Exploring the effects of transducer models when training convolutional neural networks to eliminate reflection artifacts in experimental photoacoustic images

Derek Allman[a], Austin Reiter[b], and Muyinatu Bell[a,c]

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

ABSTRACT

We previously proposed a method of removing reflection artifacts in photoacoustic images that uses deep learning. Our approach generally relies on using simulated photoacoustic channel data to train a convolutional neural network (CNN) that is capable of distinguishing sources from artifacts based on unique differences in their spatial impulse responses (manifested as depth-based differences in wavefront shapes). In this paper, we directly compare a CNN trained with our previous continuous transducer model to a CNN trained with an updated discrete acoustic receiver model that more closely matches an experimental ultrasound transducer. These two CNNs were trained with simulated data and tested on experimental data. The CNN trained using the continuous receiver model correctly classified 100% of sources and 70.3% of artifacts in the experimental data. In contrast, the CNN trained using the discrete receiver model correctly classified 100% of sources and 89.7% of artifacts in the experimental images. The 19.4% increase in artifact classification accuracy indicates that an acoustic receiver model that closely mimics the experimental transducer plays an important role in improving the classification of artifacts in experimental photoacoustic data. Results are promising for developing a method to display CNN-based images that remove artifacts in addition to only displaying network-identified sources as previously proposed.

1. INTRODUCTION

To implement photoacoustic imaging, pulsed laser light is delivered to a target that absorbs the light, undergoes thermal expansion, and converts the absorbed energy to a pressure wave. This mechanical pressure can be imaged with an ultrasound transducer and create a photoacoustic image.[1,2] Promising applications of photoacoustic imaging include visualization of surgical tools[3,4] and imaging blood vessels in the body.[5]

One limitation of photoacoustic imaging is notorious reflection artifacts.[3,6] Reflection artifacts can be caused by a source signal that travels to a hyperechoic object, which generates an additional pressure wave due to the acoustic impedance mismatch. Traditional beamforming methods use a time-of-flight model to reconstruct acoustic signals sensed at the transducer, which causes reflection artifacts to be incorrectly mapped to locations that are deeper than their point of origin in the photoacoustic image.

Our previous work[7,8] explored a deep learning method to eliminate reflection artifacts. We trained a convolutional neural network (CNN) with simulated photoacoustic channel data to both detect and classify photoacoustic sources and artifacts. The network correctly distinguished between sources and artifacts in simulated channel data. Transfer learning was then implemented to test the trained network on experimental data, and the network correctly classified 100% of sources in the experimental images, but only 54.1% of artifacts were correctly classified.[8]

When training a neural network with synthetic or simulated data, it is expected that performance will increase when the simulated data closely resembles the real environment. For example, when applying CNNs to semantically segment urban street scene images, Veeravasarapu et al.[9] observed performance increases when the accuracy of rendered scenes approached that of real world images. Similarly, generative adversarial networks (GANs)[10] were recently developed to create candidate synthetic images that are graded by a component known

as a discriminator. The discriminator determines if the generated image looks sufficiently real compared to a database of real images. Once fine tuned, a GAN is capable of creating vast amounts of synthetic images which are real in appearance. In general, the use of simulated data to train neural networks is gaining popularity in cases when labeled training data is scarce or expensive to acquire, which is particularly true with the multiple reflection artifact locations in photoacoustic imaging.

The purpose of the work presented in this paper is to assess the extent to which our simulated data needs to match our real data in terms of noise levels, signal intensities, and acoustic receiver properties when implementing transfer learning to identify and remove artifacts in photoacoustic images. We initially simulated a continuous transducer with zero kerf and a range of sampling frequencies at only one noise level and a single photoacoustic amplitude.[7,8] We are now expanding our simulations to include training data over multiple noise levels and signal intensities, using an acoustic receiver model that more closely matches experimental ultrasound transducers. We directly compare two networks. One network was trained with multiple noise levels, signal intensities, and the same continuous receiver model used previously. The second network was trained with multiple noise levels, signal intensities, and a discrete receiver model with a kerf, element spacing, and sampling frequency that matches the specifications of the ultrasound transducer used when acquiring our experimental data.

