Photoacoustic Image Reconstruction Using Multilayer Neural Network

Joep van de Weem
Technical University Eindhoven

***Abstract*—Photoacousic Tomography (PACT) is an emerging field in Biomedical Imaging used for e.g. breast cancer detection and brain lesion detection. Current conventional reconstruction techniques use Delay and Sum (DAS) based backprojection algorithms to reconstruct images, however these suffer from PACT's practical limitations. The main limitations are: Limited bandwidth, limited view, lossy medium and hetergeneous medium. This paper proposes an end-to-end deep learning approach for the reconstruction of PACT images constisting of two seperate neural networks. The first network suppresses sidelobes caused by beamforming using a beamforming network (BFN) tackling the limited view problem. It also extends the bandwidth of the bandlimited data using a bandwidth extension network (BWN). These are trained using and analysed using four training strategies to evaluate the performamnce. First proposed BFNs are trained using capon beamformed *in vivo* data and evaluated. Secondly the BWN is tested using DAS images based on simulated data and as training target the original intensity map. The third is training the entire network on only simulated data and lastly the network is trained on both simulated and enhanced *in vivo* data. From the results of these networks it can be concluded that the BWN network proves powerful in finding the underlying intensity map. The combined use with the beamforming network shows good results in enhancing the contrast near the center of the image. In the results it can be observed that the main limiting factor right now is the beamforming network as it does not enhance the negative part of the signal. It does, however, show great results in enhancing the resolution of the image.**

I. Introduction

Photoacoustic Computed Tomography (PACT) uses the photoacoustic effect to ultrasonically image the optical contrast. It achieves this by using non-ionizing laser pulses to heat up certain parts, leading to a transient thermoelastic that can be received using ultrasound transceivers. It combines the functional optical contrast of diffuse optical Tomography and the high spatial resolution of ultrasonography. It is shown to be a helpful tool in brain lesion detection [1], hemodynamics [2] and breast cancer diagnosis [3].

Conventional PACT usually uses the Delay and Sum (DAS) based backprojection algorithms [4] to reconstruct images. Adaptive approaches, such as coherence factor weighing [] and Capon beamforming [] show improvements in contrast, but are more computationally expensive. Other approaches use Deep Learning algorithms to find the underlying intensity map. This paper tries to combine these two methods into one by using Deep Learning to make an Beamforming Network (BFN) for contrast enhancements and a Bandwidth Expansion Network (BWN) to find the underlying intensity map from the beamformed bandlimited data.

As is stated in [5] the main practical limitations of PACT are: limited view, limited bandwidth, lossy medium and heterogeneous medium. The limited view problem causes an enhancement of the sidelobes and distortion of the images morphology. The limited bandwidth problem is caused by the bandlimited ultrasound sensors and results in blurring and an incomplete image of the true intensity in the image. The lossy medium problem causes a lower signal intensity in deeper regions and the heterogeneous medium causes blurring or a distortion of the true morphology. The proposed network tries to tackle the limited view and bandwidth problem.

In conventional DAS beamforming the output is a uniform weighted sum of the delayed received signals. In this case the delay is based on the distance between the source and the individual sensors and is usually estimated by using a fixed speed of sound value. Existing computationally expensive methods that use a better estimate of the speed of sound map are available [6] for more accurate delaying, however these often require known speed of sound maps and computationally expensive algorithms. After aligning the received channel signals, these are weighted following a predetermined apodization scheme and summed to yield the final image. The downside of such a method is the inherent design choice between resolution and contrast. Adaptive beamforming algorithms such as the Capon beamformer overcome these issues through data-adaptive weighing of the components at each point in the image.

Current Deep Learning techniques used in the reconstruction of PACT data focus on artifact removal [7], point source reconstruction [8] and reconstruction of ill-posed problems [9]. Alternatively, end-to-end solutions are proposed in [10] where they propose Y-NET which uses both the sensor data together with the DAS image to find the underlying intensity map leading to great results.

These solutions mainly focus on the limited bandwidth problem to find the underlying intensity map from the backprojected image. However these methods (with the exception of Y-NET) use the regular DAS backprojected images as the input. Recent developments in Deep Learning for ultrasonography [11] show great results using learned adaptive beamforming methods [12] to increase the contrast of the reconstructed image.

In this paper an end-to-end deep learning solution is proposed that both optimizes the beamformer apodization data-adaptively and extends the limited bandwidth of the ultrasonic receivers. The apodization is determined using a network that

is based on recent developments in Ultrasonography and is used to minimize the sidelobes and improve the contrast and resolution effectively tackling the limited view problem. The limited bandwidth problem is tackled using a Neural Network that is used to find the underlying intensity map by expanding the bandwidth of the band limited sensors.

In this paper ||.|| will describe the L2 or euclidean norm. Bold letters (e.g. **x**) will indicate vectors, x^H indicates complex transpose or Hermetian of x and the tr(.) denotes the mathematical trace operation.

In Section II, we first describe the existing beamforming methods. Next the preparation of the training data explained in section III, afterwards the deep neural networks structure is described in section IV and tested in section V followed by a conclusion VI and discussion in section VII.

II. Existing methods

A. Delay and Sum

Beamforming is a technique that is used to acquire signals from a specific direction using multiple sensors. In the case of PACT these multiple sensors are ultrasonic transducers and the signal that is acquired is the transient thermoelastic expansion caused by the laser. To reconstruct an image we basically want to know the value at each pixel according to each sensor in the array. A common and simple way to achieve this is by using Delay And Sum (DAS) beamforming.

In DAS beamforming the data is first delayed by applying by applying a delay that can be calculated using:

$$\Delta\mathbf{T}[x,y] = \frac{||\mathbf{r}_s - [x,y]^T||}{c} + t_0 \tag{1}$$

Where $\Delta\mathbf{T}_{s,g}$ propagation time from each sensor r_s to a grid point at $[x,y]^T$, c is the speed of sound in the medium and t_0 is a fixed delay that could be part of the system or as compensation for the time it takes for the particle to start moving given an impulse.

After this pre-delay, the value of all the sensors at a given pixel is multiplied with a certain weight and then added.

$$I[x,y] = \mathbf{w}[x,y]^h\mathbf{y}[x,y] \tag{2}$$

where I is the value at pixel [x, y], **x** is the delayed value from each sensor ($\mathbf{x}[x,y] = \mathbf{s}[\Delta T[x,y]]$) **w** is the weight at location [x, y]. The weight is in the simplest cases equal to a rectangular window or a more refined shape like a Hanning window. Adaptive apodization methods that use the statistics of an image are often used to improve this.

