PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Deep learning the sound of light to guide surgeries

Muyinatu A. Lediju Bell

Muyinatu A. Lediju Bell, "Deep learning the sound of light to guide surgeries," Proc. SPIE 10868, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XVII, 108680G (26 February 2019); doi: 10.1117/12.2521315

SPIE.

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

Deep learning the sound of light to guide surgeries

Muyinatu A. Lediju Bell[a,b,c]

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

ABSTRACT

Photoacoustic imaging utilizes light and sound to make images by transmitting laser pulses that illuminate regions of interest, which subsequently absorb the light, causing thermal expansion and the generation of sound waves that are detected with conventional ultrasound transducers. The Photoacoustic and Ultrasonic Systems Engineering (PULSE) Lab is developing novel methods that use photoacoustic imaging to guide surgeries with the ultimate goal of eliminating surgical complications caused by injury to important structures like major blood vessels and nerves that are otherwise hidden from a surgeons immediate view. This paper summarizes our recent work to learn from the physics of sound propagation in tissue and develop acoustic beamforming algorithms that improve image quality, using state-of-the-art deep learning methods. These deep learning methods hold promise for robotic tracking tasks, visualization and visual servoing of surgical tool tips, and assessment of relative distances between the surgical tool and nearby critical structures (e.g., major blood vessels and nerves that if injured will cause severe complications, paralysis, or patient death). Impacted surgeries and procedures include neurosurgery, spinal fusion surgery, hysterectomies, and biopsies.

1. INTRODUCTION

Photoacoustic imaging is an emerging medical imaging modality that uses light and sound to make images.[1–3] Light is transmitted to the body, and structures with higher optical absorption than surrounding tissue (e.g., blood vessels) preferentially absorb this light. Optical absorption leads to thermal expansion and contraction, which generates ultrasound waves that can be detected with conventional ultrasound transducers. This emerging technique has great potential to guide surgeries by avoiding accidental injury to major blood vessels, which is one active area of research for the Photoacoustic and Ultrasonic Systems Engineering (PULSE) Lab.[4–8] We have also shown that this imaging technique can be used to target desired blood-rich regions, such as the cancellous core of pedicles during spinal fusion surgeries.[9]

One of the most outstanding challenges with photoacoustic images is the presence of artifacts that severely degrade image quality. Several groups have explored approaches to mitigate these artifacts, including wavelength-dependent techniques,[10] motion-based methods,[11,12] frequency-based methods,[13,14] techniques using singular value decomposition,[15] photoacoustic-guided focused ultrasound (PAFUSion),[15–17] and short-lag spatial coherence.[18–20] Limitations of these methods include minimal potential to remove artifacts caused by bright acoustic reflections, assumptions of identical acoustic reception pathways, reduced frame rates, and the lack of compensation for potential inter- and intrapatient variability.

Our group previously demonstrated that a deep learning approach can be trained with simulated data to detect photoacoustic point sources,[21–25] including photoacoustic signals originating from an optical fiber tip housed in a needle surrounded by water,[22–24] a needle surrounded by *ex vivo* tissue,[25] and a cardiac catheter located in an *in vivo* femoral vein.[25] 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.[23] This paper summarizes our results obtained across multiple deep neural network architectures, transducer receiver models, and simulation and experimental datasets that were not included during training.

Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XVII
edited by Anita Mahadevan-Jansen, Warren S. Grundfest, Proc. of SPIE Vol. 10868
108680G · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2521315

2. METHODS

The overall goal of the proposed approach is to learn the unique shape-to-depth relationship of point-like photoacoustic sources in order to provide a deep learning-based replacement to common photoacoustic image formation steps, as illustrated in Fig. 1 (top). We tested several convolutional neural networks (CNNs) to achieve this goal. For each network, we first trained CNNs with k-Wave[26] simulated data of acoustic wavefronts emanating from point-like sources. After this training step, CNNs that achieved greater than 90% source classification accuracy were transferred to real photoacoustic data. Initially our output was point source locations. We later trained networks to output both source and artifact locations as well as classifications of the detected wavefronts. These outputs are then displayed in an image format that we call CNN-based images, which show detected point source locations and location error as an image, as shown in Fig. 1 (bottom).

