



# Deep learning-based quantitative optoacoustic tomography of deep tissues in the absence of labeled experimental data

JIAO LI,<sup>1,2</sup> CONG WANG,<sup>1</sup> TINGTING CHEN,<sup>1</sup> TONG LU,<sup>1</sup> SHUAI LI,<sup>1</sup> BIAO SUN,<sup>3,4,8</sup>  FENG GAO,<sup>1,2,9</sup> AND VASILIS NTZIACHRISTOS<sup>5,6,7,10</sup>

<sup>1</sup>School of Precision Instruments and Optoelectronics Engineering, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

<sup>2</sup>Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

<sup>3</sup>School of Electrical and Information Engineering, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

<sup>4</sup>Tianjin Key Laboratory of Process Measurement and Control, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

<sup>5</sup>Institute of Biological and Medical Imaging, Helmholtz Munich, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany

<sup>6</sup>Chair of Biological Imaging at the Central Institute for Translational Cancer Research, (TranslaTUM), School of Medicine, Technical University of Munich, Ismaningerstr. 22, D-81675 Munich, Germany

<sup>7</sup>Munich Institute of Robotics and Machine Intelligence (MIRMI), Technical University of Munich, Georg-Brauchle-Ring 60, 80992 Munich, Germany

<sup>8</sup>e-mail: sunbiao@tju.edu.cn

<sup>9</sup>e-mail: gaofeng@tju.edu.cn

<sup>10</sup>e-mail: bioimaging.translatum@tum.de

Received 10 August 2021; revised 22 November 2021; accepted 22 November 2021; published 6 January 2022

Deep learning (DL) shows promise for quantitating anatomical features and functional parameters of tissues in quantitative optoacoustic tomography (QOAT), but its application to deep tissue is hindered by a lack of ground truth data. We propose DL-based “QOAT-Net,” which functions without labeled experimental data: a dual-path convolutional network estimates absorption coefficients after training with data-label pairs generated via unsupervised “simulation-to-experiment” data translation. In simulations, phantoms, and *ex vivo* and *in vivo* tissues, QOAT-Net affords quantitative absorption images with high spatial resolution. This approach makes DL-based QOAT and other imaging applications feasible in the absence of ground truth data. © 2022 Optical Society of America under the terms of the

OSA Open Access Publishing Agreement

<https://doi.org/10.1364/OPTICA.438502>

## 1. INTRODUCTION

Optoacoustic tomography (OAT), also called photoacoustic tomography, combines high optical absorption contrast with spatial-resolution imaging deep in tissue [1]. OAT has shown clinical potential in small-animal studies and human trials focusing on the brain, arm, breast cancer, vascular and joint diseases [2–12]. It can generate high-resolution images of tissue function because it can track endogenous molecules such as hemoglobin, melanin, and lipids, as well as exogenous probes such as fluorescent agents and nanoparticles. The accurate evaluation of chromophore concentrations obtained by such imaging requires accurate estimation of the absorption coefficient  $\mu_a$  of target tissues [13–19], which is challenging because it depends on photon fluence  $\Phi$  that attenuates as the light propagates deeper into the tissue.

Quantitative optoacoustic tomography (QOAT) aims to improve the reconstruction accuracy of  $\mu_a$  by starting from conventional images of the distribution of initial pressure and treating the pressure as the product of  $\mu_a$  and  $\Phi$  [20,21]. This requires solving the non-linear, ill-posed optical inverse problem, which

imposes major challenges [22]. In one QOAT approach,  $\Phi$  is estimated by assuming homogeneous or empirical optical properties, but this assumption may not be accurate, leading to substantial reconstruction errors [23,24]. A second multimodal approach combines OAT with other imaging modes such as diffuse optical tomography or acoustic-optic tomography in order to calculate the fluence distribution, but this requires more elaborate systems and computational resources [25–27].

A third approach is iterative reconstruction to minimize error, in which an appropriate optimization strategy is chosen with the established forward model of photon transport [28–37]. This approach depends on accurate calibration between the acoustic reconstruction and optical model, which can be challenging. In previous work, we introduced a reference phantom with known optical properties into the calibration process [38], but this increases computational complexity and may be difficult to implement for biological tissues with irregular boundaries. In addition, most optimization algorithms require prior empirical information in order to define the initialization and regularization parameters, and they involve computationally expensive iterative calculations.

A promising alternative to all these approaches may be deep learning (DL). Deep neural networks (DNNs), such as convolutional neural networks (CNNs), have demonstrated great potential to remove artifacts from initial pressure  $P_0$  images caused by limited-view setups [39–44] or sparse detection [45–47]. They can also reconstruct artifact-free  $P_0$  images from raw optoacoustic signals [39,48,49], and they can estimate optical absorption coefficients [50,51], chromophore concentrations, and oxygen saturation from multispectral  $P_0$  images [52].

DNNs are typically trained in a supervised manner, which in QOAT requires data-label pairs, i.e.,  $P_0$  images and their corresponding optical absorption coefficient images. However, obtaining such experimental data with ground truth images of deep tissues is challenging. Even when labeled experimental data exist, they are usually impractical, costly, and prohibitively time consuming. Using DNNs trained on data from simulations of photon and ultrasound propagation is inapplicable to tissue imaging, due to substantial system noise, scaling errors, and other types of mismatch between experiment and model [38,53]. Although our previous work with phantoms confirmed the feasibility of applying DNNs to QOAT [54], the lack of adequate data-label pairs for tissue imaging remains a severe limitation.

In this work, we break through this limitation by translating labeled simulation data into the experimental domain to generate data-label pairs. Inspired by unsupervised image translation methods such as CycleGAN [55], we propose a simulation-to-experiment end-to-end data translation network (SEED-Net), which provides experimental datasets with ground truth images through unsupervised data translation from abundant simulation datasets of modeled tissues, such as the digital mouse, the digital brain [56], and other tissues [57,58]. Then, a QOAT-Net is presented to reconstruct high-resolution images of  $\mu_a$  for deep tissues in which a novel dual-path network is trained using abundant data-label pairs generated by SEED-Net. The specially designed dual-path network is based on the relationships among  $P_0$ ,  $\mu_a$ , and  $\Phi$ , whose characteristics are used to optimize the network. To the best of our knowledge, this is the first time that DL has been used to reconstruct the distribution of absorption coefficients for deep tissues, yielding  $\mu_a$  images of absolute quantification without sacrificing spatial resolution.