B. Capon

The Capon beamformer, first used in [13], is a common data-adaptive method used to improve the resolution in the beamformed reconstruction of images. It achieves this boost in resolution by minimizing the variance between the signal recorded at each of the sensors using weighing. To calculate the weights, first the covariance matrix must be estimated:

$$\hat{\boldsymbol{R}}[x,y] = \boldsymbol{y}[x,y]\boldsymbol{y}[x,y]^H \tag{3}$$

In which $\boldsymbol{y}$ is the pre-delayed signal received at location [x, y] and $\hat{\boldsymbol{R}}$ is the covariance matrix. This estimation of the covariance matrix can be non singular and thus the inverse might not exist. To ensure that the inverse exists, diagonal loading [14] is often used. This diagonal loading can be calculated using:

$$\epsilon = \frac{d * tr(\hat{\boldsymbol{R}})}{N} \tag{4}$$

in which d is an estimated diagonal loading factor. These factors can be used to calculate the final weights used for beamforming:

$$\boldsymbol{w} = \frac{(\hat{\boldsymbol{R}} - \epsilon\boldsymbol{I})^{-1}\boldsymbol{a}}{\boldsymbol{a}^H(\hat{\boldsymbol{R}} - \epsilon\boldsymbol{I})^{-1}\boldsymbol{a}} \tag{5}$$

These weights can be used to calculate the final intensity using (2).

C. Multilayer Capon

A multilayer approach to the Capon beamformer is proposed in [15]. It shows good results for circular arrays in Hydrophonics. This approach has perfomance that is identical to the conventional MVDR beamformer while being more robust and and having a lower white noise gain. adsfasf Since half time reconstruction is used in our approach. It would be impossible to use the multilayer beamformer directly, since two opposing sensors will never receive a signal from the same source in the final grid. This means that the subarray already needs to become bigger.

Therefore a variation of this beamformer is used. In this beamformer a window of M equidistantly spaced sensors is used instead of the two in [15]. This window is then traversed over the array to accurately estimate the covariance matrix and a partial signal vector using:

$$\boldsymbol{R}[x,y] = \frac{\sum_{i\in S}\boldsymbol{y_i}[x,y]\boldsymbol{y_i}[x,y]^H}{M} \tag{6}$$

$$\boldsymbol{y_{part}} = \frac{\sum_{i\in S}\boldsymbol{y_i}[x,y]}{M} \tag{7}$$

where S indicates the subset of sensors indexed by every $\frac{N}{M}$'th sensors. $\boldsymbol{y_i}$ is the signal from the *i*'th sensor. As is already stated in [15], increasing the number of sensors in a subarray, decreases the robustness of the solution. To improve this, diagonal loading is applied to the covariance matrix to ensure its singularity, similar as in 4. These steps are then repeated according to:

1) Estimate covariance matrix and partial $\boldsymbol{y_{part}}$ using 6 and 7 respectively
2) Calculate diagonal loading using 4
3) Calculate weight using 5
4) Calculate response of partial array using

$$\boldsymbol{y_{resp}} = \boldsymbol{w}^h\boldsymbol{y_{part}}$$

5) Repeat step 1, 2 and 3 using the response vector $\boldsymbol{y_{resp}}$ until only one value remains.

The multilayer Capon approach results in a contrast enhanced image as can be seen in Figure 3. This image shows that this

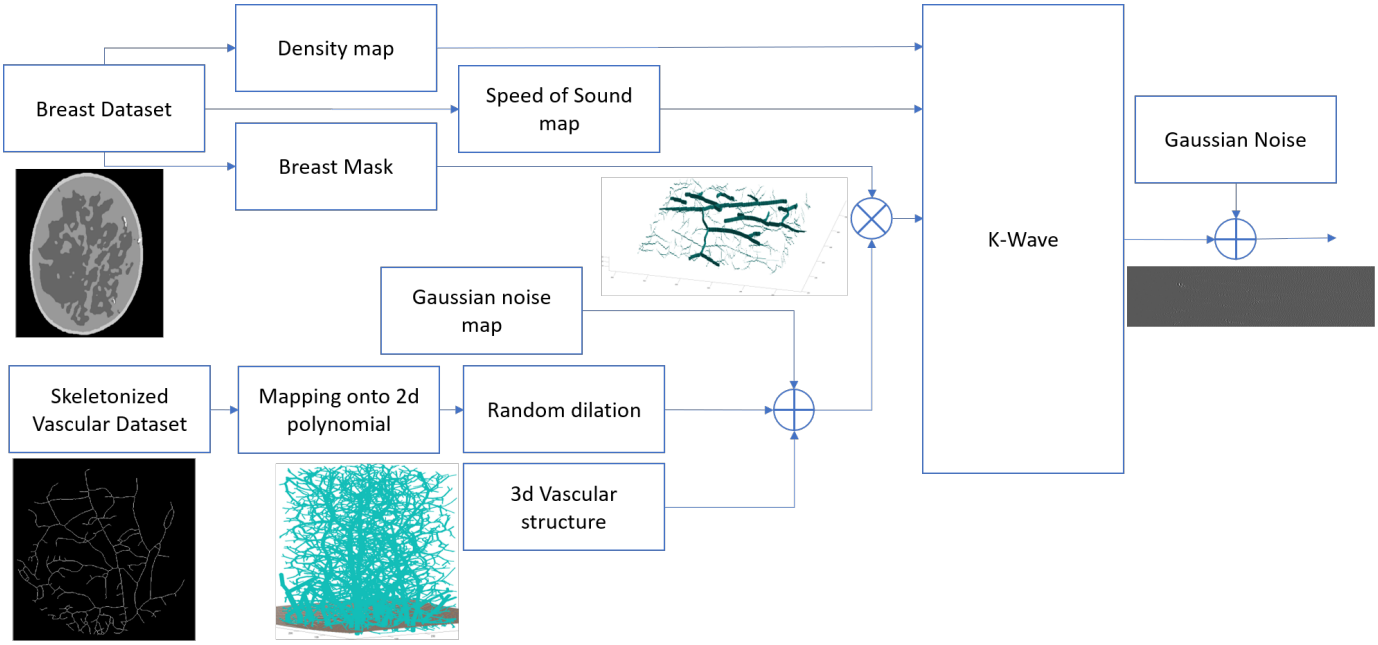


Fig. 1. A schematic overview of the steps taken to simulate realistic vascular photoacoustic data

algorithm is especially powerful when searching for data in the regions further from the center.

