Optics Letters

End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging

Chuangjian Cai,[1] Kexin Deng,[1] Cheng Ma,[2,3,4] and Jianwen Luo[1,5]

[1]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China
[2]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[3]Beijing National Research Center for Information Science and Technology, Beijing 100084, China
[4]e-mail: cheng_ma@tsinghua.edu.cn
[5]e-mail: luo_jianwen@tsinghua.edu.cn

Received 23 March 2018; revised 27 April 2018; accepted 5 May 2018; posted 9 May 2018 (Doc. ID 326784); published 4 June 2018

An end-to-end deep neural network, ResU-net, is developed for quantitative photoacoustic imaging. A residual learning framework is used to facilitate optimization and to gain better accuracy from considerably increased network depth. The contracting and expanding paths enable ResU-net to extract comprehensive context information from multispectral initial pressure images and, subsequently, to infer a quantitative image of chromophore concentration or oxygen saturation (sO_2). According to our numerical experiments, the estimations of sO_2 and indocyanine green concentration are accurate and robust against variations in both optical property and object geometry. An extremely short reconstruction time of 22 ms is achieved.

OCIS codes: (170.6960) Tomography; (170.3010) Image reconstruction techniques; (170.3880) Medical and biological imaging; (170.5120) Photoacoustic imaging.

https://doi.org/10.1364/OL.43.002752

Photoacoustic (PA) imaging can form images of optical contrast with fine spatial resolution, high sensitivity, and good specificity in the optically diffusive regime [1]. Based on a series of multispectral PA images, quantitative PA imaging (QPAI) generates images of chromophore concentrations. Quantification enriches our capabilities to estimate absolute concentrations of administered contrast agents, saturation of oxygen, etc., thus offering better understanding of the biology problems being studied [2]. The conventional QPAI method [2] ignores the wavelength dependence of light fluence and employs linear fitting for estimating oxygen saturation (sO_2) (linear unmixing), which introduces substantial errors. Quantification accuracies can be improved by deep-tissue fluence correction, but most of these methods rely on overly ideal assumptions such as piecewise constant optical properties, *a priori* knowledge of scattering coefficients, and homogeneous (and known) background optical properties. Moreover, these traditional methods add significant computing burden if the pixel count is large [1,3]. Diffuse optical tomography can help estimate the fluence distribution at the expense of loss of high spatial frequencies and increased system complexity and cost [4]. The multispectral PA data cube embeds hidden features about the pursued quantitative information. Kirchner *et al.* proposed local context encoding (LCE), a machine learning method based on a random forest regressor, for QPAI [5]. However, LCE only relies on the measured PA signals in the local neighborhood and neglects other measured signals that contain important information. With a deep neural network (DNN) representing a complex mapping, deep learning (DL) can detect and "interpret" the critical features comprehensively. Thus, this suggests an alternative approach to solving QPAI problem using DL, which has been well demonstrated in segmentation and registration, labeling and captioning, computer-aided detection and diagnosis, etc. [6].

In this Letter, to the best of our knowledge, the first DNN framework, i.e., ResU-net, for QPAI is proposed. ResU-net takes the entire initial pressure images at different wavelengths as inputs, so the reconstruction makes the best of all the measured signals. In order to prevent the deep network from degradation, the residual learning mechanism is adopted. In ResU-net, comprehensive context information is extracted from the multispectral initial pressure images, for the purpose of quantification of chromophore concentration or sO_2.

In QPAI, the dimensionalities of the inputs and outputs are high. The inputs are initial pressure images acquired at different wavelengths. The outputs are quantitative images such as sO_2 maps or probe concentration images. To solve the QPAI problem using DL, the employed end-to-end neural network should be capable of (1) measuring the object's profile, (2) performing optical inversion, (3) suppressing noise, (4) estimating the types of major absorbing chromophores and their absorption spectra, and (5) conducting linear decomposition. This large-scale, nonlinear, and complicated problem requires special network designs. Convolutional neural networks (CNNs), which are composed of layers that filter the input to obtain useful information, are excellent at working with images. U-net is a fully CNN [7] composed of a contracting path to capture comprehensive context and a symmetric expanding path to enable precise localization, and was originally designed for biomedical image segmentation. This indicates that the framework of the

