# Deep learning to detect catheter tips *in vivo* during photoacoustic-guided catheter interventions

*Invited Presentation*

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*—Catheter guidance is typically performed with fluoroscopy, which requires patient and operator exposure to ionizing radiation. Our group is exploring robotic photoacoustic imaging as an alternative to fluoroscopy to track catheter tips. However, the catheter tip segmentation step in the photoacoustic-based robotic visual servoing algorithm is limited by the presence of confusing photoacoustic artifacts. We previously demonstrated that a deep neural network is capable of detecting photoacoustic sources in the presence of artifacts in simulated, phantom, and *in vivo* data. This paper directly compares the *in vivo* results obtained with linear and phased ultrasound receiver arrays. Two convolutional neural networks (CNNs) were trained to detect point sources in simulated photoacoustic channel data and tested with *in vivo* images from a swine catheterization procedure. The CNN trained with a linear array receiver model correctly classified 88.8% of sources, and the CNN trained with a phased array receiver model correctly classified 91.4% of sources. These results demonstrate that a deep learning approach to photoacoustic image formation is capable of detecting catheter tips during interventional procedures. Therefore, the proposed approach is a promising replacement to the segmentation step in photoacoustic-based robotic visual servoing algorithms.**

## I. Introduction

Catheter guidance is often used in cardiac interventions, with catheter tip visualization provided by fluoroscopy. However, fluoroscopy requires ionizing radiation and lacks depth information within the planar projection images. Photoacoustic imaging coupled with robotic visual servoing is an alternative which does not require ionizing radiation [1]. Photoacoustic imaging is implemented by sending laser pulses through an optical fiber that is housed inside the cardiac catheter, which generates a photoacoustic effect at the catheter tip (i.e., light is converted to sound), and the resulting sound is received by an ultrasound transducer array [2], [3].

To implement visual servoing, point-like photoacoustic signals generated at the catheter tip are segmented and the positions of the segmented signals are used to command a robot that is holding an ultrasound transducer to keep it centered on the photoacoustic signal at all times [1]. This approach is beneficial for freeing the hands of operators from both holding an ultrasound transducer and searching for the photoacoustic signal associated with the catheter tip. One additional benefit is that a global reference frame for catheter tip positions can be provided when the external ultrasound transducer (used to receive the photoacoustic images) is affixed to the end effector of a robotic arm.

Despite these major benefits, one challenge that we have encountered with photoacoustic-based visual servoing is that the segmentation algorithms that enable visual servoing are highly susceptible to artifacts and noise in the images that are used to guide the robot to its next position. Therefore, we propose to apply our novel deep learning photoacoustic point source detection technique [4], [5] to identify catheter tips in the presence of these problematic artifacts in raw channel data, prior to implementing the beamforming step that is currently needed to form photoacoustic images.

## II. Methods

Two independent convolutional neural networks (CNNs) were trained with simulated data created with the k-Wave [6] acoustic simulation package. One CNN was trained with a dataset corresponding to linear array [7] and the second CNN was trained with a dataset corresponding to a phased array [8]. The k-Wave [6] simulation software was used to create photoacoustic channel data from point sources and reflection artifacts that can be confused for photoacoustic sources after beamforming. The goal of the training step was to locate true point sources and differentiate them from reflection artifacts.

Images of the raw photoacoustic channel data that were used for training contained one source and one reflection artifact after varying medium sound speeds, target locations, and noise levels. Each CNN was trained using the Faster Region-based CNN algorithm [9] Resnet-101 [10] architecture to identify source and reflection artifacts in the channel data for 4,000 simulated test images per transducer architecture.

The trained networks were then transferred to *in vivo* data acquired while guiding a catheter through a swine femoral vein. This *in vivo* experiment was performed with approval from the Johns Hopkins University Animal Care and Use Committee. A total of 279 and 140 images were acquired with the linear array and phased array transducers, respectively.

## III. Results

Fig. 1 shows the results of testing each trained network on simulated and *in vivo* data. The linear array network correctly classified 99.1% and 88.8% of sources for the simulated and *in*

978-1-7281-1151-3/19/$31.00 ©2019 IEEE

*vivo* datasets, respectively. The phased array network correctly classified 84.3% and 91.4% of sources in the simulated and *in vivo* datasets, respectively. These high classification rates show that a deep network trained with only simulated data can transfer knowledge from the simulated domain to *in vivo* images with no additional training. In addition, these networks are suitable for eliminating artifacts as all networks produced less than 2-7.9% misclassification rate in all test sets.

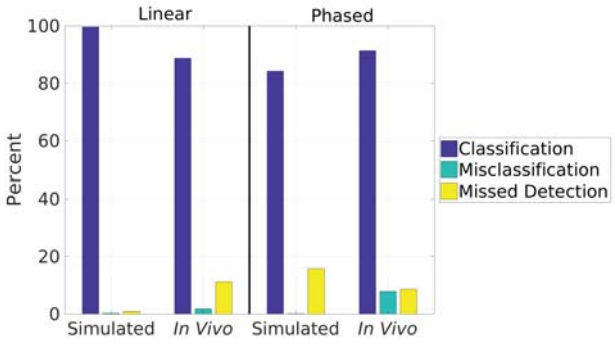


Fig. 1: Classification results for the linear and phased array networks where dark blue bars indicate classification rate, light blue bars indicate misclassification rate, and yellow bars indicate missed detection rate.

