Robust Photoacoustic Beamforming Using Dense Convolutional Neural Networks

Emran Mohammad Abu Anas[1]([✉]), Haichong K. Zhang[1], Chloé Audigier[2], and Emad M. Boctor[1,2]

[1] Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
eanas1@jhmi.edu
[2] Radiology and Radiological Science, Johns Hopkins University, Baltimore, MD, USA

Abstract. Photoacoustic (PA) is a promising technology for imaging of endogenous tissue chromophores and exogenous contrast agents in a wide range of clinical applications. The imaging technique is based on excitation of a tissue sample using short light pulse, followed by acquisition of the resultant acoustic signal using an ultrasound (US) transducer. To reconstruct an image of the tissue from the received US signals, the most common approach is to use the delay-and-sum (DAS) beamforming technique that assumes a wave propagation with a constant speed of sound. Unfortunately, such assumption often leads to artifacts such as sidelobes and tissue aberration; in addition, the image resolution is degraded. With an aim to improve the PA image reconstruction, in this work, we propose a deep convolutional neural networks-based beamforming approach that uses a set of densely connected convolutional layers with dilated convolution at higher layers. To train the network, we use simulated images with various sizes and contrasts of target objects, and subsequently simulating the PA effect to obtain the raw US signals at an US transducer. We test the network on an independent set of 1,500 simulated images and we achieve a mean peak-to-signal-ratio of $38.7\,\mathrm{dB}$ between the estimated and reference images. In addition, a comparison of our approach with the DAS beamforming technique indicates a statistical significant improvement of the proposed technique.

Keywords: Photoacoustic · Beamforming · Delay-and-sum
Convolutional neural networks · Dense convolution
Dilated convolution

1 Introduction

Photoacoustic (PA) is considered as a hybrid imaging modality that combines optical and ultrasound (US) imaging techniques. The underlying physics of this

imaging technology is based on the PA effect that refers to the phenomenon of generation of acoustic waves following a short light pulse absorption in a soft-tissue sample. To exploit the PA effect and enable imaging of that soft-tissue, a light source (laser or light emitting diode) is employed to excite the soft-tissue, and simultaneously an US transducer is used to collect the instantaneously generated acoustic signal. In contrast to the acoustic properties-based pure US imaging technique, the PA imaging modality provides functional information (e.g., hemoglobin in blood and melanin in skin) of the anatomy. Based on this fact, the key applications of PA imaging have been found in detection of ischemic stroke, breast cancer or skin melanomas [3,7]. In addition to tissue chromophores, the PA technique has shown its ability to image exogenous contrast agents in a number of clinical applications including molecular imaging and prostate cancer detection [1,18].

The most common approach to reconstruct a PA image from the received US signal (channel data) is the delay-and-sum (DAS) beamforming technique due to its simple implementation and real-time capability. In short, the output of the DAS method is obtained by averaging the weighted and delayed versions of the received US signals. The delay calculation is based on an assumption of an US wave propagation with a constant speed of sound (SoS), therefore, it compromises the image quality of a DAS beamformer [5]. To improve the beam-forming with PA imaging, a significant number of works have been reported using, for example, minimum variance [11], coherence factor [13], short-lag spatial coherence beamforming [4], adaptive beamforming [15] and double-stage delay-multiply-and-sum beamforming [12]. Though these approaches have shown their potential to improve the beamforming, they are less robust to tackle the SoS variation in different applications.

In the recent years, deep learning based approaches have demonstrated their promising performance compared to the previous state-of-the-art image processing approaches in almost all areas of computer vision. In addition to vision recognition, deep neural techniques have been successfully applied for beamforming [2,9,10,14] of US signal. Luchies et al. [9,10] presented deep neural networks for US beamforming from the raw channel data. However, their proposed network is based on fully connected layers that are prone to overfit the network parameters. Nair et al. [14] recently proposed a deep convolutional neural networks (CNN)-based image transformation approach to map the channel data to US images. In fact, their output US image is a binary segmentation map instead of an intensity image, therefore, it does not preserve the relative contrast among the target objects that is considered quite important in functional PA imaging.