## 2. METHODS

### A. Concept and Structure of QOAT-Net

The SEED-Net [Fig. 1(a)] aims to translate images from the simulation domain into the experiment domain without human labeling (unsupervised learning), thus establishing an interconnection between a large amount of standardized simulation data and experimental data for the system under study. The architecture consists of two generative adversarial networks (GANs):  $G_{S \rightarrow E}$  and  $G_{E \rightarrow S}$ , each of which is trained to translate an image from the simulation domain into the experiment domain and vice versa. Each GAN consists of a generator network and a discriminator network. The generator network is given an input image and produces an output image using CNN architecture with residual blocks [Fig. S1(a) in Supplement 1]. The discriminator is a classifier that receives an image and predicts whether it is genuine or created by a generator [Fig. S1(b) in Supplement 1]. In the SEED-Net, each generator takes images from its respective domain (simulation or experiment) and creates images in the opposite domain (experiment or simulation). While each discriminator is

trained to distinguish generated images from real ones, the generators in turn are trained to fool the discriminators. To ensure true data translation, a cyclic constraint is imposed in which the generated images by  $G_{S \rightarrow E}$  or  $G_{E \rightarrow S}$  are put into the generators of the corresponding domain ( $G_{E \rightarrow S}$  or  $G_{S \rightarrow E}$ ) and the output must be identical to the original image used to create the generated one. The trained generator ( $G_{S \rightarrow E}$ ) can then generate abundant experimental datasets, which are translated from rich digital images and now contain known optical parameters as ground truth. These generated datasets with the known ground truth can be used as data-label pairs for training the second sub-network.

To quantify  $\mu_a$ , the QOAT-Net [Fig. 1(b)] solves the QOAT inverse problem, in which  $P_0$  depends not only on  $\mu_a$  but also on  $\Phi$ . A novel dual-path network based on U-Net is constructed (Fig. 2), where the top path focuses on the generation of  $\Phi$ , while the bottom path recovers  $\mu_a$  in the tissues. Images of  $\mu_a$  are obtained from rich models of digital biological tissues [56–60] and serve as the ground truth for training the QOAT-Net as well as for simulating initial pressures  $P_0^S$  using the Monte Carlo method [61,62].  $P_0^S$  is translated to generated experiment  $P_0^G$  via the established data translation network ( $G_{S \rightarrow E}$ ), and generated  $P_0^G$  is input into the QOAT-Net. Dividing generated  $P_0^G$  by  $\mu_a$  yields calculated  $\Phi$ . In this way, the loss function comprises  $\mu_a$ ,  $P_0$ , and  $\Phi$ , which is more compatible with the mathematical model of QOAT than the single U-Net architecture [54].