D. Speed Of Sound correction

Neither of the aforementioned beamforming methods can be used to compensate for heterogeneities in the speed of sound. Delaying methods using more complex delay estimation using e.g. Fast Marching Methods or other Eikonal solvers are shown to reduce artifacts in [6], [16] and [17]. These methods estimate the paths that the sound traverses to get to each of the sensors using a prior known speed of sound map. However these methods are often too computationally expensive to implement and rely on prior known speed of sound maps to estimate the ray paths.

The speed of sound heterogeneities also cause errors in phase between data from each sensor. This discrepancy in phase is detrimental to the intensity of the summed signal and therefore bad for overall contrast. This can be compensated for by demodulating the network in an I and Q component and weighting these components differently to cause a phase shift that can minimize the error.

Another problem with heterogeneities in speed of sound and differences in depth can cause large displacements for sources seen from two opposing sensors in circular arrays. A common technique to get rid of a part these displacements is by using half-time reconstruction [18], in which only the data up until the middle of the image is used. This means that sensors that are far away from the pixel of interest, will no longer be used to calculate the pixel value. The downside of this method is that part of the data is discarded.

III. Data Preparation

Machine learning solutions are well known for requiring labelled data to achieve good generalization. The problem here is that there is no labelled data available. In this paper we try to tackle this problem by using simulated data generated using the K-wave toolbox [19] together with enhanced *in vivo* data to make sure that the network does not overfit on the simulated data.

A. Data Generation

In this particular case PACT is used for angiography in the breast using a circular ultrasonic receiver array similar to the one described in [3]. The simulated data needs to be as close to the *in vivo* images received by the physical system as possible. Therefore some factors are taken into account, namely: tissue Speed of sound heterogeneity, tissue density heterogeneity, vascular structure, shape of array, sensor bandwidth, physical response of the system, intensity map of the system, depth. The steps taken to achieve this are described in Figure 1.

The speed of sound heterogeneity and density heterogeneity are mapped using the numerical phantoms of the breast generated in [20]. The heterogeneity in speed of sound will generate circular artifacts similar to those found in the *in vivo* data. These artifacts are further described in [21].

Since PACT has a high optical absorption coefficient for blood vessels, the intensity map used for simulation will mainly consist of the vascular structure. This vascular structure consists of a three-dimensional structure of vessels with varying sizes. To achieve this, two datasets are used; one for microvascular structures and one for the main vessel structures. For the microvascular structure a generated 3d model of a vascular structure found in [22] was used. This image was

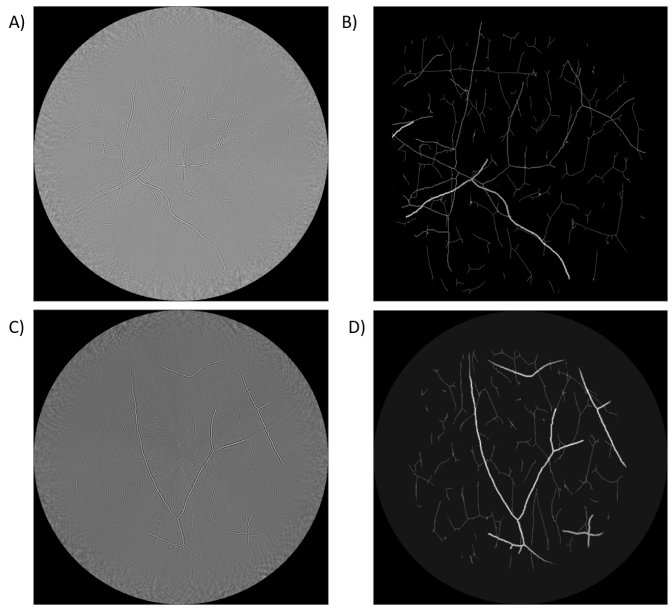


Fig. 2. Simulated data with (a, c): regular delay and summed data and (b, d): intensity map used to simulate.

skeletonized to make the vessels as thin as possible. For the main vessel structure a 2d retinal dataset [23] was used. This retinal dataset was first skeletonized, then mapped to a randomly generated two dimensional polynomial to achieve a continuous vessel structure in 3-dimensional space. Lastly it was dilated in 3-dimensional space with a random amount to simulate the varying thicknesses found in the *in vivo* data.

The standard physical response of the system can be described by looking at how the physical system reacts to an impulse. Since the *in vivo* data was made on the same setup as used in [3], the impulse response given in the supplementary material was used for this system. The bandwidth of the system is also represented in this waveform, since it is received (and thus filtered,) by the actual system.

The intensity of the laser causes a Gaussian intensity over the breast and a exponentially decaying intensity over the depth of the breast. This is simulated by applying a Gaussian response over the x-y plane of the breast and an inverse exponential function in the z direction over the intensity map.

The noise of the system is generated using a randomly generated speckle map that is more intense on the middle of the breast together with an uncorrelated white nose map over the final result.

The intensity maps are inherently 3-dimensional and it is thus chosen to apply an exponential decay in both directions from the sensor location in the z-direction to add more uncertainty to vessels at a different depth. Afterwards it is summed over the z-direction to get a two dimensional intensity map that can be used for training targts.

Examples of the resulting images can be seen in figure 2.

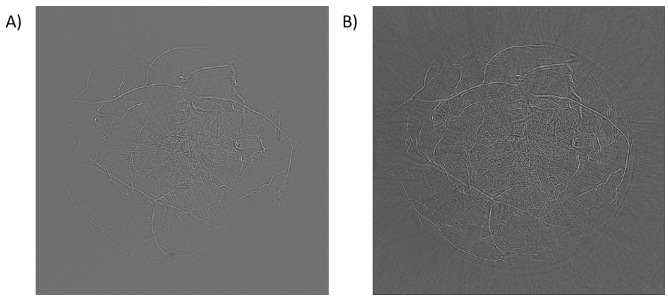


Fig. 3. *in vivo* where A) is regular delay and summed data and B) is the beamformed data using the Capon beamformer. Both of the signals use spatially interpolated sensor data for reconstruction. The contrast of the images is enhanced for the purpose of visualization

B. in vivo pre-processing

The *in vivo* data available consisted of a stack of 255 images from one breast and 1 from another. From the image stack 10 sufficiently different images are chosen for training. The simulated data consisted of 156 images. The training was done using an increased probability for *in vivo* images of 5:1 (*in vivo*:simulated). this meant that 45:125 images were used per epoch for training and 2:31 for validation. The two *in vivo* images consist of one for each breast.