In this work, we propose a deep CNN-based approach to reconstruct a PA image from the raw US signals. Unlike the techniques in [9,10], we use fully convolutional networks for the beamforming of the US signals that reduces the problem of overfitting the network parameters. In addition, our proposed network maps the channel data to an intensity image that keeps the relative contrast among the target objects. The network consists of a set of densely connected convolutional layers [6] that have shown their effectiveness to eliminate

the gradient vanishing problem during training. Furthermore, we exploit dilated convolution [17] at higher layers in our architecture to allow feature extraction without loss of resolution. The training of the network is based on simulation experiments that consist of simulated target objects in different SoS environments. Such a variation in SoS during training makes the proposed network less sensitive to SoS changes when mapping the channel data to a PA image.

2   Methods

Figure 1(a) shows the architecture of our proposed deep CNN that maps an input channel data to an output PA image. Note that the channel data refers to the pre-beamformed RF data in this whole work. The sizes of the input and output images of the network are 384×128. The network consists of five dense blocks representing convolution at five different scales. In each dense block, there are two densely connected convolutional layers with 16 feature maps in each layer. The size of all of the convolutional kernels is 9×9 and each convolution is followed by rectified linear unit (ReLU). The key principle of dense convolution is using all of the previous features at its input, therefore, the features are propagated more effectively, subsequently, the vanishing gradient problem is eliminated [6]. In addition to dense convolution, we use dilated convolution [17] at higher layers of our architecture to overcome the problem of losing resolution at those layers. We set the dilation factors for the dense block 1 to 5 as 1, 2, 4, 8 and 16, respectively (Fig. 1(a)). The dilated convolution is a special convolution that
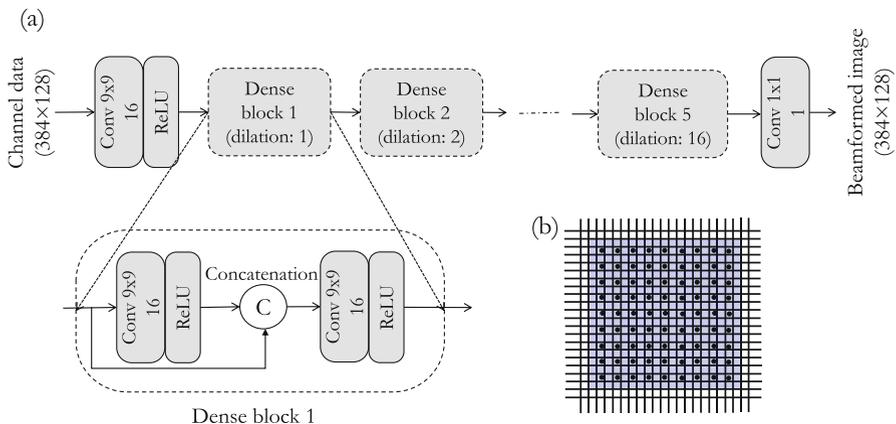
Fig. 1. The proposed beamforming approach. (a) The neural network architecture to map the channel data to a PA image. The network consists of five dense blocks, where each dense block consists of two densely connected convolutional layers followed by ReLU. The difference among five dense blocks lies in the amount of dilation factor. (b) Effect of dilation on the effective receptive field size. A dilated convolution of a kernel size of 9×9 with a dilation factor of 2 indicates an effective receptive field size of 19×19.

allows the convolution operation without decreasing the resolution of the feature maps but using the same number of convolutional kernel parameters. Figure 1(b) shows an example of a dilated convolution for a kernel size of 9×9 with a dilation factor of 2 that represents an effective receptive field size of 19×19. A successive increase in the dilation factor across layer 1 to 5, therefore, indicates a successively greater effective receptive field size. At the end of the network, we perform 1×1 convolution to predict an output image from the generated feature maps. The loss function of the network consists of the mean square losses between the predicted and target images.