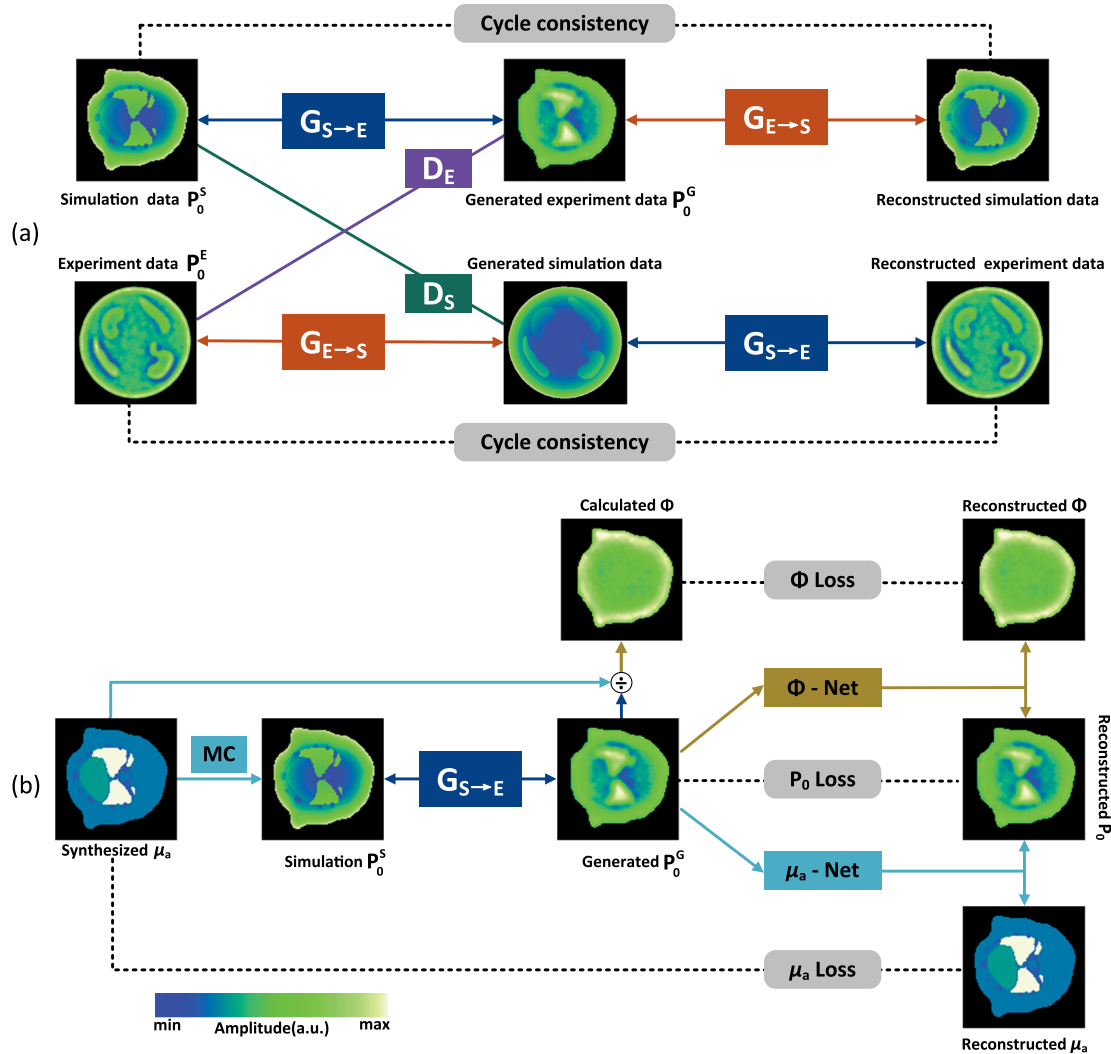
### B. SEED-Net Architecture

Our network architecture consisted of two GANs [63], one of which was trained to transfer an image from simulation domain to experiment domain ( $G_{S \rightarrow E}$ ), while the other was trained to transfer an image in the opposite direction ( $G_{E \rightarrow S}$ ) [Fig. 1(a)]. Each GAN consisted of a generator network and a discriminator network. The generator network produced an output image from an input image, while the discriminator ( $D_S$  or  $D_E$ ) classified images as genuine or created by the generator. To ensure realistic data translation, the following cyclic constraint was imposed: generators of the corresponding domain ( $G_{E \rightarrow S}$  or  $G_{S \rightarrow E}$ ) were fed generated images from  $G_{S \rightarrow E}$  or  $G_{E \rightarrow S}$  and had to generate images identical to the original one. The key advantage of this approach is that it does not require samples of the same tissues in both the simulation and experiment domains. As a result, one can train the network using unlabeled experiment data from an arbitrary OAT system as well as simulation data or publicly available labeled data similar to real experimental data expected for the situation under study.

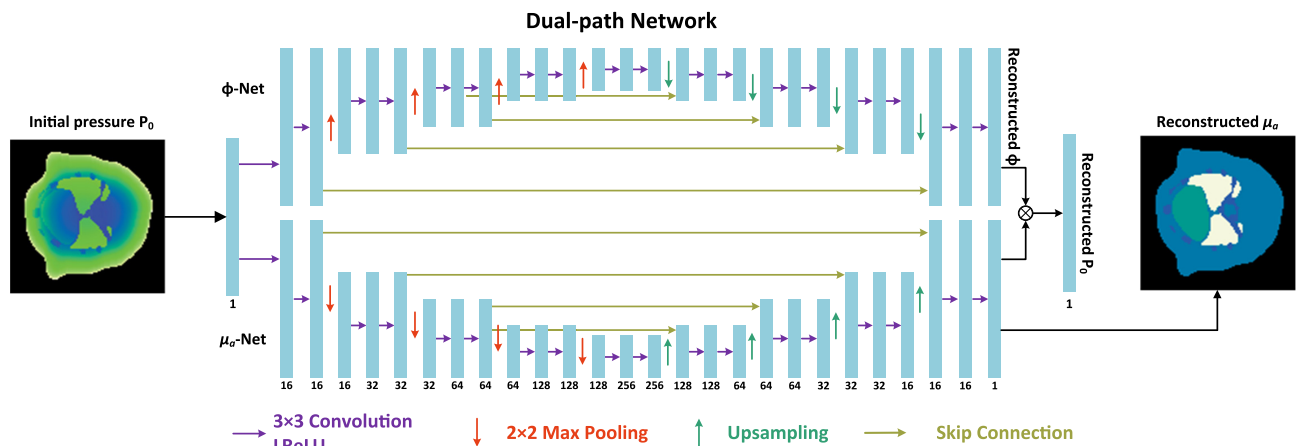
The generators were formulated into an encoder–decoder framework with residual blocks [Fig. S1(a) in Supplement 1]. The encoder downsampled high-dimensional data into embedded representations, whereas the decoder upsampled high-dimensional input. We used stride convolution for downsampling instead of a pooling layer, and we used deconvolution for upsampling. Each convolutional and deconvolutional layer was followed by batch normalization [64] and a leaky rectified linear unit (LReLU) activation function [65], defined as

$$\text{LReLU} = \begin{cases} x & x > 0 \\ 0.1x & x \leq 0. \end{cases} \quad (1)$$

Other than the first and last layers, which used  $7 \times 7$  kernels with stride  $1 \times 1$ , all convolutional and deconvolutional layers used  $3 \times 3$  kernels with sliding stride  $2 \times 2$ . We constructed the bottleneck of the generators by concatenating nine convolution



**Fig. 1.** Principle of absorption coefficient reconstruction based on deep learning. (a) Procedure of the SEED-Net. Cycle consistency denotes the cycle consistency loss in this network.  $G_{S \rightarrow E}$  generates experimental data from simulation, while  $G_{E \rightarrow S}$  does the opposite. Discriminators  $D_S$  and  $D_E$  are trained to differentiate original and generated images. (b) Process of the QOAT-Net. Monte Carlo (MC) simulation calculates simulation initial pressures  $P_0^S$  from synthesized  $\mu_a$ . Generated experiment pressures  $P_0^G$  are produced from the  $P_0^S$  via the data translation network ( $G_{S \rightarrow E}$ ) and input into the dual-path network to obtain reconstructed images of photon fluence  $\Phi$  and  $\mu_a$ , whose product is reconstructed  $P_0$ .  $P_0^G$  is divided by  $\mu_a$  to calculate  $\Phi$ . The Grüneisen parameter  $\Gamma$  is considered to be spatially constant and equal to 1 in the present study.