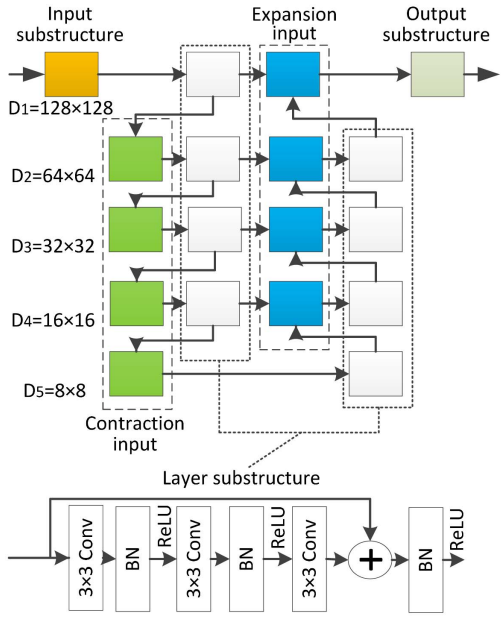

Fig. 1. Basic architecture of ResU-net. Each box corresponds to a residual learning substructure. $D_1, \ldots, D_5$ denote image dimensions, which are constant in each row. BN denotes batch normalization.

U-net is suitable for QPAI. A large number of stacked layers can enrich the level of features. However, too many layers will introduce the problem of vanishing/exploding gradients and hamper convergence. Batch normalization [8] reduces internal covariate shift and accelerates convergence. Additionally, degradation of training accuracy appears when the network depth keeps increasing. The residual learning mechanism [9] can address the degradation problem and improve accuracy from considerably increased depth. In this Letter, the residual learning mechanism is adopted to the U-net, which we named ResU-net, to address the QPAI problem.

In Fig. 1, the basic architecture of the proposed ResU-net is shown. ResU-net stacks residual learning substructures and is made up of a contracting path and a symmetric expanding path. In the contracting path, images shrink and comprehensive context information is extracted from the input initial pressure images. The feature images of different levels are fed to the expanding path. In the expanding path, the image size increases continuously through upsampling. Eventually, a high-resolution quantitative image is reconstructed. The number of images is called channel number, and the images constitute a tensor.

The details of the layer substructure are depicted in Fig. 1. The main connections are composed of convolutional layers, batch normalization layers, and rectified linear units (ReLUs). The tensor size stays constant. Besides the main connections, shortcut connections are added to perform identity mapping. In other residual learning substructures, the layers are adjusted based on the layer substructure because the tensor size is altered: the input substructure changes the channel number from the wavelength number of the initial pressure images to 32, while the output substructure changes the channel number from 32 to 1. The contraction input substructure decreases the image size to half through max-pooling and doubles the channel number. The channel numbers output from the contraction path to the expanding path are 32, 64, 128, 256, and 512 for the first, second, third, fourth, and fifth rows, respectively. The corresponding output tensors are concatenated to the ones of the main connections of the expansion input substructure.

The performance of ResU-net was validated using simulated data generated with a light propagation model, which solved the two-dimensional (2D) diffusion equation shown in Eq. (1) using the finite volume method:

$$\mu_a(\mathbf{r})\Phi(\mathbf{r}) - \nabla \cdot (D(\mathbf{r})\nabla\Phi(\mathbf{r})) = q_o(\mathbf{r}), \quad \forall \mathbf{r} \in \Omega, \tag{1}$$

where $D(\mathbf{r})$ is the optical diffusion coefficient, $\mu_a(\mathbf{r})$ denotes the optical absorption coefficient, $\Phi(\mathbf{r})$ denotes the optical fluence, and $q_o(\mathbf{r})$ is the source term. A set of initial pressure images, expressed as