3   Experiments and Training

3.1   Experiments and Materials

We perform simulation experiments to train, validate and test the proposed network. Each experiment consists of a number of simulated 2D target objects with different sizes and contrasts. In this work, we consider only circular target objects because their shapes are highly similar to those of the blood vessels in 2D planes, where most of the PA applications have been reported. In each simulation, we randomly choose the total number (between 1 and 6 inclusive) of target objects. In addition, the SoS of the background as well as of the targets is set as constant for each experiment and it is randomly chosen in the range of 1450–1550 m/s. Each target is modeled by a Gaussian function, where position of the target and its peak intensity (corresponding to contrast) are randomly chosen. The size of the target is controlled by the standard deviation of the Gaussian function, and it is also chosen randomly within a range of 0.2–1.0 mm. We have performed a total of 5,000 simulations to obtain the simulated images. For each image, we generate the channel data considering a linear transducer with 128 elements at the top of the simulated image (simulating PA effect) using the k-Wave simulation toolbox [16]. In addition, we have introduced white noise with Gaussian distribution on the channel data to allow the proposed network to be robust against the noise. The variance of the noise is randomly chosen as a number that is always less than the power of the signal in each experiment. Figure 2 shows an example of simulated image and corresponding channel data with 4 target objects with different sizes and contrasts.

We divide all of our images into three groups to constitute the training, validation and test sets. The images are distributed independently as 60% vs 10% vs 30% for the training, validation and test sets, respectively. Therefore, the total number of training, validation and test images are 3,000, 500 and 1,500, respectively, in this work.

3.2   Training and Validation

We use the TensorFlow library (Google, Mountain View, CA) with Adam [8] optimization technique to train the proposed network based on the training set.
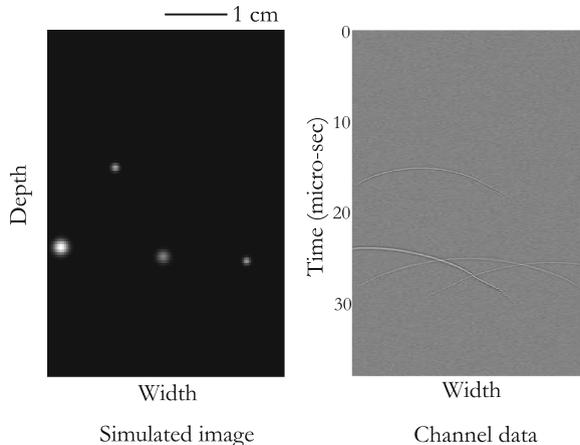
Fig. 2. An example of simulated PA image with 4 targets (left figure). We can notice the variation of sizes and contrasts among the targets. From the simulated image, we use the k-Wave simulation toolbox [16] to obtain the channel data (right figure).

A total of 8,000 epochs is used to optimize the network parameters in our GPU (NVIDIA GeForce GTX 1080 Ti with 11 GB RAM) with a mini-batch size of 16. The initial learning rate is set to 10^{-3} and there is an exponential decay of the learning rate after each successive 2,000 epochs with a decay factor of 0.1.

While the training set is used to optimize the network parameters, the validation set in our work is used to fix the hyper-parameters of our network including the size of the convolutional kernel (9×9), the number of convolutional layers (2) in each dense block, the number of feature maps (16) in each dense convolution and the initial learning rate (10^{-3}) in the Adam optimization.

4   Evaluation and Results

4.1   Evaluation

To evaluate the proposed beamforming approach using the test set, we use the peak-signal-to-noise-ratio (PSNR) that is based on the mean square losses between the estimated and reference images in decibels (dB) as:

$$\text{PSNR} = 20 \log_{10} \left(\frac{I_{\max}}{\sqrt{MSE}} \right) \quad \text{dB}, \tag{1}$$

where,

$$MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(\text{I}_{\text{ref}}(m,n) - \text{I}_{\text{est}}(m,n) \right)^2 .$$

Here, I_{ref} and I_{est} (both sizes of $M \times N$) indicate the reference and estimated images, respectively; and I_{max} represents the maximum intensity in the reference image.

