A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions

Allman, Derek, Assis, Fabrizio, Chrispin, Jonathan, Lediju Bell, Muyinatu

Derek Allman, Fabrizio Assis, Jonathan Chrispin, Muyinatu A. Lediju Bell, "A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions," Proc. SPIE 10878, Photons Plus Ultrasound: Imaging and Sensing 2019, 108785E (27 February 2019); doi: 10.1117/12.2510993

SPIE.

Event: SPIE BiOS, 2019, San Francisco, California, United States

A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions

Derek Allman[a], Fabrizio Assis[b], Jonathan Chrispin[b], and Muyinatu A. Lediju Bell[a,c,d]

[a]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
[b]Division of Cardiology, Johns Hopkins Medical Institutions, Baltimore, MD, USA
[c]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[d]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Interventional cardiac procedures often require ionizing radiation to guide cardiac catheters to the heart. To reduce the associated risks of ionizing radiation, our group is exploring photoacoustic imaging in conjunction with robotic visual servoing, which requires segmentation of catheter tips. However, typical segmentation algorithms are susceptible to reflection artifacts. To address this challenge, signal sources can be identified in the presence of reflection artifacts using a deep neural network, as we previously demonstrated with a linear array ultrasound transducer. This paper extends our previous work to detect photoacoustic sources received by a phased array transducer, which is more common in cardiac applications. We trained a convolutional neural network (CNN) with simulated photoacoustic channel data to identify point sources. The network was tested with an independent simulated validation data set not included during training as well as *in vivo* data acquired during a pig catheterization procedure. When tested on the independent simulated validation data set, the CNN correctly classified 84.2% of sources with a misclassification rate of 0.01%, and the mean absolute location error of correctly classified sources was 0.095 mm and 0.462 mm in the axial and lateral dimensions, respectively. When applied to *in vivo* data, the network correctly classified 91.4% of sources with a 7.86% misclassification rate. These results indicate that a CNN is capable of identifying photoacoustic sources recorded by phased array transducers, which is promising for cardiac applications.

1. INTRODUCTION

Cardiac interventional procedures are often performed to diagnose and treat cardiac arrhythmias. These procedures require catheter guidance from an insertion point in the patient's thigh to the heart via the femoral vein. Guidance is typically performed with fluoroscopy[1,2] or intracardiac echocardiography,[3] however, there are several challenges with these techniques. Fluoroscopy exposes both patients and operators to ionizing radiation, and it is difficult to determine the depth of the catheter in these x-ray projection images. While intracardiac ultrasound imaging generally provides suitable views of a cardiac catheter, it requires additional fluoroscopy, electromagnetic tracking, and skilled operators in order to provide a global reference frame.[4] Ultrasound imaging is also plagued by acoustic clutter[5] (which can obscure the catheter tip) and shadowing from the ribs (which further limits localization potential).[6]

Photoacoustic imaging coupled with robotic visual servoing was previously investigated as a method to guide biopsy needles in phantoms and *ex vivo* tissue samples.[7,8] Photoacoustic imaging utilizes pulsed laser light to excite optically active absorbers in a region of interest. These absorbers convert the optical energy to acoustic energy, in the form of mechanical pressure waves, via the photoacoustic effect, which can be sensed by a standard ultrasound transducer and reconstructed to create a photoacoustic image.[9,10] A robotic arm can then be used to hold the ultrasound transducer as a visual servoing algorithm segments the tip of the optical fiber in the beamformed image.[7,8] The robot would then track the fiber tip by guiding the probe to the desired location.

The robotic photoacoustic imaging method described above has the potential to overcome limitations with existing catheter guidance techniques by limiting radiation exposure as well as providing a global frame of reference via the robotic arm. However, reflection artifacts resulting from highly echoic structures, like bone, are

Photons Plus Ultrasound: Imaging and Sensing 2019, edited by Alexander A. Oraevsky, Lihong V. Wang,
Proc. of SPIE Vol. 10878, 108785E · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2510993

problematic in photoacoustic imaging as they cause bright reflections in the beamformed image. The segmentation step of the visual servoing algorithm is susceptible to these reflection artifacts as it utilizes brightness information to identify the fiber tip in the image.