$$p_0(\mathbf{r}) = \Gamma H(\mathbf{r}), \tag{2}$$

was obtained. In Eq. (2), p_o denotes the PA initial pressure, Γ denotes the Grüneisen parameter (assumed to be a constant), and $H(\mathbf{r}) = \mu_a(\mathbf{r})\Phi(\mathbf{r})$ is the absorbed energy density. p_0 acquired at different wavelengths, together with the corresponding chromophore concentration and/or sO_2 distributions, were used to train the network with the mean square error loss function and the Adam algorithm [10] implemented on Tensorflow. The training (2,048 samples) and testing (256 samples) data were obtained with randomly created maps of optical properties, simulating different physiological states and applications. The batch size for each iteration was 16, while the iteration number was 24,000. The training time was approximately 3 h on a 12 GB NVIDIA Titan GPU. After being well trained, it only took 22 ms on average to reconstruct a quantitative image. The superior performances of ResU-net were demonstrated using the following three numerical experiments. In all these simulations, we worked with 2D images closely representing those obtained using focused ultrasound arrays such as in Ref. [11].

In Experiment 1, simulations of p_0 of arbitrary tissues at different wavelengths (700–800 nm, step size 5 nm) were carried out to evaluate the quantification accuracy of sO_2. This experiment embodies the crucial application of imaging sO_2 in deep tissue, which bears significant physiological and pathological relevance such as assessing tumor growth, progression, and metastasis, as well as resistance to therapies [2]. In this experiment, light absorption was assumed to be caused by oxy-hemoglobin (HbO2) and deoxy-hemoglobin (Hb). A circular structure of 1 cm radius (128 × 128 pixels) was used. In order to evaluate the system's robustness to optical property variations, random maps of $sO_2(\mathbf{r})$, hemoglobin concentration $c_H(\mathbf{r})$, and a reduced scattering coefficient $\mu_s'(\mathbf{r})$ were generated, the values of which followed a Gaussian distribution (N). $sO_2(\mathbf{r}) \sim N(sO_2^{mean}, sO_2^{std})$, in which sO_2^{mean} followed uniform distribution U(80%, 90%) and sO_2^{std} is $5\% \cdot \mu_s'(\mathbf{r}) \sim N(\mu_s'^{mean}, \mu_s'^{std})$, $\mu_s'^{mean} \sim U(5\text{cm}^{-1}, 10\text{cm}^{-1})$ and $\mu_s'^{std} = 3\ \text{cm}^{-1}$. $\mu_a(\mathbf{r})$ were determined by $c_H(\mathbf{r})$ and $sO_2(\mathbf{r})$. The created $\mu_a(\mathbf{r})$ at 800 nm (isosbestic point of hemoglobin) followed $N(\mu_a^{mean}, \mu_a^{std})$, in which $\mu_a^{mean} \sim U(0.2\ \text{cm}^{-1}, 0.4\ \text{cm}^{-1})$ and $\mu_a^{std} = 0.05\ \text{cm}^{-1}$. Abrupt transitions between optically different regions were smoothed. White Gaussian noise (signal-to-noise [SNR} = 40 dB) was further superimposed to the output multi-wavelength p_o.

Table 1. Statistics of Relative Errors

	Experiment 1		Experiment 2	Experiment 3
	ResU-net	Linear unmixing	ResU-net	ResU-net
Mean	0.76% (sO_2)	36.90% (sO_2)	3.26% (ICG)	0.51% (sO_2) 7.51% (ICG)
Standard deviation	0.18% (sO_2)	1.22% (sO_2)	1.32% (ICG)	0.25% (sO_2) 3.02% (ICG)

There were 256 testing samples for each experiment. The mean and standard deviation of the relative error, which is the ratio of the L2 norm of the error to the true value for each sample, are listed in Table 1. For Experiment 1, ResU-net has a very low reconstruction error for the sO_2 map (mean error 0.76%) and very stable performance (standard deviation 0.18%). As shown, ResU-net outperforms linear unmixing (mean 36.90%).

