Deep neural networks to remove photoacoustic reflection artifacts in *ex vivo* and *in vivo* tissue

Derek Allman,* Fabrizio Assis,† Jonathan Chrispin,† and Muyinatu A. Lediju Bell*§‡
*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
§Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
‡Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
†Division of Cardiology, Johns Hopkins Medical Institutions, Baltimore, MD, USA

***Abstract*—Deep neural networks trained with simulated data are capable of distinguishing sources from reflection artifacts in photoacoustic data. Our group recently introduced this concept with simulated and experimental waterbath and phantom data. In this novel approach, channel data is used as an input to learn the spatial impulse response of pressure waves from point-like sources and differentiate true sources from reflection artifacts. We hypothesize that this is possible based on learned differences in the unique shape-to-depth relationship of point sources. The work presented in this paper builds on previous demonstrations to investigate the feasibility of this approach when applied to *ex vivo* tissue and *in vivo* data from a pig catheterization procedure. Three networks were trained with k-Wave simulated data: two residual deep network architectures (i.e., Resnet-50 and Resnet-101) and the previously implemented VGG16 deep network architecture, which does not use residual learning. These networks classified sources correctly in over 82% of images in the *ex vivo* chicken breast, liver, and steak datasets and over 64% of images in the dataset acquired from an *ex vivo* chicken thigh containing bone. When applied to *in vivo* data, the Resnet-50 and Resnet-101 architectures classified 83.3% and 88.8% of sources correctly, respectively, while the VGG16 architecture performed more poorly, classifying 14.5% of sources correctly. In addition, the residual network architectures had <2% misclassification rate, whereas the VGG16 architecture had a maximum 23.53% misclassification rate for all datasets. These results indicate that residual networks architectures are better suited to *in vivo* source detection and artifact elimination using our approach.**

I. Introduction

Photoacoustic imaging uses pulsed laser light to illuminate a region of interest, which absorbs the light, undergoes thermal expansion and converts the absorbed optical energy into mechanical pressure waves [1]. The resulting pressure signals are received with standard ultrasound imaging equipment for reconstruction into a photoacoustic image. Photoacoustic imaging has several promising applications including the detection and treatment of cancer [1]–[3] as well as surgical guidance and navigation [3]–[10].

Reflection artifacts are a major limitation to the clinical utility of photoacoustic imaging. These artifacts are caused by pressure waves from a photoacoustic source reflecting off of highly echoic objects in the surrounding tissue region, such as bone. Traditional beamforming techniques reconstruct these multipath reflection artifacts at greater depths than they truly appear, which can lead to misinterpretations by clinicians.

Previous work from our group demonstrates that a deep neural network can identify point sources from raw channel data [11], differentiate these sources from signals associated with reflection artifacts, and thereby remove reflection artifacts in photoacoustic images [12]–[14]. We trained deep neural networks to distinguish between sources and artifacts by learning the spatial impulse response function of photoacoustic point sources in simulated photoacoustic channel data. These deep neural networks correctly classified true sources in simulated data and additionally transferred learned knowledge to experimental photoacoustic data acquired from the circular cross section of a needle tip (i.e. a point-like source) located in a waterbath setup and the circular cross section of brachytherapy seeds embedded in a plastisol phantom.

Our previous work utilized the VGG16 deep network architecture [15]. Recently however, residual neural networks [16] have become the new state of the art for object recognition tasks. The depth of neural networks has been correlated with performance in various machine learning tasks. Non-residual networks, often referred to as plain deep networks, suffer from the vanishing gradient problem, which particularly occurs when the network becomes too deep and thus renders training intractable. Residual learning can be used to overcome this issue. In residual learning, skip connections are utilized to pass information from shallower layers to deeper layers allowing for greater information flow through the network. Residual learning has allowed for training of networks which exceed 100 layers in depth and have also been shown to improve performance over plain neural networks in practice [16].

This paper investigates the ability of our deep beamforming technique to transfer learned knowledge from simulated data to both *ex vivo* data from four tissue samples and *in vivo* data from a pig catheterization procedure. In addition, we investigate the potential of residual networks to improve our technique by directly comparing three network architectures (i.e., VGG16, Resnet-50, and Resnet-101) and their ability to transfer knowledge from the simulated domain to these *ex vivo* and *in vivo* test cases.