The following four network architectures were trained in our previous work:

- AlexNet network architecture[21,27]
- Faster R-CNN architecture composed of a deep fully convolutional network (i.e., the VGG16 network architecture[28] and a Region Proposal Network[29]) and a Fast R-CNN detector,[30] as illustrated in Fig. 1 (top),[22–24] which was used within the Caffe framework.[31] Faster R-CNN[29] was also used within the Detectron software[32] with the VGG-16 network architecture[25]
- Faster R-CNN[29] used within the Detectron software[32] with Resnet-50[33] (i.e., a residual network with 50 layers) replacing the VGG-16 architecture[25]
- Faster R-CNN[29] used within the Detectron software[32] with Resnet-101[33] (i.e., a residual network with 101 layers) replacing the VGG-16 architecture[25]

The following six types of data were tested in our previous work:

1. k-Wave simulated data from point targets[21,23,24]
2. experimental data from a phantom containing a cylindrical rubber rod to mimic a blood vessel[21]
3. experimental data from a phantom containing brachytherapy seeds[24]
4. experimental data from a needle containing an optical fiber inserted in a water bath[22,24]
5. experimental data from a needle containing an optical fiber inserted in *ex vivo* tissues from a chicken breast, bovine liver, steak, and whole chicken thigh containing bone[25]
6. experimental data from a cardiac catheter containing an optical fiber inserted in an *in vivo* porcine femoral vein[25]

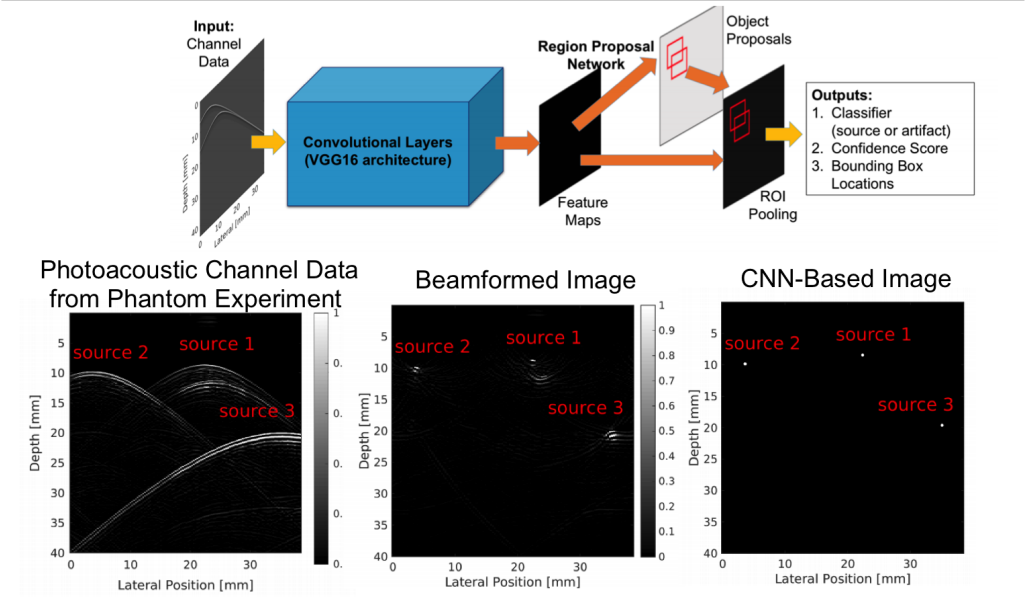

Figure 1: Example network architecture and example experimental images of channel data, beamformed data and CNN-based images

In some of these cases, two types of transducer models were implemented, with various noise levels, multiple sources in a single recording, and multiple medium sound speeds. In this paper, we compare the performance of each test set based on reported source classification, misclassification, and missed detection rates, as well as distance error of correct detections. Whenever possible, artifact classification, misclassification, and missed detection rates as well as precision, recall, and area under the curve were also reported, but these metrics are omitted from this paper. More details regarding our training and validation methods, training and validation test sets, and experimental imaging equipment are available in the papers cited above.

