Towards a Fast and Safe LED-Based Photoacoustic Imaging Using Deep Convolutional Neural Network

Emran Mohammad Abu Anas[1(✉)], Haichong K. Zhang[1], Jin Kang[1], 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. The current standard photoacoustic (PA) technology is based on heavy, expensive and hazardous laser system for excitation of a tissue sample. As an alternative, light emitting diode (LED) offers safe, compact and inexpensive light source. However, the PA images of an LED-based system significantly suffer from low signal-to-noise-ratio due to limited LED-power. With an aim to improve the quality of PA images, in this work we propose to use deep convolutional neural networks that is built upon a previous state-of-the-art image enhancement approach. The key contribution is to improve the optimization of the network by guiding its feature extraction at different layers of the architecture. In addition to using a high quality target image at the output of the network, multiple target images with intermediate qualities are employed at in-betweens layers of the architecture to guide the feature extraction. We perform an end-to-end training of the network using a set of 4,536 low quality PA images from 24 experiments. On the test set from 15 experiments, we achieve a mean peak signal-to-noise ratio of 34.5 dB and a mean structural similarity index of 0.86 with a gain in the frame rate of 6 times compared to the conventional approach.

Keywords: Photoacoustic · LED · Laser
Convolutional neural networks · Densenet · Super-resoluton

1 Introduction

Photoacoustic (PA) is an emerging interventional imaging modality based on the photoacoustic phenomenon of generation of acoustic waves following light absorption in a soft-tissue sample. The primary applications of the PA technique include imaging of tissue chromophore (e.g. blood vessel) and exogenous contrast agents [3,6]. The standard work-flow of PA imaging starts with excitation of a sample using an intense short light pulse, followed by local thermo-elastic expansion due to sudden temperature rise. As a consequence of thermal expansion,

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11073, pp. 159–167, 2018.
https://doi.org/10.1007/978-3-030-00937-3_19

wideband acoustic signals are generated, and an ultrasound receiver is then used to collect the signal, which is usually known as PA signal.

For sources of PA imaging, the commercially available systems usually prefer Nd:YAG, Ti:Sapphire or dye laser [1], and they are capable to generate high energy laser pulse at biologically relevant wavelengths. Due to the high intense light source, a laser enclosure is recommended to install in the system to prevent the operator from incident irradiation. In addition to expensive and bulky laser system, such enclosure makes the system more cumbersome and does not allow the operator directly contacting with the sample [3].

Light emitting diode (LED) is a potential alternative that offers compact, safe and inexpensive illumination system in contrast to the conventional laser source. However, due to the limited output power, the PA signal of an LED-based system significantly suffers from low signal-to-noise-ratio (SNR) which in turn degrades the quality of the reconstructed PA images. To improve the SNR with LED-based system, the currently available technology acquires multiple (e.g. a few thousands) frames for the same sample and performs an averaging over them. In fact, the quality of a PA image is proportional to the number of frames used for averaging; an example is shown in Fig. 1(a) for two phantom PA images. Though a simple averaging over thousand of frames improves the image quality, it reduces the effective frame rate of PA images and more importantly, it often makes the PA images prone to motion artifact in *in vivo* applications (marked by circles in Fig. 1(b)). Therefore, it is recommended to use a less number of frames for averaging and perform standard signal processing to improve the SNR. The signal processing approach could be based on adaptive denoising, empirical mode decomposition or wavelet transform [2,3].

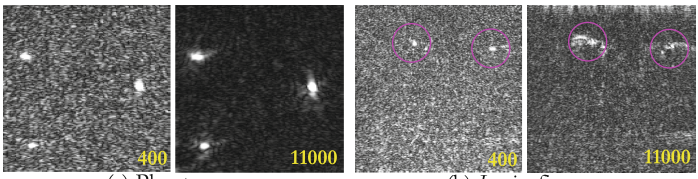

Fig. 1. Effect of averaging number (400 vs. 11000) on (a) a phantom and (b) an *in vivo* (proper digital arteries of fingers) examples. For the *in vivo* example, higher number of averaging frames introduces motion artifact (marked by circles).