II. Methods

A. Simulated Datasets

Two datasets were generated corresponding to channel data acquired with imaging depths of 4.5 cm and 10 cm. These datasets were simulated in k-Wave [17]. Noting the importance of simulating an accurate receiver model during training [12],

the discrete receiver was modeled after the Alpinion L3-8 linear array ultrasound transducer, and the parameters for this model are reported in Table I.

A total of 19,992 photoacoustic channel data images were simulated for each dataset. Each image contained one true 0.1 mm-diameter source and one reflection artifact appearing with the same diameter as the source. Reflection artifacts were generated using the technique detailed in our previous work, shifting true sources deeper into the image [12]–[14] by the Euclidean distance, Δ, between the source location, (x_s, z_s), and the reflector location, (x_r, z_r), defined as:

$$|\Delta| = \sqrt{(z_s - z_r)^2 + (x_s - x_r)^2} \tag{1}$$

Additional parameters used to simulate the dataset are reported in Table II.

B. Deep Network Training

Three network architectures were trained with the simulated datasets: VGG16 [15], Resnet-50 [16], and Resnet-101 [16]. VGG16 is a plain deep network used in previous work and will represent our basis of comparison to Resnet-50 and Resnet-101 which are residual network architectures with 50 and 101 layers, respectively. Each network was used in conjunction with the Faster R-CNN algorithm [18] used within the Detectron software [19]. Each network was trained to detect and classify the peak of incoming acoustic waves as either source or artifact. Training was performed with 80% of the dataset, while the remaining 20% was used for validation. Faster R-CNN outputs a list of object detections for each class (i.e., source or artifact), along with the object location in terms of bounding-box pixel coordinates as well as a confidence score between 0 and 1 for each image. A total of 6 different networks were trained, one for each architecture and image depth combination.

The plain deep network, VGG16, was trained using an NVIDIA Titan X (Pascal) GPU for 100,000 iterations, corresponding to 5 epochs in total, and was initialized using a network pre-trained with the ImageNet dataset [20]. Training using this configuration took approximately 5.5 hours. The residual networks were trained using 2 NVIDIA Titan X (Pascal) GPUs for 30,000 iterations, corresponding to 3 epochs in total. Similar to the VGG16 the network, the residual networks were initialized using a network pre-trained with the ImageNet dataset [20]. The base learning rate used was 5×10^{-3} and decayed to 5×10^{-4} at iteration 15,000, and 5×10^{-5} at iteration 20,000. Training in this configuration took approximately 3 hours. Once trained, these networks provided detection results in 0.068 s, which translates to a frame rate of 14.7 Hz.

TABLE I: Simulated Acoustic Receiver Parameters

	Value
Kerf (mm)	0.06
Element Width (mm)	0.24
Sampling Frequency (MHz)	40

TABLE II: Range and Increment Size of Simulation Variables

Image Depth	Point Target Parameters	Min	Max	Increment
4.5 cm	Depth Position (mm)	5	25	0.25
	Lateral Position (mm)	5	30	0.25
	Channel SNR (dB)	-5	2	random
	Object Intensity (multiplier)	0.75	1.1	random
	Speed of Sound (m/s)	1440	1640	6
10 cm	Depth Position (mm)	50	95	0.25
	Lateral Position (mm)	5	30	0.25
	Channel SNR (dB)	-5	2	random
	Object Intensity (multiplier)	0.75	1.1	random
	Speed of Sound (m/s)	1440	1640	6

C. Ex Vivo Experiments

The trained networks were first tested on *ex vivo* photoacoustic channel data acquired from the circular cross section of a needle tip inserted in four tissue samples: (1) chicken breast, (2) liver, (3) steak, and (4) a chicken thigh containing bone. The hollow core biopsy needle contained a 1 mm core diameter optical fiber. The optical fiber was coupled to a Quantel (Bozeman, MT) Brilliant laser operating at 750 nm, and this setup enables the signals in the photoacoustic channel data to appear as if they originated from a point-like source. The point-like photoacoustic response from the needle tip was recorded with an Alpinion (Bothell, WA) E-Cube 12R scanner connected to an L3-8 linear array ultrasound transducer which was held in place by a Sawyer Robot (Rethink Robotics, Boston, MA). These datasets were previously acquired for photoacoustic-based visual servoing of the needle tip [21]. Seventeen channel datasets were acquired for each of the chicken breast, liver, and steak experiments, while 31 channel data sets were acquired in the chicken thigh. These channel data were acquired at an imaging depth of 4.5 cm. Therefore, these *ex vivo* data were tested with the 4.5 cm-deep networks.