2. METHODS

We generated two datasets, one corresponding to the continuous receiver model and another corresponding to the discrete receiver model. Each transducer model was simulated in k-Wave.[11] A schematic diagram of the two receiver models is shown in Fig. 1. The continuous transducer has a kerf of 0, an element width of 0.1 mm, and a total of 350 transducer elements, while the discrete receiver model has a kerf of 0.06 mm, an element width of 0.24 mm, and a total of 128 elements (same specifications as our Alpinion L3-8 linear array transducer). In addition, the continuous receiver sampled the photoacoustic response with 48 - 54.6 MHz sampling frequency depending on the speed of sound of the medium (which is the default setting for the k-Wave simulation), while the discrete transducer used a fixed 40 MHz sampling frequency.

Each dataset included 19,992 photoacoustic channel data images, each containing a 0.1 mm photoacoustic source and one reflection artifact. Photoacoustic sources were simulated using the parameters in Table 1. Reflection artifacts were generated using our previously described technique,[8] where source wavefields were shifted deeper into the image according to the Euclidean distance, Δ, between the source and the reflector, as defined by the equation:

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

(a)   (b)
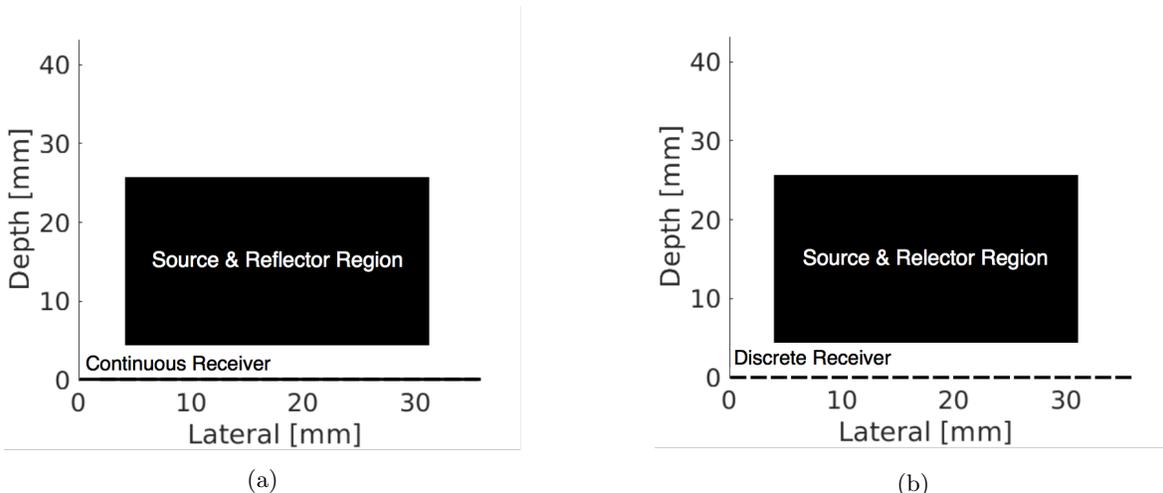
Figure 1: Schematic diagrams of the (a) continuous and (b) discrete receiver models used for simulating photoacoustic reception. Note that the discrete receiver model is not drawn to scale in this diagram (e.g., there are actually 128 transducer elements and the kerf is smaller than the kerf is shown here.)

Table 1: Range and Increment Size of Simulation Variables for Discrete and Continuous Receiver Models

Parameter	Min	Max	Increment
Number of Sources	1	10	1
Depth Position (mm)	5	25	0.25
Lateral Position (mm)	5	30	0.25
Channel SNR (dB)	-5	2	random
Signal Intensity (multiplier)	0.75	1.1	random
Speed of Sound (m/s)	1440	1640	6