For data preparation the data is first interpolated [24] to remove some of the limited view artifacts. Afterwards its pre-delayed to a grid of 256x1024 ($R \times \phi$) resulting in a 1024x256x1024 grid ($elements \times R \times \phi$. Memory is a big bottleneck in the system, therefore both the input and output data is divided into patches to save memory usage. This is possible due to the The spatial invariance of convolutional networks. These patches have a base size of 32 x 128. However, there needs to be an added compensation since the receptive field of the network is larger. This compensation is applied to the input by adding surrounding data to the patch making it (32 + F_R) x (128 + F_ϕ) where F_x indicates the receptive field of the network in the "x" direction. The patches are batched together in pairs of 8.

The setup used for receiving the *in vivo* processing mainly comprises of an illumination laser, an ultrasonic transducer array, signal amplification/acquisition modules, a linear scanning stage and a patient bed. For more information on this system it is suggested to look at the methods section of [3].

Since the simulated data is still an approximation of reality, the *in vivo* data is also used in the training of the network to improve the overall generalization. However, there is no underlying intensity map available. To get these ground truth images. The Capon beamformer in II is used to increase the contrast in the images as well as the interpolation of the data to reduce the aliasing artifacts [24]. as it requires only half of the sensors to be used at each of the pixels. These images are then used to train the beamforming part of the network.

IV. Neural Network

The proposed Neural Network structure consists of two parts, a deep beamformer and a bandwidth expander. The

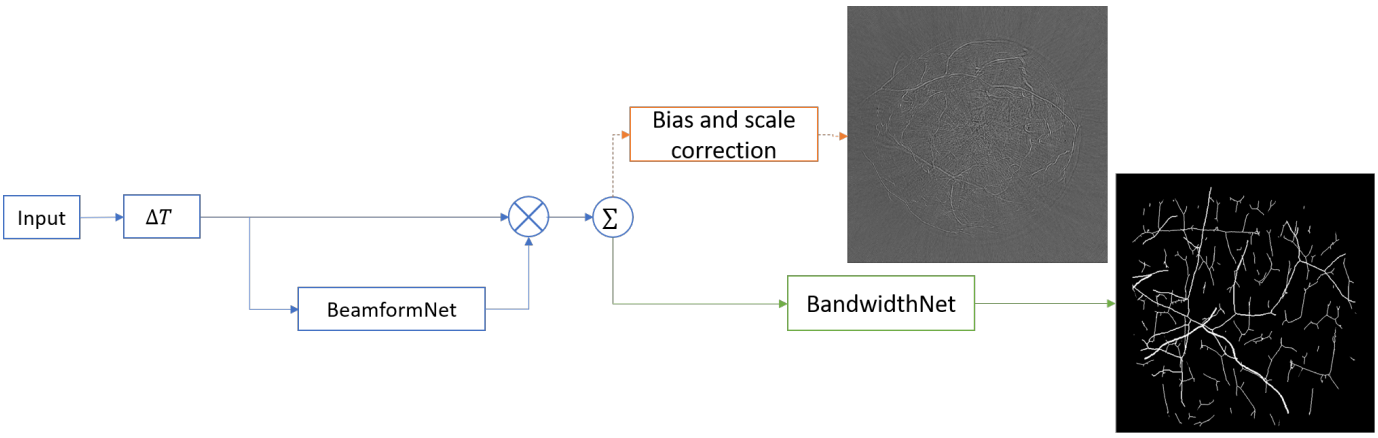


Fig. 4. Proposed Network Structure consisting of a Beamforming network (BFN) and a Bandwidth Extension Network (BWN). The former of which finds the weights based on the pre-delayed input data and the second extends the bandlimited data to get the underlying intensity mask

former of which tries to tackle the limited view problem by decreasing the sidelobes and improvig the overall resolution by adaptively weighing the data (BFN), and the second expands the bandwidth of this image to tackle the limited bandwidth problem (BWN). The setup for these networks is described in Figure 4.

The networks use, similar to [12] Antirectifier layers [] in the network to keep the bipolar information in the network. The Antirectifier consists of a concatination of the two Rectified Linear Unit (ReLU) Activation functions. One regular ReLU and one ReLU of the negative of the input data.

A. BFN

The BFN finds the ideal weights for the delayed signal. The idea for this network structure is based on the recent advancements in Deep Learning for Ultrasound image reconstruction where it is showing promising results.

The idea is that the multiple convolutional layers will be able to find the signal while ignoring the noise based on the data of sensors around it and apply the weights accordingly.

1) 1D-Network: As is discussed in previous chapters, the Capon beamformer solely uses the relation between sensors to obtain the result. To also look at this relation between sensors, a one dimensional kernel along the sensor direction can be used. the kernel sizes are shown in Table I.

The used kernel sizes lead to a receptive field of 36 sensors. This receptive field is still small compared to the total interpolated sensor size of 512 sensors per pixel, however a bigger receptive field leads to more memory usage, which is a bottleneck in the system.

The advantage of this method is that it does not discard any data. The downside is that due to memory constraints it is difficult to obtain a large receptive field.

2) Pseudo-3D Network: This network expands on the idea of the 1D-Network by also using the other two spatial dimension to use as context for calculating the weights. The idea is that the network would be able to see if there is a signal found next to the pixel of interest and subsequently conclude that it is also more probable that there is a signal present in the pixel.

To save on the heavy memory requirement that full 3-dimensional kernels bring, a Pseudo-3d kernel [25] is used. This kernel could consists of a e.g. (3x3x1) kernel followed by a (1x1x5) kernel (with dimensions being $R \times \phi \times N$) leading

Layer	1DNet	AverageNet	PseudoNet	BWN
1		AvgPool(8, 1, 1)		Conv(5, 5) - 32
2				Antirectifier
3	Conv(9, 1, 1) - 16	Conv (11, 1, 1) - 16	Conv(9, 1, 1) - 16	Conv(5, 5) - 64
4	AntiRectifier	AntiRectifier	AntiRectifier	Antirectifier
5	Conv(9, 1, 1) - 16	Conv (11, 1, 1) - 32	Conv(1, 5, 5) - 16	Conv(5, 5) - 64
6	ReLU	ReLU	ReLU	ReLU
7	Conv(9, 1, 1) - 16	Conv (11, 1, 1) - 16	Conv(9, 1, 1) - 16	Conv(5, 5) - 64
8	ReLU	ReLU	ReLU	ReLU
9	Conv(9, 1, 1) - 1	Conv (11, 1, 1) - 16	Conv(9, 5, 5) - 1	Conv(5, 5) - 32
10		ReLU		ReLU
11		Conv (11, 1, 1) - 1		Conv(5, 5) - 1
12	Input * Layer9	Layer1 * Layer11	Input * Layer9	ReLU