As an alternative to the error-prone segmentation step in visual servoing, we propose a deep learning approach to detect the optical fiber tip, which is based on our previous work.[11–15] We previously demonstrated that a deep learning approach can be trained with simulated data to detect photoacoustic point sources,[11–15] including photoacoustic signals originating from an optical fiber tip housed in a needle surrounded by water,[12–14] a needle surrounded by *ex vivo* tissue,[15] and a cardiac catheter located in an *in vivo* femoral vein.[15] Our previous work also demonstrates the importance of correctly modeling the ultrasound receiver when implementing deep learning to detect photoacoustic sources and remove reflection artifacts.[13] These major contributions were demonstrated with a linear array ultrasound transducer.[11–15] For cardiac imaging applications, phased array transducers are more desirable due to their lower acoustic frequencies (which enable increased imaging depths), their smaller physical footprint when imaging between the ribs, and their larger image field of view relative to their small physical footprint.

In this paper, we apply our previously developed deep learning artifact removal and source detection approaches[11–15] to a phased array ultrasound transducer. We train a network with simulated phased array channel data and introduce a new approach to setting up the imaging geometry for this training task. To ensure robustness to channel noise, target amplitude, and speed of sound differences, we generate a training set which includes multiple noise levels, signal amplitudes, and sound speeds. We then test our network on simulated validation data and *in vivo* data acquired during a pig catheterization procedure.

2. METHODS

2.1 Simulated Datasets for Training

A dataset was simulated corresponding to a phased array acoustic receiver architecture. The channel data was simulated in k-Wave[16] with an imaging depth of 12 cm and the additional parameters given in Table 1. The phased array receiver was modeled after the Alpinion (Bothell, WA) SP1-5 phased array ultrasound transducer, and the parameters for this model are shown in Table 2.

A total of 20,000 photoacoustic channel data samples were generated for the dataset. Each image contained one true source and one reflection artifact, each with a diameter of 0.1 mm. Reflection artifacts were generated according to the technique detailed in our previous work,[12–14] which shifts a true source signal deeper into the image by the Euclidean distance between the source location and the reflector location.

In previous work with a linear array transducer,[12–15] the input to the network was the channel data along with labels corresponding to the source or artifact location in the channel data. However, in the phased array case, the locations of the sources and artifacts can extend beyond the width of the array footprint. Therefore, the input image to the phased array network must correspond to the area of the scan-converted phased array

Table 1: Range and Increment Size of Simulation Variables

Parameters	Min	Max	Increment
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

Table 2: Simulated Acoustic Receiver Parameters

Parameter	Value
Kerf (mm)	0.08
Element Width (mm)	0.22
Sampling Frequency (MHz)	40

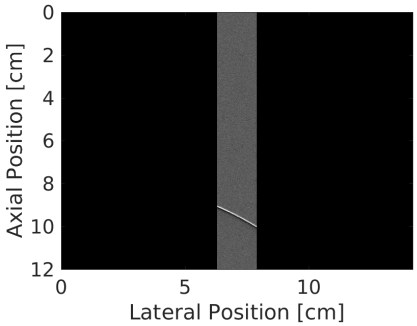


Figure 1: Example image used for training the network. The channel data is present in the middle of the image, and the black region on the left and right corresponds to the full size of a scan-converted image.

image. To implement this additional constraint, channel data was overlaid onto a black image corresponding to the size of the scan-converted phased array image, as demonstrated in Fig. 1.

2.2 Training Methods

We trained a Resnet-101 network[17] (i.e., a residual network with 101 layers) using the dataset described in Section 2.1. The Resnet-101 network was trained with the Faster R-CNN algorithm[18] with the Detectron software package.[19] The network was trained to detect and classify the peak of the incoming acoustic waves as either source or artifact. Since the location of the object generating the acoustic wave is not necessarily contained in the channel data, the peak of the wavefront may be occluded. 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.

Detections were considered correct if the intersect-over-union (IoU) of the ground truth and detection bounding boxes was greater than 0.5 and their score was greater than an optimal score. The optimal score was calculated based on the receiver-operating-characteristics (ROC) curve, which evaluates the true positive rate and false positive rate for a range of confidence score thresholds and plots one data point for each confidence threshold. Positive detections were defined as detections with an IoU of 0.5. The ROC curve expresses the overall quality of the detections made by the network.

The optimal score for each class was found by first defining a line with a slope equal to the number of negative detections divided by the number of positive detections. This line was then shifted from the ideal operating point (true positive rate of 1 and false positive rate of 0) down and to the right until it intersected with the ROC curve. The first intersection of this line with the ROC curve was determined to be the optimal score threshold. Misclassifications were defined to be a source detected as an artifact or an artifact detected as a source, and missed detections were defined as a source or artifact being detected as neither a source nor artifact.