Continuous   **Discrete**

(a)   (b)   (c)   (d)
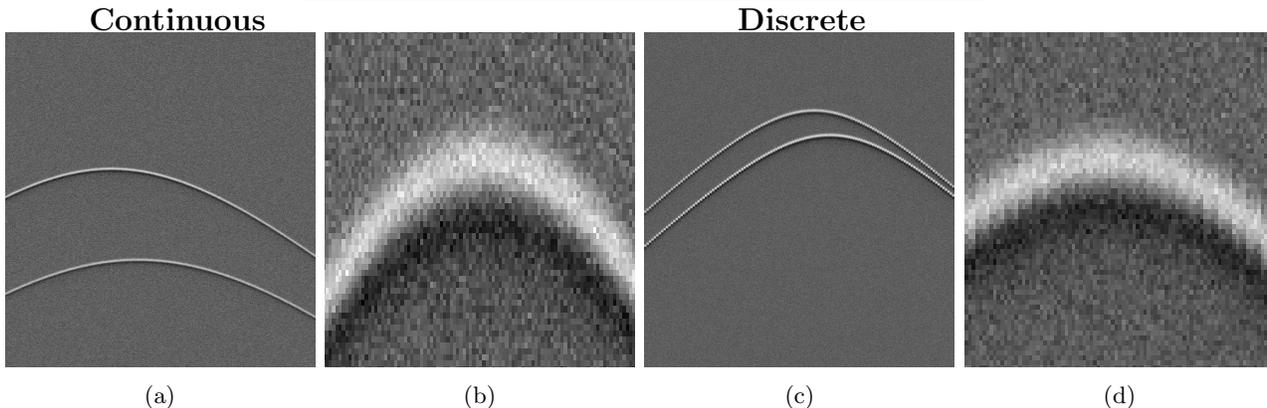
Figure 2: Examples of channel data generated with the (a,b) continuous and (c,d) discrete acoustic receiver models. The zoomed views in (b) and (d) show portions of the wavefronts that appear in (a) and (c), respectively. Note that the wavefront in the zoomed version from the discrete receiver model contains streaks and other subtle differences that are not present in the continuous case.

where (x_s, z_s) is the 2D spatial location of the source and (x_r, z_r) is the 2D spatial location of the reflector. Fig. 2 shows example images simulated with each transducer model.

One network was trained for each dataset using the Faster R-CNN algorithm,[12] VGG16 CNN network architecture.[13] The network was trained to detect and classify the peaks of the incoming acoustic waves as either sources or artifacts for 100,000 iterations. For each of dataset, 80% of the images were used for training and the remaining 20% of the images were saved for testing.

For each image, the Faster R-CNN algorithm outputs a list of object detections for each corresponding class, source or artifact, along with the object location in terms of bounding-box pixel coordinated as well as a confidence score (between 0 and 1).

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

The optimal score for each class and each network 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 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.

To implement transfer learning, experimental images of a needle submerged in water where utilized. The needle had a hollow core and a 1 mm core diameter optical fiber was inserted into the needle. One end of the optical fiber coincided with the tip of the needle. The needle was placed in the imaging plane between the

transducer and a sheet of acrylic and the entire apparatus was submerged into a waterbath. The other end of the optical fiber was coupled to a Quantel (Bozeman, MT) Brilliant laser. The laser light from the fiber tip creates a photoacoustic signal in the water which propagates in all directions. This signal travels both directly to the transducer, creating the source signal, and to the acrylic which reflects the signal to the transducer, creating the reflection artifact. Seventeen channel data images were captured, each after changing the location of the transducer while keeping the laser and acrylic spacing fixed. The mean channel SNR of the experimental data was measured as 0.1dB and the artifacts were labeled by hand after observing the B-mode image. This same experimental dataset was used for testing in our previous work.[8]

We evaluated classification, misclassification, and missed detection rates for this experimental dataset. Results were compared to the classification, misclassification, and missed detection rates obtained when the same networks were applied to the simulated data that was saved for testing only.

3. RESULTS

The classification results for the two CNNs applied to both simulated and experimental test data are shown in Fig. 3. The CNN trained with the continuous receiver model is labeled as "Continuous", while the CNN trained with the discrete receiver model is labeled as "Discrete" in Fig. 3. The experimental dataset contained a total of 17 true sources and 34 reflection artifacts across the 17 channel data images.