TABLE I

TABLE CONTAINING THE STRUCTURE OF ALL THE NETWORKS, CONVOLUTIONAL KERNELS ARE DESCRIBED AS (ELEMENT $\times R \times \phi$). THE " - 16" MEANS THAT THAT LAYER HAS 16 FILTERS. ALL THE LAYERS ARE CIRCULARLY PADDED IN ELEMENT AND PHI DIRECTION AND ZERO PADDED IN THE R DIRECTION IF NEEDED.

to a receptive field of (3x3x5) while using only 14 weights instead of the 45 for a single (3x3x5) layer.

However, even after reducing the size with the pseudo-3d kernels, using multiple 3d convolutional layers on this already big network still lead to a memory bottleneck. To minimize this, a couple memory minimizing tricks are used, first the input images are split up into batches because of the spatial invariance of the convolutional network. Secondly mixed precision training [26] is used to maintain high accuracy while lowering the memory cost while training. And lastly the aforementioned half time reconstruction is used which effectively halves the amount of sensors at each pixel.

another problem with the usage of convolutional layers in the spatial domain together with a circular array is that there is more data in the center of an image generated using a Cartesian grid. Because the rays traced from the elements will start to overlap in the center. It is therefore decided to switch to a polar grid for reconstructions since the convolutional layers of a network are based on a constant resolution.

The downside of the Pseudo-3D network are that it is fairly computationally expensive and cannot easily be translated to, for instance, a denser grid like the 1DNet can. By also using the spatial information, the network will most likely only work on a grid that is equal to the grid it is trained on. The upside is that the usage of spatial information should lead to better overall performance.

3) Average Network: The Average Network (AverageNet) is loosely based on the proposed beamforming network, as it uses subarrays to calculate the final weights. These subarrays are similar to the $\boldsymbol{y_{part}}$ calculated in equation 7 which can also be achieved using an average pooling layer in the neural network.

By limiting the amount of sensors that the system needs to process, the memory requirements have become smaller, while the averaging caused the receptive field to become higher. This means that the depth and the kernel sizes of the network can be expanded. The depth of 5 layers combined with a 11 deep convolutional kernel is chosen because this leads to a receptive field in the sensor direction of 55 out of 64 averaged sensors.

The advantages of this network are its lower memory usage and its increased receptive field.

B. BWN

BWN mainly tries to tackle the limited bandwidth problem stated in [5]. The original beamformed images consist of data that is received by a bandlimited sensor. To get to the original intensity map of the vascular structure, the bandwidth of the image needs to be extended. This network uses a series of convolutional layers to find the original intensity map. The layers are described in I. The idea of this network is that it effectively deconvolves the inputted sensor data to the original data.

C. Training strategies

Several training strategies are considered for the training of the network. These training strategies are used to test the effect of certain design choices on the results. All the networks are trained for 100 epochs.

1) Only BFN: The strengths of the BFN is evaluated by training it on purely the contrast enhanced *in vivo* data. Evaluating the increase in contrast and resolution of the beamformed image gives insights into the strengths and weaknesses of each of the BFNs. The best BFN is chosen based on these results and used in the subsequent tests.

2) Only BWN: In this strategy, the BWN is only trained on the simulated data based on regular DAS images with as training targets the intensity maps used for simulation. The result of this network is compared to the results of the other strategies to see if the addition of the beamforming network actually improves the networks performance. This is tested at the hand of its Structural Similarity Index (SSIM)

$$SSIM(f, gt) = \frac{(2\mu_f\mu_{gt} + c_1)(2\sigma_{f,gt} + c_2)}{(\mu_f^2 + \mu_{gt}^2 + c_1)(\sigma_f^2 + \sigma_{gt}^2 - c_2)} \tag{8}$$

where μ_x is the expected value of x, $\sigma_{x,y}$ is the covariance of x and y and $\sigma_x^2$is the variance of x. And the Peak Signal to Noise Ratio (PSNR):

$$PSNR(f, gt) = 10\log_{10}(\frac{I_{max}^2}{MSE}) \tag{9}$$

where MSE indicates the mean square error and I_max indicates the maximum intensity. A higher PSNR hints at a higher quality image.

3) Complete network trained on simulated data: In this strategy, the complete network is trained on simulated data. This is done to evaluate whether the addition of the contrast enhanced *in vivo* data improves the generalization of the network.

4) Complete network trained on all data: This strategy uses all the data to train the entire network.