We also consider metrics for precision, recall, and area-under-the-curve (AUC). Where precision is the fraction of correct detections over the total number of positive detections, and recall is defined as the fraction of correct detections over the total number of objects which should have been detected (note that recall and classification rate are equivalent in this work). AUC was defined as the area under the ROC curve.

The deep network was trained using 2 Nvidia Titan X (Pascal) GPUs for 30,000 iterations, corresponding to 3 epochs in total, and was 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, the network provided detection results in 0.093 s per image, which translates to a frame rate of 10.8 Hz.

2.3 Validation and Testing with Simulated and In Vivo Data

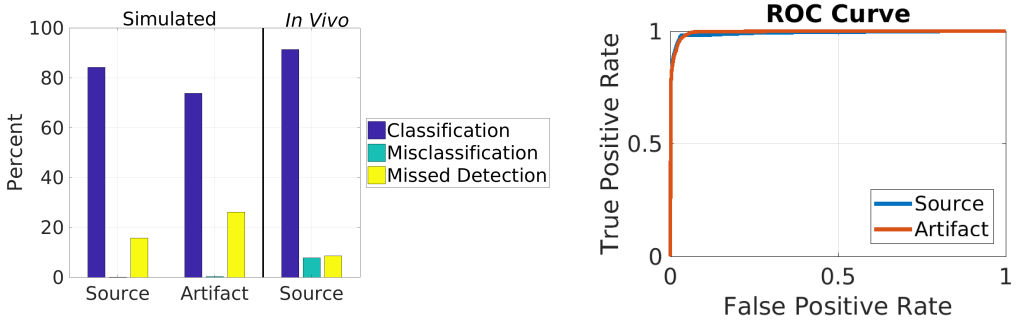
The remaining 20% of the simulated dataset which was not used during training was evaluated by calculating source and artifact classification, misclassification, and missed detection rates as well as precision, recall, AUC, and distance error of correct detections.

In addition, the trained network was 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 on an operating table in a supine position and fully anesthetized. A 1 mm core-diameter optical fiber was inserted into a 5F inner-diameter cardiac catheter (St. Jude Medical, St. Paul, Minnesota, U.S.A.), which was inserted in a 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. The fiber tip appears as a point source in the photoacoustic channel data when imaged by an Alpinion (Bothell, WA) E-Cube 12R scanner connected to an SP1-5 phased array ultrasound transducer which was held in place by a Sawyer Robot (Rethink Robotics, Boston, MA). A total of 140 channel dataset samples were acquired at an imaging depth of 12 cm. Source classification, misclassification, and missed detection rate metrics were used to compare the performance of the *in vivo* dataset to the performance of the simulated validation set.

3. RESULTS

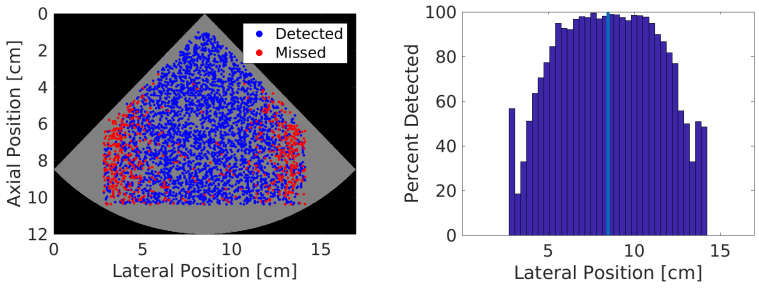
The classification results for the simulated and *in vivo* datasets are shown in Fig. 2(a). The network correctly classified 84.25% and 73.87% of simulated sources and artifacts, respectively. The network correctly classified 91.42% of 140 source samples present in the *in vivo* images. The misclassification rate was 0% for the simulated set and 7.86% for the *in vivo* set. Fig. 2(b) shows the ROC curves for simulated sources and artifacts. The AUC was 0.991 and 0.996 for sources and artifacts, respectively. Precision was 0.907 and 0.921 for sources and artifacts, respectively. Recall was 0.843 and 0.739 for sources and artifacts, respectively.

Fig. 3(a) shows source detections and missed source detections relative to their locations within the imaging plane after scan conversion. Qualitatively, the network classifies sources at a higher rate near the middle of the transducer. Fig. 3(b) shows a histogram of the source classification rate as a function of the lateral position in



(a) (b)

Figure 2: (a) Classification results for the simulated validation and *in vivo* test datasets. (b) Source and artifact ROC curves for the simulated dataset.



(a) (b)

Figure 3: (a) Map of source detections and missed source detections for the simulated validation dataset overlaid on the scan converted image field of view. (b) Histogram showing the source classification rate as a function of lateral position with the light blue line indicating the center of the transducer.