In the simulated case, when transitioning from the continuous to discrete receiver, source classification fell from 97.1% to 91.6% while artifact misclassification rose from 3.82% to 12.6%. This indicates an overall decrease in the network's ability to distinguish a source from an artifact, as the network classifies true sources less often and misclassifies artifacts as sources more often when the discrete receiver is adopted. However, artifact classification rose from 86.2% to 93.2% and source misclassification fell from 14.9% to 11.0% with the discrete receiver model applied to the simulated data. Thus, contrary to the network's decrease in performance with respect to source detections, the network trained with the discrete transducer model appears to be a better network when classifying artifacts, as this network classifies artifacts correctly more often in addition to misclassifying sources as artifacts less often.

For the experimental dataset, the two CNNs classified all source in the images correctly (100% source classification rate). The discrete receiver network classified more artifacts correctly (89.7%) when compared to the
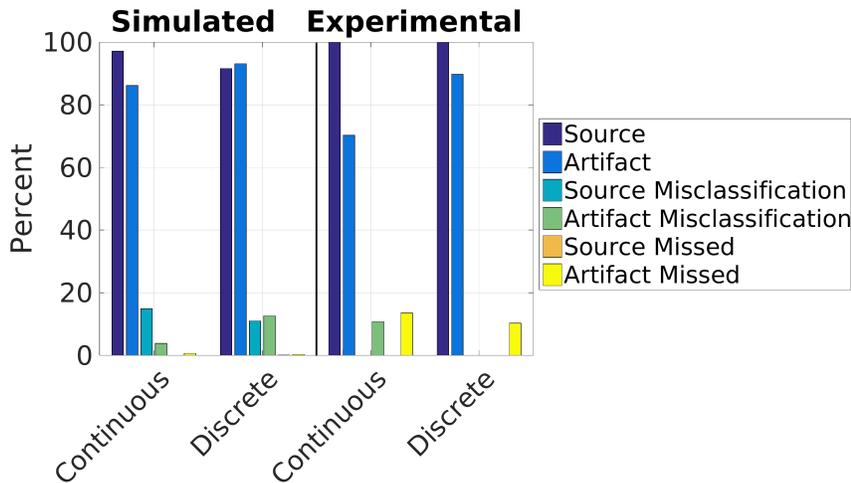
Figure 3: Classification results for networks trained with the continuous and discrete acoustic receiver models applied to both simulated and experimental data. The dark and medium blue bars show the accuracy of source and artifact detections, respectively. The light blue and green bars show the misclassification rate for sources and artifacts, respectively. The dark and light yellow bars show the missed detection rate for sources and artifacts, respectively.
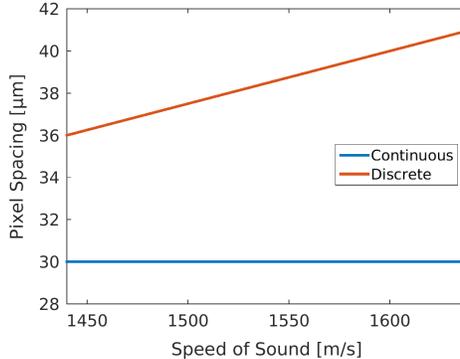
Figure 4: Plot of pixel spacing in the depth dimension (defined as the speed of sound divided by the sampling frequency) of the images created with the continuous and discrete acoustic receiver models.

continuous receiver network (70.3%), which follows the trend observed for artifact classification in the simulated data. In both cases (simulated and experimental), the network trained with the discrete receiver model performs better when classifying artifacts. We also note that more artifacts are missed when the trained network is transferred to experimental data, likely due to the presence of additional types of artifacts that were not included during training.

Fig. 4 shows a plot of pixel spacing in the depth dimension of the image (defined as the speed of sound divided by the sampling frequency) for the continuous and discrete receiver models. Note that regardless of the speed of sound in the medium, an object at a specific depth in the image will occur at the same pixel depth in the channel data for the continuous receiver, which has a sampling frequency that depends on the speed of sound (k-Wave default). This does not reflect reality. In comparison, for the discrete receiver with a fixed sampling frequency, an object at a specific depth will have a pixel depth in the channel data that is dependent on the speed of sound, which is more realistic. Fig. 4 is used in Section 4 to describe possible reasons for the initially unexpected decrease in source classification rate when the discrete receiver model was applied to simulated data.