In addition to the evaluation based on PSNR, we investigate the sensitivity of the proposed beamforming technique with respect to SoS variation across different test images. For this purpose, we divide the whole range (1450–1550 m/s) of SoS distribution in the test set into 10 different non-overlapping sets, followed by computation of the PSNR in that non-overlapping region. Furthermore, we compare our approach with widely accepted DAS beamforming technique. Note that the SoS parameter in the DAS method is set to 1500 m/s in this comparison. Finally, we report the computation time of the proposed method in GPU to check its real-time capability.
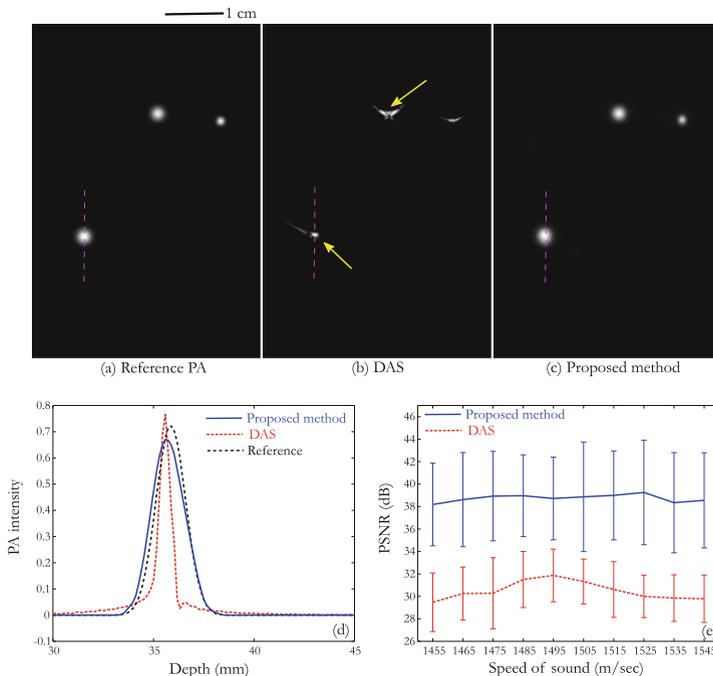
4.2   Results

Based on 1,500 test images, we achieve a PSNR of 38.7 ± 4.3 dB compared to that of 30.5 ± 2.4 dB for the DAS technique. A student t-test is performed to determine the statistical significance between the obtained results of these two techniques, and the obtained p-value $\ll 0.01$ indicates the superiority of the proposed method.

Figures 3(a–c) present a qualitative comparison among reference, DAS- and our CNN-beamformed images, respectively. In this particular example, we observe the distortion of the circular targets (marked by arrows) by the DAS beamforming technique. In contrast, the presented technique preserves the shapes and sizes of the targets. For a better visualization, we plot the PA intensity along a line (marked by dotted lines in Figs. 3(a–c)) in Fig. 3(d) and it indicates the promising performance of our proposed beamforming method, i.e., the width of the object is well preserved in our reconstruction.

Figure 3(e) shows a comparison between our and DAS methods in PSNR with respect to SoS variation based on the test set. As mentioned earlier, for a better interpretation of the sensitivity of the beamforming approaches, we divide the whole range (1450–1550 m/s) of SoS into 10 different non-overlapping sets, and we use the mean SoS of each set along the x-axis in this figure. The comparison between our and DAS beamforming techniques indicates less sensitivity of the DAS technique on the SoS variation in the test images. It is also interesting to note that the best performance of the DAS method is obtained for a SoS in the 1490–1500 m/s range. It is expected as this SoS is closer to the inherent SoS (1500 m/s) assumption in the DAS technique.

The run-time of the proposed CNN-based beamforming method is 18 ms in our GPU-based computer.

Fig. 3. Results and comparison of our proposed CNN-based beamforming method. (a–c) A comparison between DAS and our technique with respect to the reference PA image. The distortions in the targets are well visible in the DAS-beamformed image (marked by arrows). (d) The PA intensity variation along depth. This particular intensity variation corresponds to the dotted lines in (a–c). (e) Sensitivity of our and DAS methods with respect to SoS variation. For this figure, we divide the whole range of SoS (1450–1550 m/s) distribution in the test set into 10 different non-overlapping sets, and the x-axis represents the mean values of these sets. We can observe less sensitivity of our technique on SoS changes, in contrast, the DAS method shows its best performance at SoS near 1500 m/s.

5   Discussion and Conclusion

In this work, we have presented a deep CNN-based real-time beamforming approach to map the channel data to a PA image. Two notable modifications in the architecture are the incorporation of dense and dilated convolutions that lead to an improved training and a feature extraction without losing the resolution. The network has been trained using a set of simulation experiments with various contrasts and sizes of multiple targets. On the test set of 1,500 simulated images, we could obtain a mean PSNR of 38.7 dB.