In recent years, deep neural networks based approaches have shown promising performance in various applications compared to the previous state-of-the-art signal and image processing techniques. In addition to image classification, deep networks have been successfully used for image denoising [9] and image resolution improvement [7,8] that closely fit to our problem of image quality improvement.

Inspired from the success of neural networks, in this work, we present a deep convolutional neural networks (CNN)-based approach to improve the quality of

reconstructed PA images. Our architecture is built on a previous state-of-the-art image enhancement approach [8] that uses a series of dense convolutional layers to improve the quality of a 2D image. In addition, we propose to improve the optimization of the network by guiding its feature extraction using a sequence of target images. The target images are maintained to have an increasing order of image quality, and they are employed at different layers of the architecture to guide the feature extraction for an improved prediction of the image quality.

2 Methods

2.1 Architecture

Figure 2 shows our CNN-based architecture to improve the quality of reconstructed PA images; it takes a low quality PA image as input and at the end it generates an improved version of the given PA image. For convenience, in Fig. 2 we indicate the number of feature maps (or channels) for each convolutional layer as '*xx*' in 'Conv *xx*'. In addition, the successively increasing feature

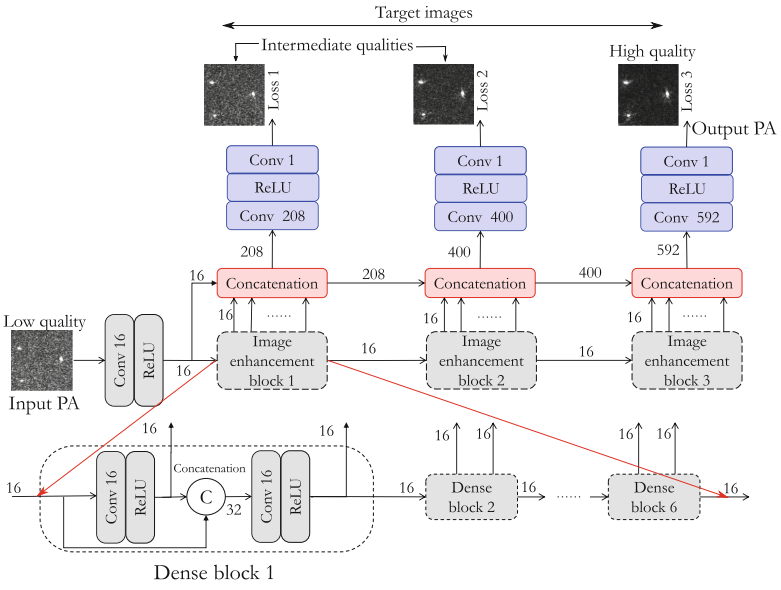

Fig. 2. The proposed architecture to improve the quality of a PA image. The input image is processed through three image enhancement networks, where each unit consists of six dense blocks. Furthermore, each dense block includes two dense 3×3 convolutional layers followed by rectified linear units. To generate an output image for each image enhancement block, all the features from the dense blocks are concatenated. Finally, a sequence of three target images with an increasing order of image quality are used to compute the mean square losses for three image enhancement blocks and these losses contribute equally to the total loss function.

maps are mentioned in the figure. The proposed architecture consists of three sequentially connected image enhancement blocks; for each block we use the architecture in [8] that was primarily proposed to improve the resolution of 2D images. Note that unlike [8] we do not use any upsamling layers in the network since both input and target images are of the same sizes (224×224 pixels) in our case. Following the architecture pattern in [8], we use six dense blocks to build each image enhancement network. Furthermore, each dense block includes two densely connected 3×3 convolutional layers followed by rectified linear units (ReLU). In principle, a dense convoluional layer uses all the features from its previous layers as inputs, as a result, it allows feature propagation more effectively and eliminates the vanishing gradient problem [5]. Note that instead of exactly using the same hyper-parameters (i.e. number of dense blocks and number of convoluional layers in each dense block) suggested in [8], we utilize the validation set to determine them (more about the validation set in Sect. 3.3 (Materials)). To produce an output image for each image enhancement block, all the generated features from the dense blocks are concatenated as shown in Fig. 2, and finally a convolution with one feature map is performed.