3. RESULTS

Table 1 summarizes many of the major results reported in our previous publications. Although many of our papers report results from both training and validation datasets, the results in Table 1 are reported for test datasets only, with the exception of the simulation results for the Resnet50 and Resnet101 architectures, which report validation dataset results. The number of images included in each grouping of datasets, as well as the source classification, misclassification, and missed detection rates are reported. Where available, the mean axial and lateral errors are additionally reported. The entries reported as not applicable (i.e., N/A) are absent because the AlexNet architecture does not classify objects, thus this network does not output classification results that can be used to calculate classification, misclassification, and missed detection rates. The blank entries (i.e., -), are absent because there was no ground truth to assess distance errors in *ex vivo* and *in vivo* datasets.

Overall, the classification rates ranged from 92-99.62% for simulated data, and the greatest classification performance with simulated data was achieved with the network architecture that included Resnet101. Similarly, for simulated data, the lowest misclassification rate (0.28%) was also achieved with Resnet101. Similar performance was achieved with the experimental water bath and phantom data when using the Faster R-CNN architecture with the plain VGG16 convolutional neural network. This success demonstrates two major breakthroughs for the field of deep learning applied to photoacoustic image formation. First, simulations of acoustic wave propagation can be used to successfully train deep neural networks. Second these networks transfer well to experimental data that were not included during training.

Table 1: Summary of Deep Learning Results Obtained with Multiple Networks and Datasets

	# of Images Tested	**Classification Rate**	**Misclassification Rate**	**Missed Detection Rate**	**Mean Axial Error (mm)**	**Mean Lateral Error (mm)**
AlexNet						
Simulated[21]	2,412	N/A	N/A	N/A	0.28	0.37
Vessel Phantom[21]	1	N/A	N/A	N/A	2	2
Faster R-CNN - VGG16						
Simulated - Continuous, 1 source, multiple noise levels[23,24]	3,998	97%	15%	<0.7%	0.12	0.20
Simulated - Discrete, 1 source, multiple noise levels[23,24]	3,998	92%	11%	<0.7%	0.12	0.17
Water bath - Discrete[24]	17	100%	0%	0%	0.24	0.27
Brachytherapy Phantom (only results in training range)[24]	15	97%	3%	3%	<0.38	<0.38
Ex Vivo Tissue[25]	82	84-100%	0-23.5%	0-16%	-	-
In Vivo[25]	279	14.5%	<2%	85.5%	-	-
Faster R-CNN - Resnet50						
Simulated (validation dataset)	3,998	97.93%	0.38%	2.08%	0.101	0.097
Ex Vivo Tissue[25]	82	71-100%	0%	0-29%	-	-
In Vivo[25]	279	83%	<2%	<15.9%	-	-
Faster R-CNN - Resnet101						
Simulated (validation dataset)	3,998	99.62%	0.28%	0.9%	0.103	0.088
Ex Vivo Tissue[25]	82	65-100%	0%	0-35%	-	-
In Vivo[25]	279	89%	<2%	<15.9%	-	-

After achieving success with simulated and experimental phantom data, these networks were then tested with data from *ex vivo* and *in vivo* tissue. The initial network with the VGG16 architecture did not perform as well on these *ex vivo* and *in vivo* datasets, as reported in Table 1. However, performance increased when the VGG16 architecture was replaced with a residual network architecture. In particular, classification rates with the residual networks ranged from 83-89% for the *in vivo* data and the misclassification rates were <2%, while the missed detection rates were <15.9%. The majority of the *ex vivo* datasets experienced similar performance, with the exception of steak tissue,[25] which produced significantly more artifacts than the other *ex vivo* tissues that were tested. These artifacts and results from the *ex vivo* steak images were solely responsible for the lower classification rates of 65-71% that are reported in Table 1 for the *ex vivo* tissue.[25]

With the exception of AlexNet applied to experimental data, the remaining network architectures and datasets produced submillimeter axial and lateral target location errors. We relate these location errors to the resolution of our deep learning-based imaging system, where resolution is defined as an integer multiple of (i.e., typically 2 or 3 times) the location errors.[24,25] Based on this relationship, the CNN-based images have promising potential to provide better resolution than traditional beamformed images, particularly as image depth increases. We have shown this for image depths as deep as 10 cm with experimental target depths of approximately 6-8 cm.[25]