D. In Vivo Experiments

The networks were additionally tested on *in vivo* data acquired during a pig catheterization procedure, which was performed with approval from the Johns Hopkins University Animal Care and Use Committee. The pig was positioned supine on an operating table and was fully anesthetized. To gain vascular access two 9F vascular sheaths were placed in the right femoral vein and artery using an ultrasound-guided micro-puncture technique. A bolus of heparin was administered after the sheath was secured in place. A 1 mm core diameter optical fiber was inserted into a 5F inner diameter, 7F outer diameter, 28 inch long cardiac catheter (St. Jude Medical, St. Paul, Minnesota, U.S.A.), which was inserted in the femoral vein sheath and advanced toward the heart. The optical fiber was coupled to a Phocus Mobile laser (Opotek, Carlsbad, California, U.S.A.) laser operating at 750 nm and imaged with an Alpinion (Bothell, WA) E-Cube 12R scanner connected to an L3-8 linear array ultrasound transducer which was held in place by a Sawyer Robot (Rethink Robotics, Boston, MA). Similar to the images of the needle tip, this setup enabled the circular cross section of the fiber tip within

the catheter to appear as a point-like photoacoustic source. A total of 279 channel datasets were acquired at an imaging depth of 10 cm. Therefore, these *in vivo* data were tested with the 10 cm-deep networks.

III. Results

The classification results for the three networks applied to the *ex vivo* and *in vivo* channel data is shown in Fig. 1. The VGG16 architecture correctly classified 100% of sources in the *ex vivo* chicken breast, liver, and steak datasets and 83.9% of the *ex vivo* chicken thigh dataset. In addition, the VGG16 architecture displayed no misclassifications for the chicken breast and steak datasets and a misclassification rate of 23.5% and 8.9% for the steak and chicken thigh datasets, respectively. Performance generally decreases with increasing levels of reflection and reverberation artifacts, considering that the presence of reflection and reverberation artifacts were minimal in the chicken breast data, greater in the liver data, and greatest in the steak and chicken thigh data (see Fig. 2).

The Resnet-50 and Resnet-101 network architectures correctly classified all chicken breast sources. The performance of these residual network architectures was worse in the remaining *ex vivo* tissue, where Resnet-50 and Resnet-101 correctly classified sources at rates of 82.35% and 94.11%, respectively, in the liver dataset, 82.35% and 88.24%, respectively, in the steak dataset, and 70.97% and 64.52%, respectively, in the chicken thigh dataset. There were no misclassifications in any of these *ex vivo* tissue samples with the residuals networks.

Despite the poor performance of the *in vivo* data with the VGG16 network architecture (i.e., 14.5% source classification and 85.5% missed detections), the residual network architectures correctly classified 83.3% (Resnet-50) and 88.8% (Resnet-101) of sources in the *in vivo* data. In addition, the misclassification rates for all three networks was less than 2%, while the missed detection rates for the residual networks was less than 15.9%.

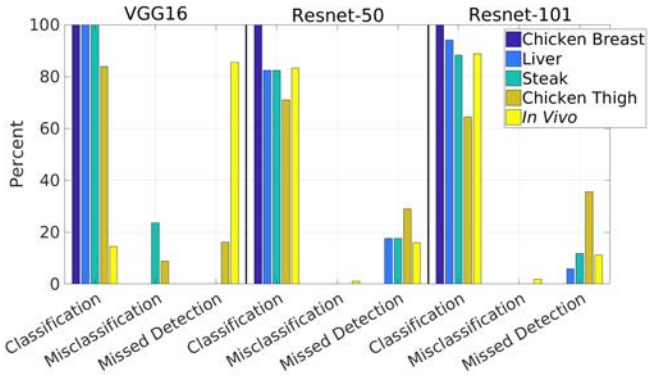

Fig. 1: Classification results for the four tissue samples and the *in vivo* data after testing with the VGG16, Resnet-50, and Resnet-101 networks. VGG16 does not transfer well to *in vivo* data, but the residual networks are more promising when tested with *in vivo* data.