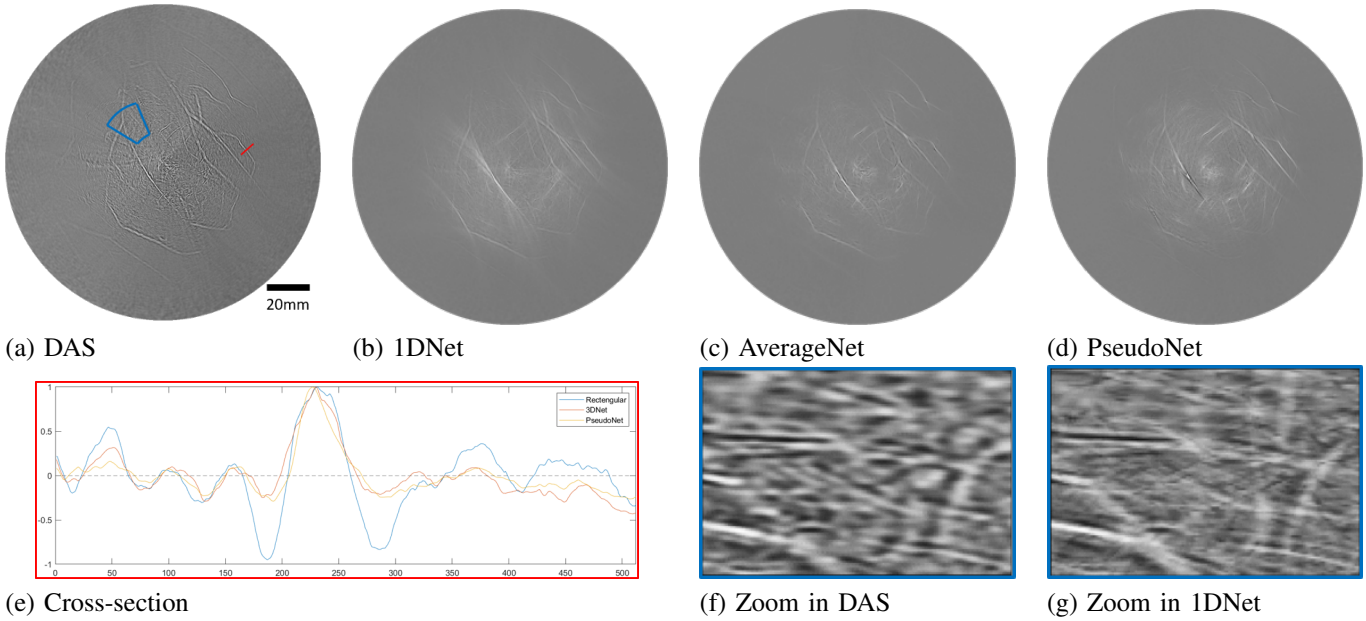
D. Speed of sound Correction

As was stated before, the phase discrepancy caused by heterogeneous speed of sound can be detrimental to the final mapping of the signal. Splitting the result into two components, an I and a Q component using a QI transform and weighing them differently to align the signal from all the sensors can be used to compensate for this discrepancy and improve the overall signal to noise ratio. This is an interesting part to look into for improving the result of the beamforming, however due to memory limitations, it was decided not to apply this in the final design.

V. Results

The results were achieved by implementing the neural networks using a PyTorch [27] backend on a computer with an Nvidia Titan X graphics card. For training, both networks use an Adam optimizer. The beamforming network has a learning rate of $\mu = 0.001$ and the bandwidth network has a learning rate of $\mu = 0.0001$. The learning rates differ due to stability issues during training.

The results are split into the four training strategies described in section IV-C.



(a) DAS (b) 1DNet (c) AverageNet (d) PseudoNet

(e) Cross-section (f) Zoom in DAS (g) Zoom in 1DNet

Fig. 5. Output of beamforming network trained using the *in vivo* data as input. In this image (a) is generated using regular DAS beamforming, (b) is generated using the 3D Net, (c) using the AverageNet and (d) using the PseudoNet. (e) shows the cross-section of the vein marked by red in (a) for the DAS, 1DNet and PseudoNet examples and (f) and (g) show a zoomed in version of the blue area in (a) using DAS and 1DNet respectively. The contrast of the images is enhanced in post-processing for the purpose of visualization.

A. *Only BFN*

Since it is hard to pin numbers to the performance of this network, a more qualitative approach is taken. The Networks are enhanced based on their visual clarity, the amount of clutter and their ability to uncover veins in regions both far and close to the center. The results can be seen in figure 5.

1) 1DNet: The results for 1DNet are promising. the network results can be seen in figure 5(c) and (e). On a perspective of visual clarity it suffers a bit compared to the DAS image, however the amount of visual clutter is also decreased significantly. Another proof of this reduction in visual clutter can be seen in figure 5(e) where it can be seen that the 1DNet has a significantly lower noise as compared to the the rectangular window together with slightly less wide main lobe. This image also shows the downside of the method, the negative part of the signal surrounding the main lobe are suppressed by the network effectively reducing its contrast. This is likely due to the networks use of only one bipolar activation function (the antirectifier) followed by several strictly positive ones (the ReLU). Another problem that can be seen is the networks apparent streaking artifacts that are caused by its unawareness on the x-y plane.

An argument can be made however that the network significantly increases he resolution of the image. this can clearly be seen in figure 5(f) and (g). These two figures show an image reconstructed on a tighter grid. It shows a significantly less blurred and clearer image when zoomed in. Together with the visibility vascular structures that where previously more difficult to see.

2) AverageNet: The AverageNet also shows great results as it has similar results as the 1DNet in the field of visual clarity, while improving it by removing the streaking. However a new artifact that can be noticed is the irregular black outline surrounding the vascular structures. The problem is its irregularity compared to the consistent black outline surrounding the vascular structures in the other results.

The lack of streaking indicates that either the added depth of the network or the increase in receptive field in the element direction counter it. The issue of suppressing the negative part of the signal however persists.

3) PseudoNet: PseudoNet shows the most promising results. The overall visibility and clutter reduction is similar to that of 1DNet. However it does not have the streaking artifacts that 1DNet has, making it a more promising network to use. Its biggest downside however is getting it to work together with the BWN. The PseudoNet is the most memory and computationally expensive network, making it work together with the BWN would require a decrease in computational power as it requires up to 10 seconds to fully reconstruct an image compared to the 3 and 2 seconds of the 1DNet and AverageNet respectively.

Overall the networks mainly improved the resolution of the image. it decreases the noise as shown in 5(e) and it slightly improved the overall capability to find vascular structures near to the center that were previously difficult to find. However all of the networks are shown to be less capable when it comes to finding structures further from the center as none of the networks show any signal in the bottom or top left of the image where a vascular structure can clearly be seen in the DAS image.

The Network chosen and used for subsequent tests is the 1DNet because of its consistency compared to the AverageNet

methods	PSNR	SSIM
BWN only	11.42	**0.5415**
BWN + BFN trained on simulation data	**12.22**	0.5341
BWN + BFN trained on all data	12.06	0.3966

TABLE II
RESULTS FROM THE VARIOUS STRATEGIES

and its lower computational cost as compared to PseudoNet.

B. Only BWN

The bandwidth network is first fully trained on DAS beamformed simulated data to evaluate the resulting image without added contrast enhancements. The resulting images can be seen in the second column of figure 6 and the table with the results of a separate validation set can be seen in II. In these images it can be seen that the network is already showing good results when compared to the ground truth data for both the simulated and the *in vivo* image.

These results show that the bandwithNet on its own already shows good results on The simulated data. Its SSIM is even the highest found between the three training strategies. Its performance however is limited when moving further away from the center as can be seen in 6(f), where it shows less clear vascular structures in the regions further from the center.

The network on its own also gave good results on the *in vivo* data. It is able to recognise most of the vascular structures found. Its weaknesses lie in the center of the image and in the further regions of the image. both of these regions show that the system is either fitting to artifacts or not seeing the veins.