To illustrate the performance of ResU-net, Figs. 2–4 show results representing average reconstruction accuracies. Figures 2(a) and 2(b) are sO_2 maps, reconstructed using ResU-net and linear unmixing, respectively. Apparently, for linear unmixing, the central region of the image suffers from a large estimation error due to spectral coloring associated with light absorption and scattering [Fig. 2(d)]. The small quantification error of ResU-net in Fig. 2(c) demonstrates that ResU-net can compensate well for spectral coloring.

Molecular imaging is another important application of PAI [12]. Quantification of the absolute concentrations of contrast agents has always been desired, yet extremely challenging [13]. In Experiment 2, 1–3 indocyanine green (ICG) labeled targets, whose shapes and sizes are representative of tumors, were introduced, and ResU-net was used to calculate the ICG concentration. Radiative decay was not considered here. ICG concentrations were assigned according

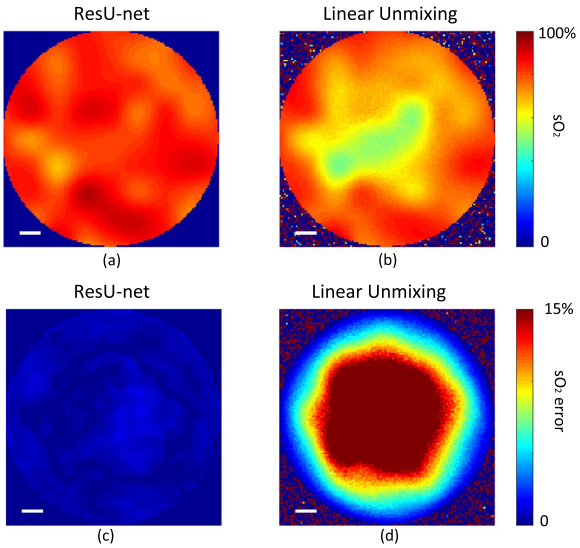

Fig. 2. sO_2 reconstruction results for Experiment 1. Each row shares the same color bar. The sO_2 reconstruction results of (a) ResU-net and (b) linear unmixing, together with their corresponding absolute errors (c) and (d) being shown. Scale bars: 2 mm.

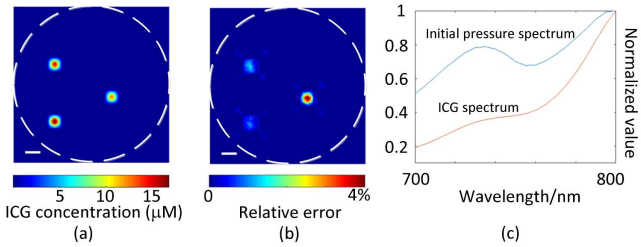

Fig. 3. Reconstruction results of the ICG concentrations in Experiment 2. (a) Reconstruction result of ResU-net. (b) Absolute value of the estimation error relative to the maximum ICG concentration. (c) Normalized ICG molar extinction coefficient and the initial pressure spectrum of the rightmost target. The dashed circle indicates an outline. Scale bar: 2 mm.

to $c_{\mathrm{ICG}} \sim \mathrm{N}(c_{\mathrm{ICG}}^{\mathrm{mean}}, c_{\mathrm{ICG}}^{\mathrm{std}})$, $c_{\mathrm{ICG}}^{\mathrm{mean}} \sim U(10\ \mu\mathrm{M}, 20\ \mu\mathrm{M})$, and $c_{\mathrm{ICG}}^{\mathrm{std}} = 2\ \mu\mathrm{M}$. The background tissue properties were set following the same procedures in Experiment 1. According to Table 1, ResU-net shows good accuracy (mean 3.26%). Figures 3(a) and 3(b) further demonstrate the performance by showing absolute errors (relative to the maximum real c_{ICG}) below 5%. In addition, the normalized spectrum of the ICG molar extinction coefficient at the given concentration, and the detected initial pressure of the rightmost target (absorption stems mainly from ICG) are plotted in Fig. 3(c). This verifies that spectral coloring significantly influences the target's spectrum.