**Fig. 2.** Dual-path network architecture features two parallel mapping architectures based on U-Net. The top path generates photon fluence  $\Phi$ , while the bottom path reconstructs absorption coefficient  $\mu_a$ . LReLU, leaky rectified linear unit.

residual blocks, with each block containing two  $3 \times 3$  convolutional layers and a residual connection [66]. Based on the notation  $k \times m \times n$  to denote  $k$  channels of feature maps in a spatial size of  $m \times n$ , the size of input images was  $1 \times 256 \times 256$ . Feature map size changed from one convolution layer to the next as follows:  $1 \times 256 \times 256 \rightarrow 32 \times 256 \times 256 \rightarrow 64 \times 128 \times 128 \rightarrow 128 \times 64 \times 64 \rightarrow 256 \times 32 \times 32$  (residual blocks)  $\rightarrow 128 \times 64 \times 64 \rightarrow 64 \times 128 \times 128 \rightarrow 32 \times 256 \times 256 \rightarrow 1 \times 256 \times 256$ . Thus, the output of the generator was a domain-translated image of  $1 \times 256 \times 256$ .

The discriminators had a CNN architecture comprising four down blocks, each of which contained a convolutional layer with sliding stride  $2 \times 2$ , a batch normalization layer, and an LReLU activation function [Fig. S1(b) in Supplement 1]. The down block was followed by a convolutional layer with stride  $1 \times 1$ . Convolution kernels throughout the discriminators were set to  $4 \times 4$ . Such a patch-level discriminator design has fewer parameters than a full-image discriminator and can work on images of arbitrary size in a fully convolutional fashion [67].

### C. Dual-Path Network Architecture

The QOAT-Net had a novel dual-path architecture based on the widely used U-Net [68] to quantify  $\mu_a$  from initial pressure (Fig. 2). Two U-Nets were organized in parallel. The top path generated  $\Phi$ , while the bottom path estimated  $\mu_a$ . Both paths contained a contraction stream to downsample the input image size and a symmetric expansive stream to restore the original image size. A skip connection was employed to pass information between layers of the corresponding level. The contraction stream consisted of four downsampling blocks, each of which was composed of a max pooling layer and two convolutional layers. The max pooling layer contained a  $2 \times 2$  kernel with sliding stride of  $2 \times 2$ . The convolutional layer contained a  $3 \times 3$  kernel with sliding stride  $1 \times 1$  and was followed by an LReLU activation function. The  $k$ th downsampling block, which mapped feature map  $x_k$  onto feature map  $x_{k+1}$ , was given by

$$x_{k+1} = \text{MP} \{ \text{LReLU} \{ \text{CONV} [ \text{LReLU} \{ \text{CONV} [x_k] \} ] \} \}, \quad (2)$$

where  $\text{CONV}[\cdot]$  is a convolution operator that includes bias terms, and  $\text{MP}\{\cdot\}$  denotes the max pooling layer. The expansive stream consisted of four symmetrical upsampling blocks, each of which included an upsampling layer and two convolutional layers with an LReLU activation function. Skip connections were established between layers of equal resolution in the downsampling and upsampling layers in order to compensate for the loss in spatial resolution resulting from the multiple downsampling operations. The  $k$ th upsampling block, which mapped a feature map  $y_k$  into feature map  $y_{k+1}$ , was given by

$$y_{k+1} = \text{LReLU} \{ \text{CONV} [ \text{LReLU} \{ \text{CONV} [ \text{MERGE}(x_{k+1}, \text{UPS}(y_k)) ] \} ] \}, \quad (3)$$

where  $\text{UPS}(\cdot)$  is the upsampling operator,  $\text{MERGE}(\cdot)$  denotes a skip connection that merges the number of channels, and  $x_{k+1}$  is the output of the downsampling convolutional layer symmetric to the current upsampling layer. Feature map size changed through the downsampling blocks and upsampling blocks as follows:  $1 \times 256 \times 256$  (input)  $\rightarrow 16 \times 128 \times 128 \rightarrow 32 \times 64 \times 64 \rightarrow 64 \times 32 \times 32 \rightarrow 128 \times 16 \times 16 \rightarrow 64 \times 32 \times 32 \rightarrow 32 \times 64 \times 64 \rightarrow 16 \times 128 \times 128 \rightarrow 1 \times 256 \times 256$  (output).

### D. Training

Calculations were carried out in a Python 3.7.1 environment on an Intel 3.50-GHz Core i7 PC with 32 GB of RAM and an Nvidia Titan Xp GPU. PyTorch 1.6.0 was used for designing and testing the SEED-Net and the QOAT-Net.

The total loss function of the SEED-Net consisted of the adversarial loss and the cycle consistency loss [55], defined as

$$L_{CG} = L_{GAN}(G_{S \rightarrow E}, D_E) + L_{GAN}(G_{E \rightarrow S}, D_S) + \lambda L_{cyc}(G_{S \rightarrow E}, G_{E \rightarrow S}), \quad (4)$$

where  $L_{GAN}(\cdot, \cdot)$  denotes the adversarial loss of a generator and its discriminator,  $L_{cyc}(\cdot, \cdot)$  denotes the cycle consistency loss of two generators, and  $\lambda$  controls the relative importance of the two losses.

The loss function of the QOAT-Net was defined as

$$L = \alpha \text{MSE}(\Phi^{\text{output}}, \Phi^{\text{ground-truth}}) + \beta \text{MSE}(\mu_a^{\text{output}}, \mu_a^{\text{ground-truth}}) + \text{MSE}(P_0^{\text{output}}, P_0^{\text{ground-truth}}) + \gamma \|\Phi^{\text{output}}\|_F + \delta \text{TV}(\mu_a^{\text{output}}), \quad (5)$$

where  $\Phi^{\text{ground-truth}}$  denotes the ground truth  $\Phi$ ;  $\Phi^{\text{output}}$  denotes the output of the top path;  $\mu_a^{\text{ground-truth}}$  denotes the ground truth  $\mu_a$ ;  $\mu_a^{\text{output}}$  denotes the output of the bottom path;  $P_0^{\text{ground-truth}}$  denotes the ground truth initial pressure  $P_0$ ; and  $P_0^{\text{output}}$  denotes the output of dual-path network, which is the multiplication of  $\Phi^{\text{output}}$  and  $\mu_a^{\text{output}}$ .  $\text{MSE}(\cdot, \cdot)$  denotes pixel-wise mean squared error between two images. Considering the gradual behavior of  $\Phi$ , we added the Frobenius norm of  $\Phi^{\text{output}}$  as a regularization term to the loss function

$$\|\Phi^{\text{output}}\|_F = \sqrt{\sum_p \sum_q |\Phi_{p,q}^{\text{output}}|^2}, \quad (6)$$

where  $p$  and  $q$  are pixel indices. The TV operator of  $\mu_a$  was defined as

$$\text{TV}(\mu_a^{\text{output}}) = \sum_p \sum_q \sqrt{(\mu_{a,p+1,q}^{\text{output}} - \mu_{a,p,q}^{\text{output}})^2 + (\mu_{a,p,q+1}^{\text{output}} - \mu_{a,p,q}^{\text{output}})^2}. \quad (7)$$