4. DISCUSSION

Our recent series of publications demonstrate the power and potential of using deep learning to fundamentally reconsider traditional approaches to photoacoustic image formation. Early success with deep learning alternatives have proven to be promising with classification rates that exceed 80% in most cases from a variety of datasets. These datasets span simulated and more importantly experimental data (including *in vivo* data from a pig catheterization procedure). The results of the point source location errors also demonstrate that this fundamentally new approach has the potential to rival spatial resolution measurements from traditional photoacoustic image formation procedures, particularly as depth increases. Our trained code and a few of our datasets are freely available to enable future comparisons.[34]

In addition to improving photoacoustic image quality with deep learning approaches, we are also pioneering similar deep learning concepts to improve ultrasound image quality.[35,36] These ultrasound-based deep learning approaches will inherently benefit photoacoustic imaging because ultrasound images are typically needed to provide anatomical context for interventional photoacoustic images. Current ultrasound imaging methods suffer from similar challenges with acoustic clutter and poor image quality[37] that can potentially be overcome with similar deep learning alternatives to ultrasound image formation. The proposed method applied to photoacoustic images can then be overlaid on traditional ultrasound and/or photoacoustic images.

5. CONCLUSION

This paper summarizes the most recent results of PULSE Lab's research efforts to apply deep learning to detect photoacoustic point sources and mitigate problematic photoacoustic artifacts, with the added advantage of improving photoacoustic image resolution. Our CNN-based images can be displayed independently or overlaid on traditional beamformed images. Many possibilities lie ahead to integrate deep learning with photoacoustic image formation for interventional guidance of surgical procedures.

ACKNOWLEDGEMENTS

This work is partially supported by NIH Trailblazer Award R21-EB025621 and NSF CAREER Award 1751522. Thanks to Derek Allman for his assistance with completing Table 1.

REFERENCES

[1] Xu, M. and Wang, L. V., "Photoacoustic imaging in biomedicine," *Review of Scientific Instruments* **77**(4), 041101 (2006).

[2] Bouchard, R., Sahin, O., and Emelianov, S., "Ultrasound-guided photoacoustic imaging: current state and future development," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **61**(3), 450–466 (2014).