A comparison of our result with that of the DAS beamforming method indicates a significant improvement achieved by the proposed technique. In addition, we have demonstrated how our method preserves the shapes and sizes of various targets in PA images (Figs. 3(a–d)).

We have investigated the sensitivity of the proposed and DAS beamforming methods on the SoS variation in test images. Since there is a inherent SoS assumption in the DAS method, it has shown its best performance at SoS near its SoS assumption. In contrast, our proposed beamforming technique has demonstrated less sensitivity to SoS changes in the test set (Fig. 3(e)).

Future works include training with non-circular targets, and testing on phantom and *in vivo* images. In addition, we aim to compare our method with neural networks-based beamforming approaches.

In conclusion, we have demonstrated the potential of the proposed CNN-based technique to beamform the channel data to a PA image in real-time while preserving the shapes and sizes of the targets.

References

1. Agarwal, A., et al.: Targeted gold nanorod contrast agent for prostate cancer detection by photoacoustic imaging. J. Appl. Phys. **102**(6), 064701 (2007)
2. Antholzer, S., Haltmeier, M., Schwab, J.: Deep learning for photoacoustic tomography from sparse data. arXiv preprint arXiv:1704.04587 (2017)
3. Beard, P.: Biomedical photoacoustic imaging. Interface Focus (2011). https://doi.org/10.1098/rsfs.2011.0028
4. Bell, M.A.L., Kuo, N., Song, D.Y., Boctor, E.M.: Short-lag spatial coherence beamforming of photoacoustic images for enhanced visualization of prostate brachytherapy seeds. Biomed. Optics Express **4**(10), 1964–1977 (2013)
5. Hoelen, C.G., de Mul, F.F.: Image reconstruction for photoacoustic scanning of tissue structures. Appl. Opt. **39**(31), 5872–5883 (2000)
6. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, p. 3 (2017)
7. Kang, J., et al.: Validation of noninvasive photoacoustic measurements of sagittal sinus oxyhemoglobin saturation in hypoxic neonatal piglets. J. Appl. Physiol. (2018)
8. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Luchies, A., Byram, B.: Deep neural networks for ultrasound beamforming. In: 2017 IEEE International Ultrasonics Symposium (IUS), pp. 1–4. IEEE (2017)
10. Luchies, A., Byram, B.: Suppressing off-axis scattering using deep neural networks. In: Medical Imaging 2018: Ultrasonic Imaging and Tomography, vol. 10580, p. 105800G. International Society for Optics and Photonics (2018)
11. Mozaffarzadeh, M., Mahloojifar, A., Orooji, M.: Medical photoacoustic beamforming using minimum variance-based delay multiply and sum. In: Digital Optical Technologies 2017, vol. 10335, p. 1033522. International Society for Optics and Photonics (2017)
12. Mozaffarzadeh, M., Mahloojifar, A., Orooji, M., Adabi, S., Nasiriavanaki, M.: Double-stage delay multiply and sum beamforming algorithm: application to linear-array photoacoustic imaging. IEEE Trans. Biomed. Eng. **65**(1), 31–42 (2018)

13. Mozaffarzadeh, M., Yan, Y., Mehrmohammadi, M., Makkiabadi, B.: Enhanced linear-array photoacoustic beamforming using modified coherence factor. J. Biomed. Opt. **23**(2), 026005 (2018)
14. Nair, A.A., Tran, T.D., Reiter, A., Bell, M.A.L.: A deep learning based alternative to beamforming ultrasound images (2018)
15. Park, S., Karpiouk, A.B., Aglyamov, S.R., Emelianov, S.Y.: Adaptive beamforming for photoacoustic imaging. Opt. Lett. **33**(12), 1291–1293 (2008)
16. Treeby, B.E., Cox, B.T.: k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. J. Biomed. Opt. **15**(2), 021314 (2010)
17. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
18. Zhang, H.K., et al.: Prostate specific membrane antigen (PSMA)-targeted photoacoustic imaging of prostate cancer in vivo. J. Biophotonics **13**, e201800021 (2018)