Regularization parameters  $\alpha(100)$ ,  $\beta(200)$ ,  $\gamma(10^{-5})$ , and  $\delta(10^{-4})$  were applied to the loss function  $L$  to enable the appropriate balance among the five components: the MSEs of the output  $\Phi$ ; output  $\mu_a$  and output  $P_0$  with respect to their ground truth values; the Frobenius norm of the output photon fluence  $\Phi$ ; and the TV of the output absorption coefficient  $\mu_a$ . In order to prove the effect of the first three terms in the loss function, we retrained the neural network three times with only two of the three terms included. The results of inputting the same test data into those retrained networks are displayed in Fig. S2 in Supplement 1, showing the best performance using the loss function including all three terms.

Both networks were trained using adaptive moment estimation optimization [69] with a learning rate of  $10^{-4}$ . The batch size in our training stage was 16. Loss functions quickly converge to an equilibrium stage after 200 epochs, after which they remain stable, as shown in Fig. S3 in Supplement 1. The dual-path network trains in 94 min with a training dataset of 3040 pairs of simulation square images.



### 3. DATASETS AND EXPERIMENTAL SETUP

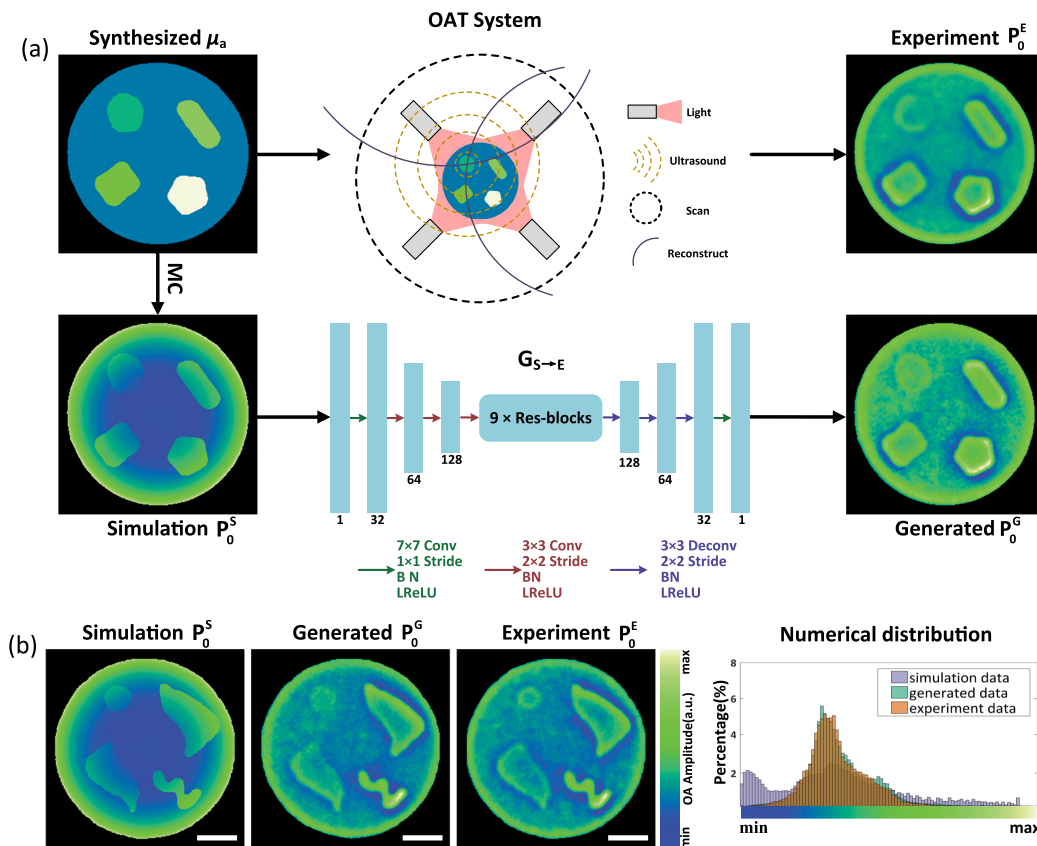
#### A. Simulation Datasets

Two numerical simulations were carried out to construct datasets. We first synthesized the absorption coefficient  $\mu_a$  and then calculated the photon fluence  $\Phi$  and initial pressure  $P_0$  using the Monte Carlo method [61,62]. The Grüneisen parameter  $\Gamma$  was considered to be spatially constant and equal to 1, so that  $P_0 = \mu_a \times \Phi$ . We assumed symmetric, vertically incident illumination with four laser outputs in the form of a line in order to simulate wide-field light. In the first simulation, square phantoms consisting of several rectangular and circular absorbers with  $\mu_a$  ranging from 0.01 to 0.2  $\text{mm}^{-1}$  were considered. The absorbers in each sample were the same size but located in different positions. Additional details of the simulation parameters are described elsewhere [54]. The second simulation involved circular phantoms comprising four tubular vessel structures with  $\mu_a$  ranging from 0.01 to 0.3  $\text{mm}^{-1}$ . The radii of inner absorbers ranged from 1 to 3 mm, while the positions were set randomly. In both simulations, we chose a constant reduced scattering coefficient of 1  $\text{mm}^{-1}$  and anisotropy parameter  $g$  of 0.9 for the background and absorbers, while  $\mu_a$  of the background was set to 0.01  $\text{mm}^{-1}$ . The circular and square datasets in QOAT-Net respectively contained 3800 pairs of training samples, of which 80% were used as training datasets, 10% as validating datasets, and 10% as test datasets.