C. Complete network trained on simulated data

this network saw a significant jump in peak signal to noise ratio as compared to the bandwidth only network (see II. this increase can also be seen in figure 6(c) where the veins are much clearer as before.

The *in vivo* data also shows improvements. The grey area that was seen before on 6(j) and (n) has disappeared. The smaller vascular structures have also become clearer in figure 6(k) and especially 6(o). The artifacts near the center have, on the other hand, also become clearer.

D. Complete network trained on all data

The network trained on all the data shows diminishing returns when looking at the SSIM. However it can be observed that the network is able to find more structures in the further regions of the image. Which is surprising since it was observed earlier that the BFN showed detrimental results in the further region of the image as compared to the regular DAS data.

The *in vivo* images look more clear overall. The central part of both figure 6(l) and (p) show great improvement in the center of the image. However it does suffer in finding the smaller vascular structure as compared to the network trained on the simulation data.

VI. CONCLUSION

To conclude, in this paper a new network structure that is trained on both simulated and *in vivo* data is proposed. The simulated data is generated using the K-wave toolbox and the *in vivo* data is enhanced using Capon beamforming methods. The network consists of two parts a BFN and a BWN. These networks are trained using various strategies ranging from using only the simulated data on the bandwidth network to using the entire network with all the data. The bandwidth expansion network shows great improvements in image clarity and the beamforming network shows very promising results in increasing the resolution of the images. The combination of the two networks show very good results in increasing the resolution of the center of the image. The BFN however is unable to accurately find the negative parts of the signals and thus suppresses it leading to lower contrast. However it is also shown to be a powerful tool in increasing the networks resolution.

VII. DISCUSSION

The tests show that the generated data is a good representation of reality, however it can probably still be improved. results from the beamforming networks show that there are more structures visible in the image as was first estimated, adding similar structures to the simulate data could improve the overall result on the *in vivo* data.

Another idea for the simulation data could be by generating the complete vascular structure similar as in [22]. This could be used to more easily implement 3-dimensional structures and variable thickness of a vein in a single image.

For BandwithNet, the usage of an unfolded ISTA algorithm or a U-net structure for masking the veins as shown in [8] would probably improve the overall result of this network.

The choice of using 1DNet for the experiments with the BWN was also rushed. In hindsight the usage of the AverageNet would probably have resulted in better outcomes since it had less streaking artifacts.

One of the main limiting factors for this network is likely BFN's suppression of the negative part of the signal. The network shows a great improvement in resolution but a decrease in contrast due to the loss of the negative part. Getting the network to also enhance the negative part of the signal would likely increase the networks performance significantly.

Another possible improvement for improving the generalization of the beamforming network could be also training it on capon beamformed simulation data in addition to the enhanced *in vivo* data and intensity maps for the simulated data.

Finding speed of sound maps based on the initial data using neural networks shows promising results in ultrasound []. Another neural network can then be used to solve the complex Eikonal equation using for instance a Deep Eikonal Solver [28] to decrease the computational complexity of the network.

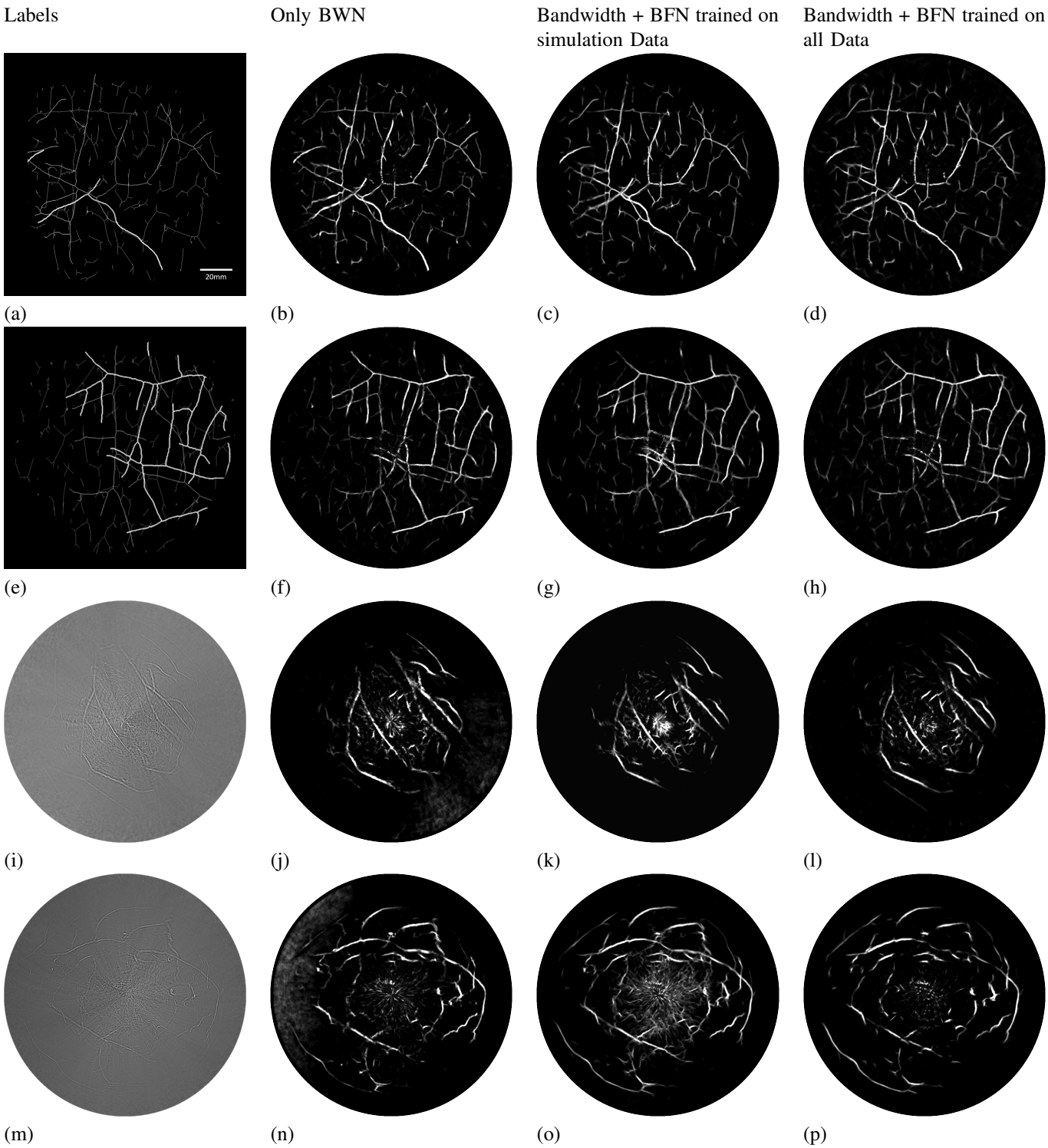


Fig. 6. Comparison of Networks and training strategies. The first column (a, e, i, m) consists of the ground truth for simulated data (a, e) and the rectangular DAS data for the *in vivo* data (i, m). The second column (b, f, j, n) has the results of the BWN trained on the rectangular DAS image and the ground truth data of the simulated intensity map. the third column (c, g, k, o) shows the results of the BFN and BWN trained on simulated data. The last column (d, h, l, p) shows the results of the BFN and BWN trained on both the simulated data and the enhanced *in vivo* data. The contrast of the images has been enhanced in post-processing for the purpose of visualization