To better model *in vivo* physiological characteristics, a digital mouse [14] was constructed in Experiment 3. Distinct from the circular structure in Experiments 1 and 2, the digital mice used in Experiment 3 had irregular geometries, including the outer profile and the organs' shape. Moreover, the optical properties of *in vivo* tissues are extremely heterogeneous. For example, μ_a of the liver can be 6 times larger than that of the heart wall, while μ_s' of the bone can be 3.6 times larger than that of the liver [15]. In order to test ResU-net's generalization ability, variations in the geometry, rotation, and body size were taken into account. For any sample in the training and testing data, an arbitrary slice of the digital mouse was chosen, then rotated and zoomed (magnification 0.7–1) randomly. The optical parameters of the organs were set referring to Ref. [15]. The weak wavelength dependence of μ_s' was considered. Assuming that optical absorption mainly results from HbO2, Hb, water, and ICG, the total absorption coefficient is calculated as

$$\mu_a(\mathbf{r},\lambda) = S_B(\mathbf{r})(\mathrm{sO}_2(\mathbf{r})\mu_{\mathrm{aHbO}_2}(\lambda) + (1-\mathrm{sO}_2(\mathbf{r}))\mu_{\mathrm{aHb}}(\lambda)) + S_W(\mathbf{r})\mu_{aW}(\lambda) + \ln(10)c_{\mathrm{ICG}}(\mathbf{r})\varepsilon_{\mathrm{ICG}}(\lambda), \quad (3)$$

where S_B and S_W are scaling factors, and $\varepsilon_{\mathrm{ICG}}(\lambda)$ is the molar extinction coefficient of ICG. Two networks were trained simultaneously for the reconstruction of the sO_2 and ICG concentration. According to Table 1, small quantification errors (0.51% for sO_2 and 7.51% for ICG) reveal that ResU-net is capable of multi-parameter quantification. In Fig. 4, the simultaneous reconstruction results of sO_2 and ICG are shown. The ICG quantification error in Experiment 3 is slightly larger than that in Experiment 2, which might be due to the complex geometry and the severe heterogeneity of the digital mouse.

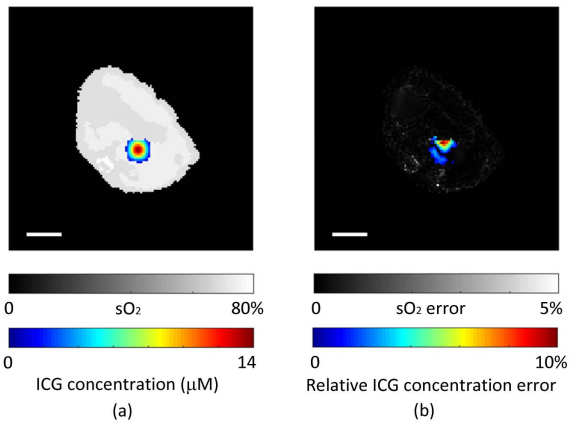

Fig. 4. Simultaneous reconstruction of sO_2 and ICG concentrations in Experiment 3. (a) Results generated by ResU-net. (b) Absolute error of sO_2 (grayscale) and absolute value of the relative ICG concentration error (pseudo-color). Scale bar: 5 mm.

Table 2. sO_2 Reconstruction Relative Errors under Different Noise Levels

SNR	40 dB	30 dB	20 dB	10 dB
Mean	0.76%	0.81%	1.43%	16.83%
Standard deviation	0.18%	0.19%	0.27%	6.64%

Different noise levels were also tested with the trained network in Experiment 1 (Table 2). The mean relative errors of both 40 and 30 dB SNRs are below 1%, which illustrates that ResU-net can perform precise optical inversion in low noise conditions. When the SNR is 20 dB, the mean relative error is only 1.43%, indicating that noise can be effectively suppressed. However, when the SNR decreases to 10 dB, the noise completely distorts the p_o images, resulting in poor quantification.