More sophisticated simulations were carried out using digital models of mouse and human brains [56–60]. These digital models were partitioned according to anatomical structures. Different sections of the models were imaged to obtain 400 images. We then carried out data augmentation by rotating each image by 45, 90, 135, 180, 225, 270, and 315 degrees clockwise to increase the size of the datasets by a factor of 8. In total, the datasets of the digital mouse and human brains respectively contained 3200 pairs of training samples, of which 80% were used as training datasets, 10% as validating datasets, and 10% as test datasets. After training the dual-path network, we blindly tested its inference by feeding it  $P_0$  images that did not overlap with the images used in the training datasets.

#### B. Phantom Datasets

Cylinder-shaped agar phantoms were constructed with a radius of 12.5 mm and height of 60 mm, inside of which four absorbers were embedded with different shapes and positions intended to mimic mouse organs. Low-melting-point agar [2.8% (m/v); congealing temperature 26°C–30°C; catalog no. A9414, Sigma] was used. To create optical scattering and absorption, different concentrations of intralipid (10%) and India ink were added to the agar [70]. To simulate tissue, background  $\mu_a$  was set as 0.01  $\text{mm}^{-1}$ , absorber  $\mu_a$  ranged from 0.015 to 0.04  $\text{mm}^{-1}$ , and the reduced scattering coefficient throughout the domain was defined as 1  $\text{mm}^{-1}$  [38]. Minor



**Fig. 3.** Implementation and validation of unsupervised data translation from the simulation to experiment domain to provide generated experiment data with the ground truth. (a) The principle of optoacoustic tomography (OAT) and the generation of experiment data. The OAT system obtains real experiment pressures  $P_0^E$ . Monte Carlo (MC) simulation produces simulation pressures  $P_0^S$  from synthesized  $\mu_a$ , then the trained generator  $G_{S \rightarrow E}$  produces generated experiment pressures  $P_0^G$  from  $P_0^S$ . BN, batch normalization; Conv, convolution; LReLU, leaky rectified linear unit. (b) Image results and distribution of image intensity with phantoms to verify unsupervised simulation-to-experiment data translation. The  $x$  axis of the histogram shows image intensity, while the  $y$  axis indicates the proportion of total intensity found at the given position on the  $x$  axis. Scale bar, 5 mm.

errors in the optical coefficients of the phantoms may occur during their production, which have little influence on the experimental results.

A total of 22 images were acquired from 22 experimental phantoms embedding targets with different shapes (Fig. S4 in Supplement 1), four of which served as test datasets, while the remainder were rotated by 360 deg in 2-degree steps to obtain 180 images from each original. Except for the test data, a total of 3240 OAT images were obtained, of which 90% were used as the training datasets and the rest were used as the validating datasets. All these images were combined with the abovementioned simulation datasets  $P_0^S$  to train the SEED-Net, giving the so-called generated experimental datasets  $P_0^G$ .

### C. Ex Vivo and In Vivo Tissue Datasets

All animal procedures in this study were reviewed and approved by the Subcommittee on Research Animal Care at Tianjin Medical University Cancer Institute & Hospital. Fresh porcine tenderloin containing two small pieces of porcine liver of higher  $\mu_a$  were fixed within agar. The tenderloin, approximately 1.5 cm thick, was defined as background. In other experiments, liver and kidney from mice were fixed in agar and imaged, or a 4-week-old healthy male KM mouse ( $\sim 20$  g) was imaged after anesthesia with 10% chloral hydrate. In the live-mouse imaging studies, cross-sectional thoracic and abdominal OAT images were sequentially acquired. Images of tenderloin, mouse liver or kidney, or mouse thorax and abdomen were reconstructed after training the dual-path network with the generated experiment mouse datasets. We used 3200 and 3800 generated data-label pairs to train the networks for the *ex vivo* and *in vivo* experiments, respectively.

### D. OAT System

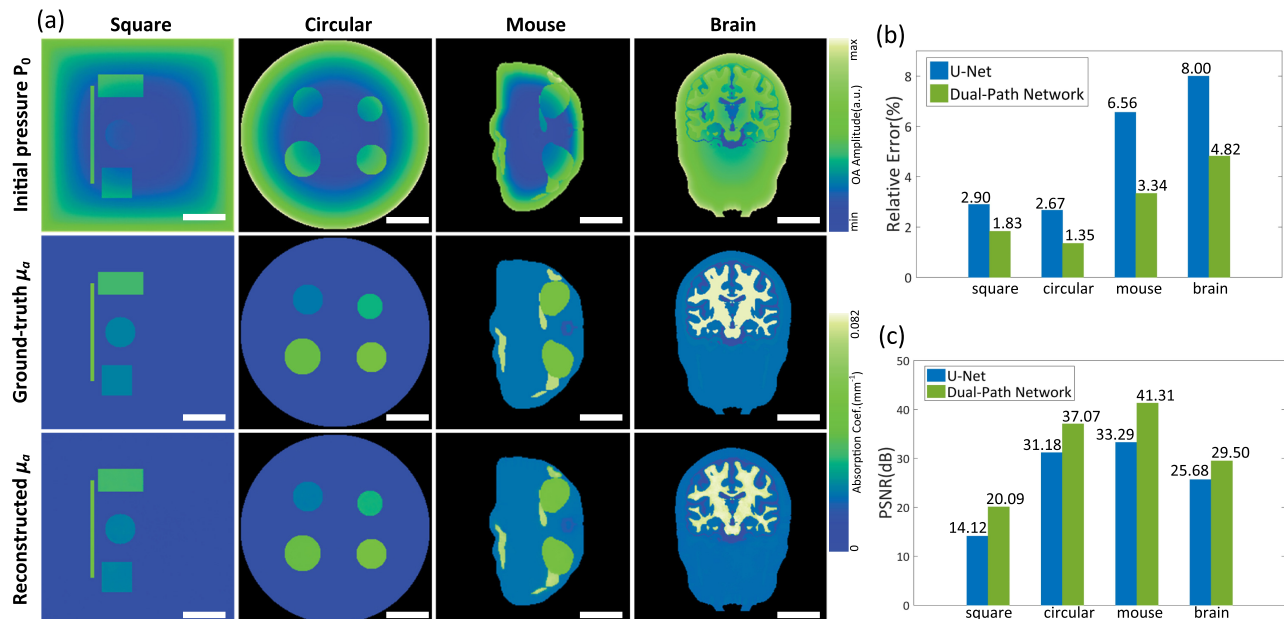
A self-built OAT system was used to image phantoms, as well as *ex vivo* and *in vivo* tissues [44]. A pulsed Nd:YAG laser followed