2.2 Loss Function

As mentioned earlier, the key contribution in this work is to guide the feature extraction of the network using a target sequence for an improved optimization. The sequence consists of three target images with an increasing order of image quality. As shown in Fig. 2, these three target images are sequentially fed to successive deeper layers. We compute the mean square losses corresponding to three target images, and subsequently we optimize the total loss function that is the average of these three individual losses. A detail description of generating a sequence of target images with an increasing order of image quality is provided in Sect. 3.4 (Training).

3 Experiments and Materials

3.1 LED Excitation Source

Our LED-based PA excitation source consisted of two LED matrices attached on both sides of the ultrasound probe. Each LED matrix included 144 LEDs arranged in four rows. The pulse repetition frequency of the LED was 1 kHz, i.e., the naive frame rate was 1 kHz. A synchronizer was employed to synchronize between LED excitation and PA signal reception.

3.2 Data Acquisition

We performed an experimental study using our LED-based PA system on 48 samples; 45 from blood mimicking phantoms and the rest 3 from volunteer fingers. For each sample, we acquired pre-beamformed PA signals for 11 s that led

to a collection of 11,000 frames of PA signals. Note that only for the phantom experiments, we could manage minimal vibration during the scanning period. After acquisition of the raw PA signals, an arithmetic averaging was performed over a number of frames (say, N). Then delay-and-sum method was used for beam-forming, followed by envelope detection to reconstruct the PA intensity image PA_N, where the subscript N indicates the number of frames used in averaging the PA signals before reconstruction.

3.3 Materials

To train, validate and test the proposed approach, we divide all of our experimentally acquired data into three groups. As mentioned earlier, for *in vivo* experiments, we could not be able to manage a steady condition during the scanning period. Therefore, the reconstructed images with high averaging frame number are affected by the motion artifacts (examples in Fig. 1(b)), subsequently, they are not used in any quantitative analysis. The training and validation sets consist of only of phantom data from 24 phantoms and 6 phantoms, respectively. And, the test set consists of 15 phantoms and 3 sets of *in vivo* data, where the latter one is used only for qualitative evaluation.

3.4 Training

To generate low quality input and high quality target PA images for the training, we exploit the positive effect of the averaging frame number (N) on the reconstructed image quality. For low quality inputs, we choose lower values of N in range of 200–4000, with a step of 200. For each chosen value of N, we divide the large set of 11,000 frames into a number of subsets, where each subset consists of N frames of PA signals. Next, within each subset of N frames, the raw PA signals are averaged, followed by reconstruction to obtain one PA image. For an example of $N = 200$, therefore, we obtain a total of 55 PA images from each sample. Calculating in the same way for all N in 200–4000 with a step of 200, we obtain a total of 189 input PA images from each experiment, subsequently 4,536 input training images from 24 experiments.

In contrast, for the target sequence, we use successive higher values of N to generate three target images with an increasing order of image quality. The values of N are, therefore, chosen from 5000–7000, 8000–10000, 11000 to generate three target images with an increasing order of image quality, where the latter one indicates the possible highest quality. Since we need higher values of N for the target images, it leads to less number of target images than input images. Therefore, it may be possible to have one target sequence correspond to more than one input images. We also perform random cropping of input (similarly for output too) images for a data augmentation in training. For the validation and test sets, we also generate input images using N in range of 200–4000 with a step of 200. However, intermediate (for N in 5000–7000 and 8000–10000) target images are not required in those cases, because we use only the final output of our network to compare with the highest quality target image PA_{11000}.

