressed sensing scheme for photoacoustic tomography. *SIAM J. Appl. Math.*, 75(6):2475–2494, 2015.
- [64] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [65] J. Schwab, S. Antholzer, R. Nuster, and M. Haltmeier. DALNet: High-resolution photoacoustic projection imaging using deep learning. *arXiv:1801.06693*, 2018.
- [66] J. Schwab, S. Pereverzyev Jr, and M. Haltmeier. A galerkin least squares approach for photoacoustic tomography, 2016. *arXiv:1612.08094*.
- [67] P. Stefanov and G. Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Probl.*, 25(7):075011, 16, 2009.
- [68] P. Stefanov and G. Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011, 16, 2009.
- [69] B. E. Treeby and B. T. Cox. k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave-fields. *J. Biomed. Opt.*, 15:021314, 2010.

- [70] C.R. Vogel and M.E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.
- [71] G. Wang. A perspective on deep imaging. *IEEE Access*, 4:8914–8924, 2016.
- [72] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio. Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography. *Phys. Med. Biol.*, 57(17):5399, 2012.
- [73] L. V. Wang. Multiscale photoacoustic microscopy and computed tomography. *Nat. Photon*, 3(9):503–509, 2009.
- [74] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517, 2016.
- [75] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun. Deep image: Scaling up image recognition, 2015. arXiv:1501.02876.
- [76] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. Deep learning computed tomography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 432–440. Springer, 2016.
- [77] M. Xu and L. V. Wang. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E*, 71(1):016706, 2005.
- [78] M. Xu and L. V. Wang. Photoacoustic imaging in biomedicine. *Rev. Sci. Instruments*, 77(4):041101 (22pp), 2006.
- [79] Y. Xu, M. Xu, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography–II: Cylindrical geometry. *IEEE Trans. Med. Imag.*, 21:829–833, 2002.
- [80] H Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. arXiv:1607.08707, 2016.