by a tunable optical parametric oscillator laser generated 10 ns pulses at a repetition rate of 10 Hz. The excitation light at 705 nm was guided into a custom-made 540-fiber bundle and split into four arms [Fig. 3(a)] to provide wide-field illumination. The light was delivered to the imaging plane with an incident energy density below the ANSI limit of 20 mJ/cm<sup>2</sup>. Acoustic signals were detected using a cylindrical-focused ultrasound transducer with the central frequency of 3.5 MHz, the focal length of 80 mm, and the diameter of 25 mm. The signals were amplified using a 50 dB amplifier and digitized using a data acquisition (DAQ) card. The scanning radius of the transducer was set to 80 mm to match the focal length. In order to ensure constant speed of sound, especially for *in vivo* experiments, the heating elements were placed in a water tank maintained at 33°C. Two-dimensional optoacoustic images were reconstructed using a universal backprojection algorithm [71]. More details of this OAT system can be found in Supplement 1.

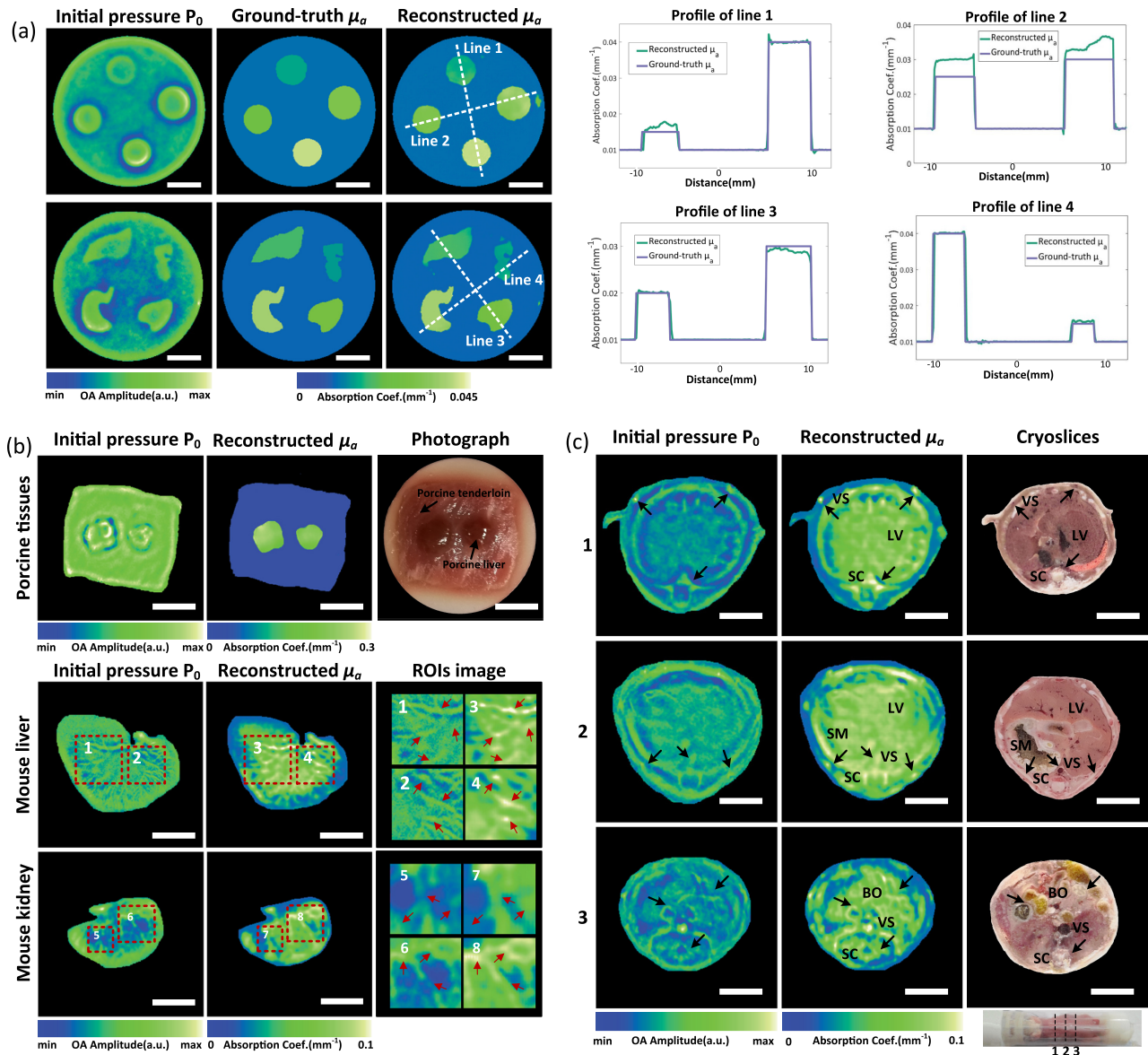
## 4. RESULTS

### A. SEED-Net Evaluation

Phantom experiments were conducted to demonstrate the validity of the SEED-Net [Fig. 1(a)] by performing simulation-to-experiment data translation after training. Real experiment  $P_0^E$  of a phantom with known  $\mu_a$  was reconstructed from optoacoustic signals obtained using a custom-built OAT system (see Section 2.D) [Fig. 3(a)]. According to the digitized  $\mu_a$  derived from the phantom, we calculated  $P_0^S$  using the Monte Carlo method [Fig. 3(a)], then translated  $P_0^S$  into the experiment domain, generating  $P_0^G$  by  $G_{S \rightarrow E}$  (see Section 2.B for architecture details). The distribution of internal brightness differed between  $P_0^G$  and  $P_0^S$ , and similar features were observed between  $P_0^G$  and  $P_0^E$ . The probability distributions of image intensity showed correspondence between  $P_0^G$  and  $P_0^E$ , which are different from  $P_0^S$  [Fig. 3(b)]. These results