REFERENCES

[1] X. Wang, Y. Pang, G. Ku, X. Xie, G. Stoica, and L. V. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain," *Nature Biotechnology*, vol. 21, pp. 803–806, July 2003. Number: 7 Publisher: Nature Publishing Group.

[2] X. Wang, X. Xie, G. Ku, L. V. Wang, and G. S. D.v.m, "Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography," *Journal of Biomedical Optics*, vol. 11, p. 024015, Mar. 2006. Publisher: International Society for Optics and Photonics.

[3] L. Lin, P. Hu, J. Shi, C. M. Appleton, K. Maslov, L. Li, R. Zhang, and L. V. Wang, "Single-breath-hold photoacoustic computed tomography of the breast," *Nature Communications*, vol. 9, no. 1, p. 2352, 2018.

[4] M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Physical Review E*, vol. 71, p. 016706, Jan. 2005. Publisher: American Physical Society.

[5] W. Choi, D. Oh, and C. Kim, "Practical photoacoustic tomography: Realistic limitations and technical solutions," *Journal of Applied Physics*, vol. 127, p. 230903, June 2020. Publisher: American Institute of Physics.

[6] J. Jose, R. G. H. Willemink, W. Steenbergen, C. H. Slump, T. G. van Leeuwen, and S. Manohar, "Speed-of-sound compensated photoacoustic tomography for accurate imaging," *Medical Physics*, vol. 39, pp. 7262–7271, Dec. 2012.

[7] A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064, p. 100643J, International Society for Optics and Photonics, Mar. 2017.

[8] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic Source Detection and Reflection Artifact Removal Enabled by Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1464–1477, June 2018. Conference Name: IEEE Transactions on Medical Imaging.

[9] M. Agranovsky and P. Kuchment, "Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography with variable sound speed," *Inverse Problems*, vol. 23, pp. 2089–2102, Aug. 2007. Publisher: IOP Publishing.

[10] H. Lan, D. Jiang, C. Yang, F. Gao, and F. Gao, "Y-Net: Hybrid deep learning image reconstruction for photoacoustic tomography in vivo," *Photoacoustics*, vol. 20, p. 100197, Dec. 2020.

[11] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE*, vol. 108, pp. 11–29, Jan. 2020. Conference Name: Proceedings of the IEEE.

[12] B. Luijten, R. Cohen, F. J. de Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. van Sloun, "Adaptive Ultrasound Beamforming using Deep Learning," *arXiv:1909.10342 [eess]*, Sept. 2019. arXiv: 1909.10342.

[13] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, Aug. 1969. Conference Name: Proceedings of the IEEE.

[14] Jian Li, P. Stoica, and Zhisong Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1702–1715, July 2003. Conference Name: IEEE Transactions on Signal Processing.

[15] S. Zhu, K. Yang, Y. Ma, Q. Yang, and X. Guo, "Robust minimum variance distortionless response beamforming using subarray multistage processing for circular hydrophone arrays," in *2016 Techno-Ocean (Techno-Ocean)*, pp. 692–696, Oct. 2016.

[16] Shengying Li, K. Mueller, M. Jackowski, D. P. Dione, and L. H. Staib, "Fast marching method to correct for refraction in ultrasound computed tomography," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pp. 896–899, Apr. 2006. ISSN: 1945-8452.

[17] C. Zhang and Y. Wang, "A reconstruction algorithm for thermoacoustic tomography with compensation for acoustic speed heterogeneity," *Physics in Medicine and Biology*, vol. 53, pp. 4971–4982, Aug. 2008. Publisher: IOP Publishing.

[18] M. A. Anastasio, J. Zhang, X. Pan, and L. V. Wang, "Half-time image reconstruction in photoacoustic tomography," *Photoacoustic Imaging and Spectroscopy*, pp. 155–163, Jan. 2017. Publisher: CRC Press.

[19] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, p. 021314, Apr. 2010.

[20] Y. Lou, "Optical and Acoustic Breast Phantoms," Apr. 2019. Publisher: Harvard Dataverse type: dataset.

[21] C. Tian, M. Pei, K. Shen, S. Liu, Z. Hu, and T. Feng, "Impact of System Factors on the Performance of Photoacoustic Tomography Scanners," *Physical Review Applied*, vol. 13, p. 014001, Jan. 2020. Publisher: American Physical Society.

[22] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes," *arXiv:1803.09340 [cs]*, Aug. 2019. arXiv: 1803.09340.

[23] L. Ding, "RECOVERY-FA19: Ultra-Widefield Fluorescein Angiography Vessel Detection Dataset," June 2019. Publisher: IEEE type: dataset.

[24] P. Hu, L. Li, L. Lin, and L. V. Wang, "Spatiotemporal Antialiasing in Photoacoustic Computed Tomography," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020. Conference Name: IEEE Transactions on Medical Imaging.

[25] Z. Qiu, T. Yao, and T. Mei, "Learning Spatio-Temporal Representation With Pseudo-3D Residual Networks," pp. 5533–5541, 2017.

[26] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, "Mixed Precision Training," *arXiv:1710.03740 [cs, stat]*, Feb. 2018. arXiv: 1710.03740.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), pp. 8024–8035, Curran Associates, Inc., 2019.

[28] M. Lichtenstein, G. Pai, and R. Kimmel, "Deep Eikonal Solvers," in *Scale Space and Variational Methods in Computer Vision* (J. Lellmann, M. Burger, and J. Modersitzki, eds.), Lecture Notes in Computer Science, (Cham), pp. 38–50, Springer International Publishing, 2019.