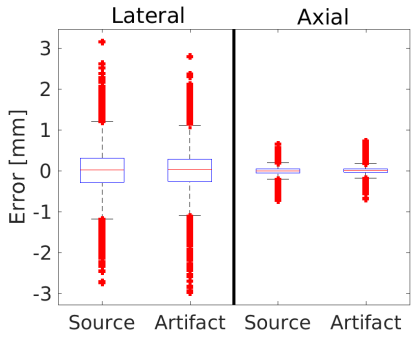


Figure 4: Summary of source and artifact location errors in the lateral and axial dimensions of the simulated validation dataset.

the image (for all axial positions shown). The classification rate is as high as 99.4% near the image center and falls as low as 18.6% as lateral distance from the image center increases.

Fig. 4 displays source and artifact axial and lateral location errors as box-and-whiskers plots. The top and bottom of each box represents the 75th and 25th percentiles of the measurements, respectively. The line inside each box represents the median measurement, and the whiskers (i.e. lines extending above and below each bot) represent the range. Outliers were defined as any value greater than 1.5 times the interquartile range and are displayed as dots. The mean absolute error in the axial dimension was 0.095 mm and 0.084 mm for sources and artifacts, respectively. The mean absolute error in the lateral dimension was 0.462 mm and 0.445 mm for sources and artifacts, respectively. In general, axial errors were consistently lower than lateral errors.

Fig. 5 shows example channel data, traditional delay-and-sum beamformed images, and our new approach that we call CNN-based images. These images are shown for both simulated (top) and *in vivo* (bottom) data. The CNN-based image contains a white circle on a black background at the center of the detected source bounding box. The radius of the circle is equal to three times the mean absolute lateral error reported for sources when testing the network with simulated channel data.

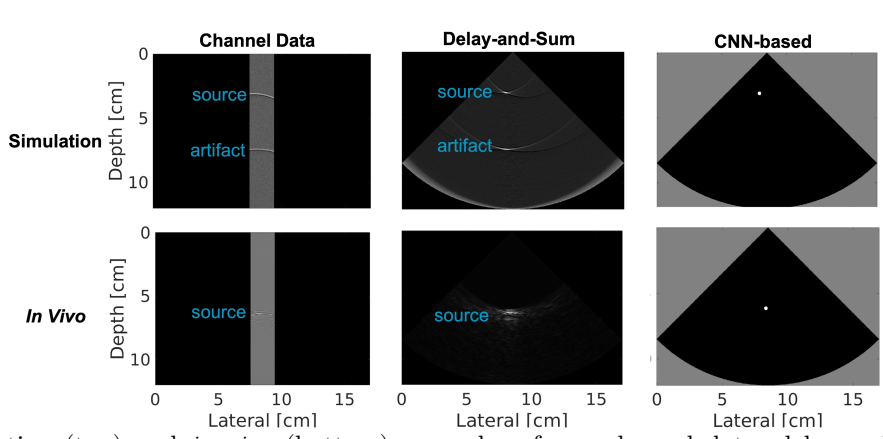


Figure 5: Simulation (top) and *in vivo* (bottom) examples of raw channel data, delay-and-sum images, and CNN-based images obtained with a phased array transducer. The CNN-based images contain a white circle presented on a black background to demonstrate the location of detected sources in the imaging field.

4. DISCUSSION

The networks tested in this paper expand our previous networks developed for linear array transducer data to phased array transducer data. We demonstrate that a deep network can correctly detect photoacoustic point-like sources in a scan converted phased array photoacoustic image. In addition, this deep network can be transferred

to data acquired during in an *in vivo* pig catheterization procedure and still maintain high classification rates of 91.4%, low misclassification rates of 7.86%, with submillimeter location errors.

It is interesting to note that the source classification rate when the network was applied to the *in vivo* dataset was higher than that of the simulated dataset (91.42% vs. 84.35%). This difference can possibly be explained by the results shown in Fig. 3, which suggests that the network classifies sources more accurately toward the center of the transducer. Similarly, the sources in our *in vivo* dataset were mostly located near the center of our transducer, and the network classified >90% of these sources correctly.