Fig. 2 shows channel data, delay-and-sum (DAS) images, CNN-based images, and network detections overlaid on DAS images. The size of the overlaid circle in the CNN-based and overlay images is proportional to the mean absolute error of the network. The mean absolute error of the network was 0.279 mm and 0.478 mm for the linear and phased array results, respectively. The red boxes in the DAS images highlight the regions zoomed in regions presented in the overlaid DAS and CNN images. This overlay allows us to compare the DAS brightness image to the detected locations generated by the network, which are shown in yellow.

## IV. Discussion & Conclusion

Deep learning is a promising approach to detect point-like sources in photoacoustic images. The approach is applicable to multiple transducer arrays, including linear and phased arrays. In both cases classification rates exceed 84% and misclassification rates are below 7.9%. Therefore, the proposed technique is well suited for the task of detecting sources and eliminating artifacts seen in photoacoustic images obtained with both linear and phased arrays.

These results demonstrate the feasibility of replacing the segmentation step in visual servoing with a CNN-based approach to image formation. While results are beneficial for

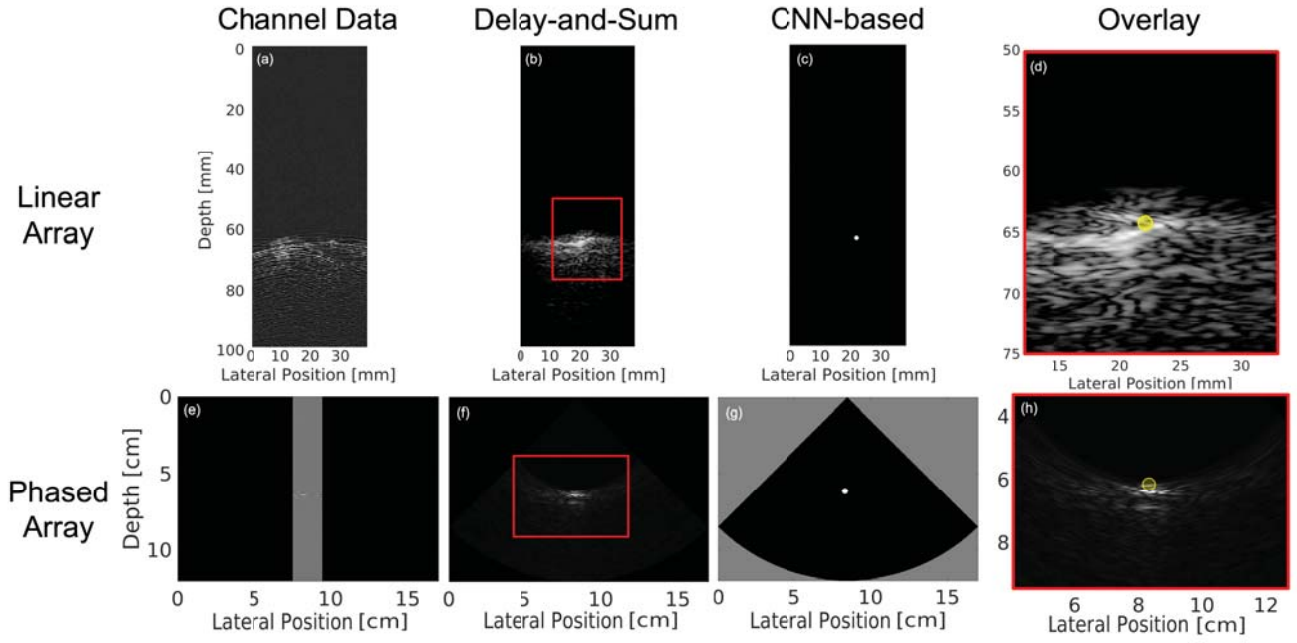


Fig. 2: (a,e) Channel data, (b,f) delay-and-sum (DAS) images, (c,g) CNN-based images, and (d,h) overlaid DAS and CNN images obtained with the (a)-(d) linear array and (e)-(h) phased array transducers. Both examples contain one source signal emanating from an optical fiber embedded in a cardiac catheter, which was imaged in the femoral vein of a swine. The DAS images contain a red box which indicates the region where the source is located. A zoomed-in version of this region is used to overlay the corresponding region from the CNN-based image, wherein the source location is shown in yellow.

guiding cardiac catheters, there are other types of catheterization and interventional procedures that have the potential to benefit from this approach.

## ACKNOWLEDGEMENTS

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

## REFERENCES

[1] Muyinatu A Lediju Bell and Joshua Shubert. Photoacoustic-based visual servoing of a needle tip. *Scientific Reports*, 8(1):15519, 2018.

[2] Minghua Xu and Lihong V Wang. Photoacoustic imaging in biomedicine. *Review of Scientific Instruments*, 77(4):041101, 2006.

[3] Paul Beard. Biomedical photoacoustic imaging. *Interface Focus*, page rsfs20110028, 2011.

[4] Derek Allman, Austin Reiter, and Muyinatu A Lediju Bell. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Transactions on Medical Imaging*, 37(6):1464–1477, 2018.

[5] Austin Reiter and Muyinatu A Lediju Bell. A machine learning approach to identifying point source locations in photoacoustic data. In *Proc. of SPIE*, volume 10064, pages 100643J–1, 2017.

[6] B. E. Treeby and B. T. Cox. k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave-fields. *J. Biomed. Opt.*, 15(2):021314, 2010.

[7] Derek Allman, Fabrizio Assis, Jonathan Chrispin, and Muyinatu A Lediju Bell. Deep neural networks to remove photoacoustic reflection artifacts in ex vivo and in vivo tissue. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2018.

[8] Derek Allman, Fabrizio Assis, Jonathan Chrispin, and Muyinatu A Lediju Bell. A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions. In *Proc. of SPIE*, 2019.

[9] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.