[3] Beard, P., "Biomedical photoacoustic imaging," *Interface focus* **1**(4), 602–631 (2011).
[4] Eddins, B. and Bell, M. A. L., "Design of a multifiber light delivery system for photoacoustic-guided surgery," *Journal of Biomedical Optics* **22**(4), 041011–041011 (2017).
[5] Gandhi, N., Kim, S., Kazanzides, P., and Bell, M. A. L., "Accuracy of a novel photoacoustic-based approach to surgical guidance performed with and without a da vinci robot," in [*Proc. of SPIE*], **10064**, 100642V–1 (2017).
[6] Gandhi, N., Allard, M., Kim, S., Kazanzides, P., and Bell, M. A. L., "Photoacoustic-based approach to surgical guidance performed with and without a da vinci robot," *Journal of Biomedical Optics* **22**(12) (2017).
[7] Allard, M., Shubert, J., and Bell, M. A. L., "Feasibility of photoacoustic guided teleoperated hysterectomies," *Journal of Medical Imaging: Special Issue on Image-Guided Procedures, Robotic Interventions, and Modeling* **5**(2), 021213 (2018).
[8] Bell, M. A. L. and Shubert, J., "Photoacoustic-based visual servoing of a needle tip," *Scientific Reports* **8**(1), 15519 (2018).
[9] Shubert, J. and Bell, M. A. L., "Photoacoustic imaging of a human vertebra: implications for guiding spinal fusion surgeries," *Physics in Medicine and Biology* (2018).
[10] Nguyen, H. N. Y., Hussain, A., and Steenbergen, W., "Reflection artifact identification in photoacoustic imaging using multi-wavelength excitation," *Biomedical Optics Express* **9**(10), 4613–4630 (2018).
[11] Jaeger, M., Harris-Birtill, D. C., Gertsch-Grover, A. G., OFlynn, E., and Bamber, J. C., "Deformation-compensated averaging for clutter reduction in epiphotoacoustic imaging in vivo," *Journal of Biomedical Optics* **17**(6), 066007 (2012).
[12] Jaeger, M., Bamber, J. C., and Frenz, M., "Clutter elimination for deep clinical optoacoustic imaging using localised vibration tagging (lovit)," *Photoacoustics* **1**(2), 19–29 (2013).
[13] Bell, M. A. L., Song, D. Y., and Boctor, E. M., "Coherence-based photoacoustic imaging of brachytherapy seeds implanted in a canine prostate," in [*SPIE Medical Imaging*], 90400Q–90400Q, International Society for Optics and Photonics (2014).
[14] Nan, H., Chou, T.-C., and Arbabian, A., "Segmentation and artifact removal in microwave-induced thermoacoustic imaging," in [*Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*], 4747–4750, IEEE (2014).
[15] Singh, M. K. A., Jaeger, M., Frenz, M., and Steenbergen, W., "In vivo demonstration of reflection artifact reduction in photoacoustic imaging using synthetic aperture photoacoustic-guided focused ultrasound (PAFUSion)," *Biomedical Optics Express* **7**(8), 2955–2972 (2016).
[16] Singh, M. K. A. and Steenbergen, W., "Photoacoustic-guided focused ultrasound (PAFUSion) for identifying reflection artifacts in photoacoustic imaging," *Photoacoustics* **3**(4), 123–131 (2015).
[17] Schwab, H.-M., Beckmann, M. F., and Schmitz, G., "Photoacoustic clutter reduction by inversion of a linear scatter model using plane wave ultrasound measurements," *Biomedical Optics Express* **7**(4), 1468–1478 (2016).
[18] Bell, M. A. L., Kuo, N., Song, D. Y., and Boctor, E. M., "Short-lag spatial coherence beamforming of photoacoustic images for enhanced visualization of prostate brachytherapy seeds," *Biomedical Optics Express* **4**(10), 1964–1977 (2013).
[19] Pourebrahimi, B., Yoon, S., Dopsa, D., and Kolios, M. C., "Improving the quality of photoacoustic images using the short-lag spatial coherence imaging technique," in [*SPIE BiOS*], 85813Y–85813Y, International Society for Optics and Photonics (2013).
[20] Alles, E. J., Jaeger, M., and Bamber, J. C., "Photoacoustic clutter reduction using short-lag spatial coherence weighted imaging," in [*Ultrasonics Symposium (IUS), 2014 IEEE International*], 41–44, IEEE (2014).
[21] 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).
[22] 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).

[23] 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).
[24] 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).
[25] 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).
[26] 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).
[27] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in Neural Information Processing Systems 25*], Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., eds., 1097–1105, Curran Associates, Inc. (2012).
[28] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations (ICLR), 2015* (2014).
[29] 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).
[30] Girshick, R., "Fast R-CNN," in [*Proceedings of the IEEE International Conference on Computer Vision*], 1440–1448 (2015).
[31] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," in [*Proceedings of the 22nd ACM International Conference on Multimedia*], 675–678, ACM (2014).
[32] Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., and He, K., "Detectron." https://github.com/facebookresearch/detectron (2018).
[33] 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).
[34] Allman, D., Reiter, A., and Bell, M. A. L., "Photoacoustic source detection and reflection artifact deep learning dataset." http://dx.doi.org/10.21227/H2ZD39 (2018).
[35] Nair, A. A., Tran, T. D., Reiter, A., and Bell, M. A. L., "A deep learning based alternative to beamforming ultrasound images," in [*2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 3359–3363, IEEE (2018).
[36] Nair, A. A., Gubbi, M. R., Tran, T. D., Reiter, A., and Bell, M. A. L., "A fully convolutional neural network for beamforming ultrasound images," in [*2018 IEEE International Ultrasonics Symposium (IUS)*], 1–4, IEEE (2018).
[37] 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).