4 Evaluation and Results

4.1 Peak Signal-to-Noise-Ratio and Structural Similarity Index

For a quantitative evaluation of the proposed approach based on the test set, we use peak signal-to-noise-ratio (PSNR) and structural similarity index (SSIM) as evaluation indices [4] that compare the output of our network with the highest quality target image PA_{11000}. In addition, we compare our results with those from simple averaging and densenet [8]-based techniques. In the simple averaging technique, we do not perform any further processing on the PA images that are reconstructed from the already averaged PA signal. In contrast, for the densenet-based technique, we process the PA images using their reported architecture that

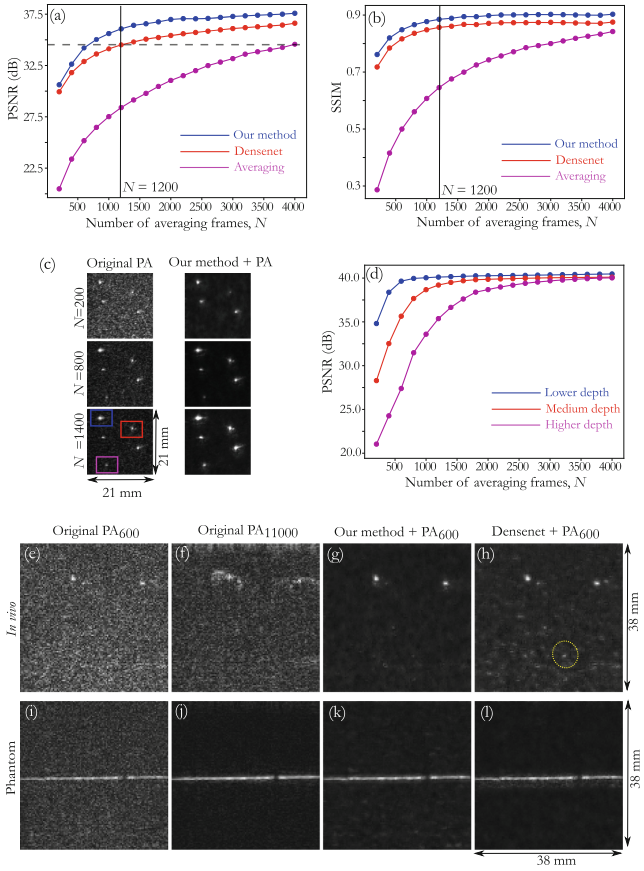

Fig. 3. Evaluation of the proposed approach. (a–b) Comparison of mean PSNR and SSIM of our method with those from simple averaging and densenet-based techniques [8] for different values of averaging frame numbers. (c–d) Effect of the imaging depth on the image quality. A qualitative comparison of our method with densenet-based technique for (e–h) an *in vivo* and (i–l) a phantom examples.

is in fact the same one shown in Fig. 2 but without using any intermediate target images during training. Note that the comparison is carried out for all values of averaging frame numbers (N) in range of 200–4000 (with a step of 200) that are used for generating low quality input images in the test set.

Figures 3(a–b) demonstrate a comparison of mean PSNR and mean SSIM (computed from 15 different experiments) of our method with those from the simple averaging and densenet-based methods for different values of N. We can notice improvement for our technique compared to two comparing methods for all values of N. In addition, we can observe comparatively higher drop in accuracy at $N < 1200$ for both of our and densenet-based methods than that at $N > 1200$. A rank-sum test is performed to measure the statistical improvement of our method with respect to two comparing methods. And we obtain p-values < 0.02 for all values of N both for PSNR and SSIM.

Another way to interpret the improvement is to compare the number of averaging frames needed to achieve a same image quality. For example, for a fixed PSNR of 34.5 dB, we achieve a gain in the frame rate of 6 times compared to the simple averaging technique ($N = 630$ vs. 4000 at dotted line in Fig. 3(a)). The corresponding mean SSIM for our method at $N = 630$ is 0.86.

4.2 Performance at Different Depths

We also investigate the performance of the proposed approach with respect to possible variations of imaging depths of targeted objects. For this purpose, we use three PA images (left column in Fig. 3(c)) of a same phantom, generated using three different values of N of 200, 800 and 1400. In addition, we present the results of our approach in the right column. To analyze the effect of imaging depth on the image quality, we select three region of interests (ROIs) around three point targets at different depths (shown in the figure). A qualitative comparison among the performance at those three ROIs indicates a successive reduced accuracy of the proposed technique for lower values of N while moving from a lower to higher depth. A corresponding quantitative analysis is presented in Fig. 3(d) that shows a comparison among the PSNRs of those ROIs for different values of N, where we can observe dependency of the performance of our method on imaging depth.