4. DISCUSSION

The networks tested in this paper are improved versions of the networks trained and tested in our previous work.[8] Generally, when transferring networks trained in simulation to operate on experimental photoacoustic data, the artifact classification rates improved from 54.1% in our previous work[8] to 70.3% and 89.7% for the new continuous and discrete receiver models, respectively, as described in Section 2. These improvements can be attributed to including a range of noise levels and signal intensities in the simulation training data as well as more closely modeling the experimental transducer.

In the simulated data domain, the decrease in source classification performance when using the discrete versus the continuous receiver model (see Fig. 3) is likely due to the change from a range of sampling frequencies with the continuous receiver to a fixed sampling frequency with the discrete receiver. Fig. 4 indicates that for the continuous, varied sampling frequency receiver there is a constant relationship between an object's actual depth (which is related to wavefront shape, considering that this shape is dependent on the spatial impulse response) and its depth in the simulated received channel data, despite changes in the speed of sound. In comparison, for the discrete, fixed sampling frequency receiver, an object at a given depth can have a range of pixel depths depending on the speed of sound of the medium, which is more realistic and also more similar to the appearance of artifacts which already presented themselves to the network with a range of pixel depths. Thus, with the continuous receiver applied to simulated data, the constant pixel spacing that is independent of sound speed enables more certainty when discriminating a source from an artifact. The constant relationship shown in Fig. 4 likely does not affect experimental data because we expect a fixed speed of sound in the water medium used in the experiments. Therefore, in the experimental data domain, we attribute the increase in artifact classification accuracy with the discrete receiver model to having the number of receiver elements, their width, and their spacing more closely resembling these properties of the experimental transducer.

In the experimental data domain, the CNN trained with the discrete receiver outperforms the CNN trained with the continuous receiver, particularly when comparing artifact classification, artifact misclassification, and the number of artifacts missed. These results agree with the expectation that knowledge gained during training with simulated data will be better suited for transferring to real scenarios when the simulated domain is more similar to the experimental domain. Our future work will incorporate experimental channel data during training to assess how this additional change will affect the performance. In addition, while we previously used the CNN output to display only sources[8] and therefore, we were not too concerned with artifact classification, these new results support the exploration of deep learning approaches that develop CNN-based images to identify and remove artifacts, in addition to our previously proposed CNN-based images that remove artifacts by only displaying identified source locations.

5. CONCLUSION

This work is the first to directly compare the performance of continuous and discrete receiver models when applying deep learning to identify reflection artifacts in simulated and experimental photoacoustic data. The training data for each receiver model included multiple noise levels, signal intensities, and sound speeds in one network. The network trained with the discrete receiver model outperformed that trained with the continuous receiver model when applied to experimental data, particularly when identifying artifacts. Results are promising for developing a method to display CNN-based images that remove artifacts in addition to only displaying network-identified sources as previously proposed.

REFERENCES

[1] Beard, P., "Biomedical photoacoustic imaging," *Interface focus* **1**(4), 602–631 (2011).

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

[3] Su, J., Karpiouk, A., Wang, B., and Emelianov, S., "Photoacoustic imaging of clinical metal needles in tissue," *Journal of biomedical optics* **15**(2), 021309–021309 (2010).

[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 (2017).

[5] Kolkman, R. G., Steenbergen, W., and van Leeuwen, T. G., "In vivo photoacoustic imaging of blood vessels with a pulsed laser diode," *Lasers in medical science* **21**(3), 134–139 (2006).

[6] Lediju Bell, M. A., Kuo, N. P., Song, D. Y., Kang, J. U., and Boctor, E. M., "In vivo visualization of prostate brachytherapy seeds with photoacoustic imaging," *Journal of Biomedical Optics* **19**(12), 126011 (2014).

[7] Reiter, A. and Bell, M. A. L., "A machine learning approach to identifying point source locations in photoacoustic data," in [*Proc. of SPIE Vol*], **10064**, 100643J–1 (2017).

[8] Allman, D., Reiter, A., and Bell, M., "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).

[9] Veeravasarapu, V., Rothkopf, C., and Ramesh, V., "Model-driven simulations for deep convolutional neural networks," *arXiv preprint arXiv:1605.09582* (2016).

[10] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Advances in neural information processing systems*], 2672–2680 (2014).

[11] 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).

[12] 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).

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