In this Letter, our method to remove the need for overly ideal assumptions was shown. Conventional methods require that all of the major absorbing chromophores, and their absorption spectra, are known in advance. However, *in vivo* tissues are composed of various optical absorbers such as hemoglobin, melanin, water, and extrinsic probes, and *a priori* knowledge about their concentrations is very limited. ResU-net can detect the absorber types and measure their spectral information during training, so explicit *a priori* knowledge is not needed. The automatic extraction of the body contour is fulfilled implicitly, which is necessary for optical inversion. Moreover, quantitative image reconstruction using ResU-net requires no iteration, so the computational burden and memory requirement are low. Our *in silico* experiments demonstrated ResU-net's high quantification accuracy and its robustness to the variations of the object's optical properties and geometry.

For *in vivo* imaging, a large amount of gold standard quantitative images is difficult to obtain, so transfer learning is important. Simulation data could be used for pretraining. In the future, more realistic properties and 3D modeling will be considered. Experiments with phantom objects or objects containing probes imbedded into tissue or live animal body facilitate learning. Moreover, acoustic detection and inversion have not been considered. Reference [16] demonstrated that DL is effective to reduce artifacts generated by filtered back-projection. Therefore, networks can compensate for imperfect detection. Otherwise, model-based and time-reversal methods are good candidates for acoustic inversion.

In conclusion, an end-to-end deep CNN, i.e., ResU-net, is proposed for QPAI. ResU-net constructs mapping from multispectral PA data to quantitative images of chromophore concentrations. In our preliminary simulations, ResU-net was able to show accurate and simultaneous reconstructions of ICG concentrations and sO_2 distributions of a digital mouse. Our method paves the way for viable and fast QPAI, and we hope that its power can be further demonstrated and developed in future studies, especially *in vivo* animal/clinical experiments.

Funding. National Natural Science Foundation of China (NSFC) (61735016, 81471665); Youth Innovation Fund of Beijing National Research Center for Information Science and Technology; 1000 Youth Talents Program in China.

REFERENCES

1. B. T. Cox, J. G. Laufer, P. C. Beard, and S. R. Arridge, J. Biomed. Opt. **17**, 061202 (2012).
2. M. L. Li, J. T. Oh, X. Xie, G. Ku, W. Wang, C. Li, G. Lungu, G. Stoica, and L. V. Wang, Proceedings of the IEEE **96**, 481 (2008).
3. F. M. Brochu, J. Brunker, J. Joseph, M. R. Tomaszewski, S. Morscher, and S. E. Bohndiek, IEEE Trans. Med. Imaging **36**, 322 (2017).
4. A. Q. Bauer, R. E. Nothdurft, J. P. Culver, T. N. Erpelding, and L. V. Wang, J. Biomed. Opt. **16**, 096016 (2011).
5. T. Kirchner, J. Gröhl, and L. Maier-Hein, "Local context encoding enables machine learning-based quantitative photoacoustics," arXiv:1706.03595 (2017).
6. J. G. Lee, S. Jun, Y. W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, Korean J. Radiol. **18**, 570 (2017).
7. O. Ronneberger, P. Fischer, and T. Brox, *International Conference on Medical Image Computing and Computer-assisted Intervention* (2015), p. 234.
8. S. Ioffe and C. Szegedy, *International Conference on Machine Learning* (2015), p. 448.
9. K. He, X. Zhang, S. Ren, and J. Sun, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), p. 770.
10. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
11. L. Li, L. Zhu, C. Ma, L. Lin, J. Yao, L. Wang, K. Maslov, R. Zhang, W. Chen, J. Shi, and L. V. Wang, Nat. Biomed. Eng. **1**, 0071 (2017).
12. J. Weber, P. C. Beard, and S. E. Bohndiek, Nat. Methods **13**, 639 (2016).
13. S. Tzoumas and V. Ntziachristos, Phil. Trans. R. Soc. A **375**, 20170262 (2017).
14. B. Dogdas, D. Stout, A. F. Chatziioannou, and R. M. Leahy, Phys. Med. Biol. **52**, 577 (2007).
15. G. Alexandrakis, F. R. Rannou, and A. F. Chatziioannou, Phys. Med. Biol. **50**, 4225 (2005).
16. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," arXiv:1704.04587 (2017).