When compared to our previous work with the linear array transducer[14] where 91.6% of sources were correctly classified, the phased array transducer network saw a decrease in classification accuracy to 83.6%. We note that classification accuracy in the simulated domain shows a dependence on the objects location in the imaging field (Fig. 3), as the classification rate varied between 99.4% and 18.6% depending on its distance from the transducer center. Despite the transducer only being 1.84 cm wide this model was able to correctly classify sources up to 5.60 cm away from the transducer center. In addition, the reported axial and lateral errors are comparable to the linear transducer case (i.e., $\leq$1 mm mean errors in both the axial and lateral dimensions). These results are promising for detecting catheter tips in photoacoustic-based visual servoing applications, particularly when considering that the catheter tip location was tested at depths as large as 10 cm from the skin surface.

5. CONCLUSION

This work is the first to apply deep neural networks to source detection and artifact removal in phased array photoacoustic imaging. After training with simulated data, we successfully transferred these networks to *in vivo* data. Promising applications of this work include robotic visual servoing during photoacoustic guidance of catheter and needle tips. Our deep learning source detection techniques have demonstrated potential to increase the robustness of the source segmentation step in these photoacoustic-based visual servoing tasks.

ACKNOWLEDGEMENTS

This work is partially supported by NIH Trailblazer Award R21-EB025621 and NSF CAREER Award 1751522.

REFERENCES

[1] Calkins, H., Kuck, K. H., Cappato, R., Brugada, J., Camm, A. J., Chen, S.-A., Crijns, H. J., Damiano, R. J., Davies, D. W., DiMarco, J., et al., "2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design: a report of the Heart Rhythm Society (HRS) Task Force on Catheter and Surgical Ablation of Atrial Fibrillation," *Heart Rhythm* **9**(4), 632–696 (2012).

[2] Yatziv, L., Chartouni, M., Datta, S., and Sapiro, G., "Toward multiple catheters detection in fluoroscopic image guided interventions," *IEEE Transactions on Information Technology in Biomedicine* **16**(4), 770–781 (2012).

[3] Kanj, M., Wazni, O., and Natale, A., "Pulmonary vein antrum isolation," *Heart Rhythm* **4**(3), S73–S79 (2007).

[4] Bartel, T., Müller, S., Biviano, A., and Hahn, R. T., "Why is intracardiac echocardiography helpful? benefits, costs, and how to learn," *European Heart Journal* **35**(2), 69–76 (2013).

[5] Lediju, M. A., Pihl, M. J., Dahl, J. J., and Trahey, G. E., "Quantitative assessment of the magnitude, impact and spatial extent of ultrasonic clutter," *Ultrasonic Imaging* **30**(3), 151–168 (2008).

[6] Sperandeo, M., Rotondo, A., Guglielmi, G., Catalano, D., Feragalli, B., and Trovato, G. M., "Transthoracic ultrasound in the assessment of pleural and pulmonary diseases: use and limitations," *La Radiologia Medica* **119**(10), 729–740 (2014).

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

[8] Bell, M. A. L. and Shubert, J., "Photoacoustic-based visual servoing of a needle tip," *Scientific Reports* **8**(1), 15519 (2018).

[9] Beard, P., "Biomedical photoacoustic imaging," *Interface Focus* , rsfs20110028 (2011).
[10] Xu, M. and Wang, L. V., "Photoacoustic imaging in biomedicine," *Review of Scientific Instruments* **77**(4), 041101 (2006).
[11] Reiter, A. and Bell, M. A. L., "A machine learning approach to identifying point source locations in photoacoustic data," in [*Proc. of SPIE*], **10064**, 100643J–1 (2017).
[12] Allman, D., Reiter, A., and Bell, M. A. L., "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).
[13] Allman, D., Reiter, A., and Bell, M. A. L., "Exploring the effects of transducer models when training convolutional neural networks to eliminate reflection artifacts in experimental photoacoustic images," in [*Proc. of SPIE*], **10494**, 10494–190 (2018).
[14] Allman, D., Reiter, A., and Bell, M. A. L., "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Transactions on Medical Imaging* **37**(6), 1464–1477 (2018).
[15] Allman, D., Assis, F., Chrispin, J., and Bell, M. A. L., "Deep neural networks to remove photoacoustic reflection artifacts in ex vivo and in vivo tissue," in [*2018 IEEE International Ultrasonics Symposium (IUS)*], 1–4, IEEE (2018).
[16] Treeby, B. E. and Cox, B. T., "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave-fields," *J. Biomed. Opt.* **15**(2), 021314 (2010).
[17] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 770–778 (2016).
[18] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," in [*Advances in Neural Information Processing Systems*], 91–99 (2015).
[19] Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., and He, K., "Detectron." https://github.com/facebookresearch/detectron (2018).
[20] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "ImageNet: A Large-Scale Hierarchical Image Database," in [*CVPR09*], (2009).