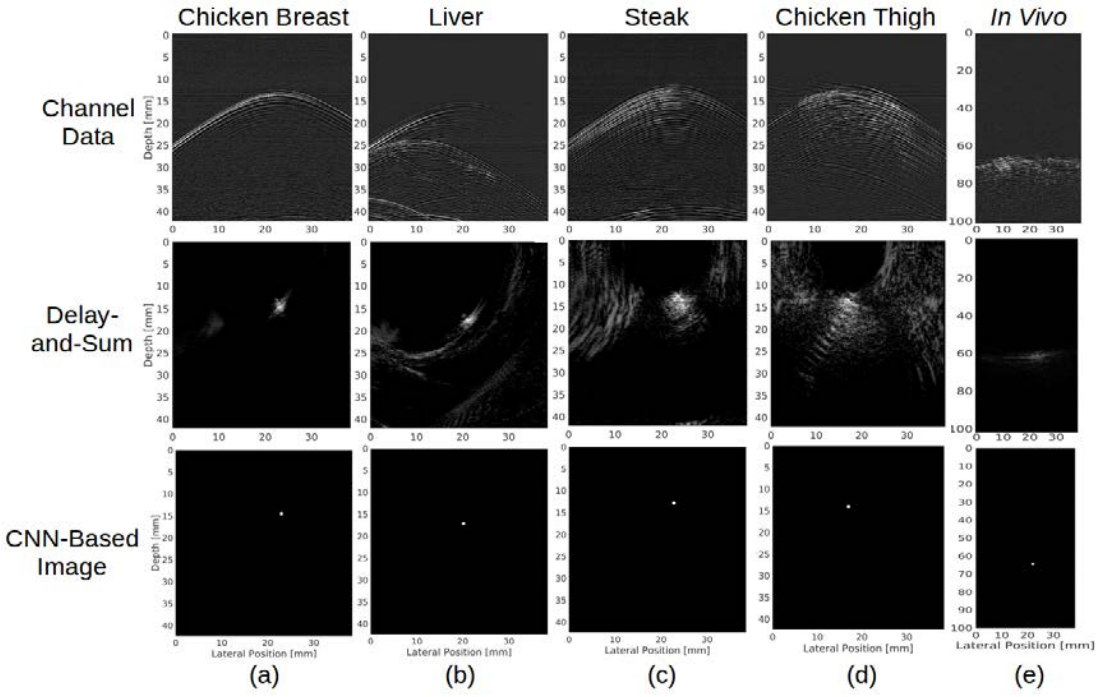

Fig. 2: Channel data corresponding to the five different experimental datasets used in this work: (a) chicken breast, (b) liver, (c) steak, (d) chicken thigh, and (e) *in vivo*. Beneath the channel data sample are the images corresponding delay and sum followed by CNN images generated using the outputs of the Resnet-101 network.

One channel data sample for each experimental dataset is displayed in Fig. 2, along with the corresponding delay-and-sum beamformed image. The CNN-based images [14] are generated from the detection results of the Resnet-101 network. The standard deviation of location errors when validating this network with simulated data were 0.038 mm and 0.058 mm for the 4.5 cm and 10 cm deep datasets, respectively. Standard deviations are presented at 10 times their true value for clarity and the location of the circle represents the detected source location at the center of the detection bounding box.

This comparison demonstrates that the networks can successfully recover the location of the optical fiber tip and display an image that is artifact free and has arbitrarily high contrast with frame rates suitable for real-time imaging.

IV. Discussion

This work advances our previous work [14] by investigating the clinical utility of our deep beamforming technique as an alternative to traditional time-of-flight based beamformers. We successfully demonstrated that a deep network trained with only simulated data can transfer learned knowledge to *ex vivo* and *in vivo* data with no additional training. Our focus is identifying point-like targets, manifested as an optical fiber housed in a hollow-core biopsy needle or a cardiac catheter.

There are three important characteristics of these network results. First, the VGG16 architecture seems to struggle with providing accurate detections *in vivo* when compared to the residual networks. One potential reason for this challenge is the depth of the channel data. When the depth of the channel data increases, sources and artifacts tend to look similar because the wavefronts appear with less curvature. Given this decreased variance in the wavefront shapes, we suspect that a deeper network (i.e., the residual network), has greater capacity to learn higher-level features. Second, all networks experienced notable performance decreases in the presence of bone, such as in the chicken thigh dataset, which contains the most reflection artifacts. As a result, the classification rates of our networks decrease. One possible solution to this challenge is temporal averaging of the network results. Despite the lower classification rates in the presence of bone, the corresponding misclassification rates remain low, indicating successful elimination of artifacts. Finally, the residual networks maintained a misclassification rate $<2\%$ for all cases, indicating their suitability for the task of artifact elimination.