4.3 Qualitative Analysis

Figures 3(e–l) show a qualitative comparison between the proposed and densenet-based [8] methods for an *in vivo* and a phantom examples. As mentioned earlier, an averaging over a higher number of frames for the *in vivo* example (proper digital arteries of fingers) in our study leads to motion artifact in PA images. Taking the PA image with less number of averaging frames as input, our proposed method is able to improve its quality and subsequently suppresses the noise better, compared to the densenet-based technique (marked by circle in Fig. 3(h)). For the phantom example in Figs. 3(i–l), we can observe satisfactory performance for both of our and densenet-based methods. Though we train the network using

only the cross-sectional images of blood vessel mimicking phantoms, we can notice its satisfactory performance on unseen 'along the axis' image.

4.4 Computation Time

The computation time obtained using NVIDIA GeForce GTX 1080 Ti is 0.05 sec for both of our and densenet-based methods.

5 Discussion and Conclusion

In this work, we have presented a real-time approach to improve the imaging quality of LED-based PA imaging technique. The key contribution in our CNN-based method is to guide its feature extraction by using a sequence of target images employed at successive layers in the architecture. We have trained the network using a set of 4,536 low quality PA images from 24 phantom experiments. On the test set from 15 experiments, we could achieve a gain in the frame rate of 6 times compared to the conventional averaging approach, with a mean PSNR of 34.5 dB and a mean SSIM of 0.86. In addition, we have demonstrated a statistical significant improvement of the proposed method compared to the state-of-the-art CNN-based image enhancement approach [8] that in turn indicates the effectiveness of our contribution of guiding the feature extraction during training.

Though we have trained the network using data from blood mimicking phantoms, we have not only observed its satisfactory performance in *in vivo* example but also noticed elimination of motion artifacts resulting from a high number of averaging frames (Fig. 3(g)). In addition, we have demonstrated its promising performance on unseen imaging planes that had not been exposed during training (Fig. 3(k)).

We have observed a comparatively reduced accuracy of the proposed approach (other methods too) at lower averaging frame numbers ($N < 1200$). We can attribute its main reason to limitation of these methods at higher imaging depth (Fig. 3(d)). In fact, the PA signal from a higher depth is affected by more noise due to increased optical scattering. As a result, we need higher averaging frame numbers to achieve the desired quality.

Future works include quantitative validation of the proposed approach with *in vivo* examples. In addition to blood vessels, we aim to include other optically interested soft-tissue and exogenous contrast agents within the imaging targets. In conclusion, we have demonstrated the potential of the proposed technique to be included in a real-time LED-based PA imaging work-flow to improve the image quality as well as to achieve a gain in the imaging speed.

Acknowledgements. We would like to thank the National Institute of Health for funding this project.

References

1. Allen, T.J., Beard, P.C.: High power visible light emitting diodes as pulsed excitation sources for biomedical photoacoustics. Biomed. Opt. Express **7**(4), 1260–1270 (2016)
2. Hariri, A., Hosseinzadeh, M., Noei, S., Nasiriavanaki, M.: Photoacoustic signal enhancement: towards utilization of very low-cost laser diodes in photoacoustic imaging. In: Photons Plus Ultrasound: Imaging and Sensing 2017, vol. 10064, p. 100645L. International Society for Optics and Photonics (2017)
3. Hariri, A., Lemaster, J., Wang, J., Jeevarathinam, A.S., Chao, D.L., Jokerst, J.V.: The characterization of an economic and portable LED-based photoacoustic imaging system to facilitate molecular imaging. Photoacoustics **9**, 10–20 (2018)
4. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 2366–2369. IEEE (2010)
5. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE CVPR, vol. 1, p. 3 (2017)
6. Jeon, M., et al.: Methylene blue microbubbles as a model dual-modality contrast agent for ultrasound and activatable photoacoustic imaging. J. Biomed. Opt. **19**(1), 016005 (2014)
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
8. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: 2017 IEEE ICCV, pp. 4809–4817. IEEE (2017)
9. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems, pp. 341–349 (2012)