V. Conclusion

We trained both plain and residual deep neural network architectures using simulated photoacoustic channel data to distinguish between sources and artifacts and transferred learned knowledge from the simulated domain to *ex vivo* and *in vivo* channel data. With no additional training, these networks successfully detected sources in *ex vivo* and *in vivo* data, which has promising applications for the development of clinical and interventional photoacoustic imaging systems.

References

[1] M. Xu and L. V. Wang, "Photoacoustic imaging in biomedicine," *Review of Scientific Instruments*, vol. 77, no. 4, p. 041101, 2006.

[2] M. A. L. Bell, N. P. Kuo, D. Y. Song, J. U. Kang, and E. M. Boctor, "In vivo visualization of prostate brachytherapy seeds with photoacoustic imaging," *Journal of Biomedical Optics*, vol. 19, no. 12, pp. 126 011–126 011, 2014.

[3] M. A. L. Bell, N. Kuo, D. Y. Song, and E. M. Boctor, "Short-lag spatial coherence beamforming of photoacoustic images for enhanced visualization of prostate brachytherapy seeds," *Biomedical Optics Express*, vol. 4, no. 10, pp. 1964–1977, 2013.

[4] M. A. L. Bell, A. K. Ostrowski, K. Li, P. Kazanzides, and E. M. Boctor, "Localization of transcranial targets for photoacoustic-guided endonasal surgeries," *Photoacoustics*, vol. 3, no. 2, pp. 78–87, 2015.

[5] M. Allard, J. Shubert, and M. A. L. Bell, "Feasibility of photoacoustic guided teleoperated hysterectomies," *Journal of Medical Imaging: Special Issue on Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 5, no. 2, p. 021213, 2018.

[6] N. Gandhi, M. Allard, S. Kim, P. Kazanzides, and M. Lediju Bell, "Photoacoustic-based approach to surgical guidance performed with and without a da vinci robot," *Journal of Biomedical Optics*, vol. 22, no. 12, p. 121606, 2017.

[7] D. Piras, C. Grijsen, P. Schütte, W. Steenbergen, and S. Manohar, "Photoacoustic needle: minimally invasive guidance to biopsy," *Journal of Biomedical Optics*, vol. 18, no. 7, pp. 070 502–070 502, 2013.

[8] W. Xia, D. I. Nikitichev, J. M. Mari, S. J. West, R. Pratt, A. L. David, S. Ourselin, P. C. Beard, and A. E. Desjardins, "Performance characteristics of an interventional multispectral photoacoustic imaging system for guiding minimally invasive procedures," *Journal of Biomedical Optics*, vol. 20, no. 8, pp. 086 005–086 005, 2015.

[9] W. Xia, E. Maneas, D. I. Nikitichev, C. A. Mosse, G. S. dos Santos, T. Vercauteren, A. L. David, J. Deprest, S. Ourselin, P. C. Beard *et al.*, "Interventional photoacoustic imaging of the human placenta with ultrasonic tracking for minimally invasive fetal surgeries," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 371–378.

[10] J. Shubert and M. A. L. Bell, "Photoacoustic imaging of a human vertebra: implications for guiding spinal fusion surgeries," *Physics in Medicine and Biology*, 2018.

[11] A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Proc. of SPIE*, vol. 10064, 2017, pp. 100 643J–1.

[12] D. Allman, A. Reiter, and M. Bell, "Exploring the effects of transducer models when training convolutional neural networks to eliminate reflection artifacts in experimental photoacoustic images," in *Proc. of SPIE*, vol. 10494, 2018, pp. 10 494–190.

[13] ——, "A machine learning method to identify and remove reflection artifacts in photoacoustic channel data," in *Proceedings of the 2017 IEEE International Ultrasonics Symposium*. International Ultrasonic Symposium, 2017.

[14] 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, no. 6, pp. 1464–1477, 2018.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations (ICLR), 2015*, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

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

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[19] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," https://github.com/facebookresearch/detectron, 2018.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[21] J. Shubert and M. A. L. Bell, "Photoacoustic based visual servoing of needle tips to improve biopsy on obese patients," in *Ultrasonics Symposium (IUS), 2017 IEEE International*. IEEE, 2017, pp. 1–4.