**Fig. 4.** Performance of the dual-path network on simulation samples. (a) Reconstructed results for four simulation samples. (b) Comparison of relative error between the dual-path network and conventional U-Net. (c) Comparison of peak signal-to-noise ratio (PSNR) between the dual-path network and U-Net. Scale bar, 5 mm.



**Fig. 5.** Performance of the QOAT-Net on biomimicking phantoms as well as *ex vivo* and *in vivo* samples. (a) Results of phantom experiments. Reconstructed  $\mu_a$  refers to absorption coefficients reconstructed by the QOAT-Net after training only with generated experimental data. The *right* panels show reconstructed  $\mu_a$  along the dotted white lines. (b) Reconstructed images of *ex vivo* porcine tissue, mouse liver, and mouse kidney. (c) Reconstructed images of three thoracic and abdominal cross sections of an anesthetized mouse taken at positions 1, 2, and 3, shown beneath a photograph of the corresponding cryoslice. Arrows identify equivalent positions across each row to highlight differences between the initial pressure and absorption coefficient. BO, bowel; LV, liver; SC, spinal cord; SM, stomach; VS, vessel. Scale bar, 5 mm.

indicated that the proposed data translation network was able to generate experimental data resembling actual experimental data.

### B. Dual-Path Network Evaluation

The top and bottom paths learned to reconstruct  $\Phi$  and  $\mu_a$ , respectively, such that feeding  $P_0$  into the dual-path network generated  $\mu_a$  (Fig. 2). The relationship between  $\Phi$  and  $\mu_a$  was preserved by forcing the deviation between input  $P_0$  and reconstructed  $P_0$  to be as small as possible (see Section 2.C for architecture details). Four simulation samples (square, circle, digital mouse, and digital brain) were chosen to evaluate the proposed dual-path network [Fig. 4(a)]. Large differences were observed between  $P_0$  and ground truth  $\mu_a$  due to heterogeneity in  $\Phi$ . However, the reconstructed

images of  $\mu_a$  determined by the dual-path network were quite consistent with the corresponding ground truth images. In fact, the proposed dual-path network led to reconstructions that showed at least 36% lower relative error and at least 15% higher peak signal-to-noise ratio than the conventional U-Net method [54] [Figs. 4(b) and 4(c)].

### C. QOAT-Net Evaluation

We demonstrated good application performance of our QOAT-Net [Fig. 1(b)] in reconstructing inner targets of a phantom mimicking different organs of the mouse when contrast was low, i.e., when the  $\mu_a$  of the target was close to that of the background [Fig. 5(a)]. When we compared the reconstructed  $\mu_a$  obtained



with or without real experiment  $P_0^E$ , we found that the best results were obtained with real experiment input (Fig. S5 in Supplement 1). Reconstruction using a network trained only with simulation datasets gave unacceptably large errors. In contrast, a training strategy involving only generated  $P_0^G$ , though potentially less accurate than a strategy based on both generated  $P_0^G$  and real experiment  $P_0^E$ , may still be reliable enough for practical applications, and may be the best choice when real labeled experiment datasets cannot be obtained. Moreover, QOAT-Net based on generated  $P_0^G$  can reconstruct various types of experimental datasets for a specific OAT system because training data generated from simulation datasets is flexible and easy to extend. Figure S6 in Supplement 1 illustrates the ability to process different data types of the QOAT-Net. In addition, comparison between the QOAT-Net and the latest non-linear iterative perturbation method [72] has been displayed in Fig. S7 in Supplement 1, showing better performance of the QOAT-Net in the noise suppression and spatial resolution. The time required by various algorithms to reconstruct  $\mu_a$  is listed in Table S2 in Supplement 1.

Encouraged by these phantom results, we compared conventional OAT results ( $P_0$ ) and QOAT results (reconstructed  $\mu_a$ ) for imaging porcine liver and tenderloin *ex vivo* [Fig. 5(b), top row]. OAT led to bright regions at the boundaries between target tissue and background, indicating poor distinction between areas with different  $\mu_a$ . QOAT, in contrast, led to starker target–background distinction, and each target was internally more homogeneous than with OAT. QOAT gave reconstructed  $\mu_a$  of  $0.2 \text{ mm}^{-1}$  for liver tissues and  $0.01 \text{ mm}^{-1}$  for tenderloin, only slightly lower than in previous reports [59,73].

Similar superiority for QOAT over OAT was observed when imaging mouse liver and kidney *ex vivo* [Fig. 5(b), middle and lower rows]. With OAT, inner vessels could not be easily differentiated from one another, but they did appear different from tissues near the boundaries. With QOAT, inner vessels could be distinguished more easily from one another. With OAT, blood vessels of the same size had different intensities, whereas with QOAT they appeared more homogeneous and more distinguishable from background.

Finally, we compared OAT and QOAT in the more complex context of an entire mouse, as a more demanding test of the (pre)clinical potential of our proposed QOAT-Net [Fig. 5(c)]. OAT provided inaccurate absorption information in three cross sections with rich anatomical content due to heterogeneous photon density. For example, image intensities in the liver regions were lower than those of external skin tissues, and some vessel regions were similarly bright as surrounding tissues. QOAT provided obviously better absorption information: vessels and organs appeared different from other tissues, and various anatomical structures appeared similar to the images from the digital mouse. For example, QOAT reconstructed  $\mu_a$  of  $0.065 - 0.075 \text{ mm}^{-1}$  in the liver area, consistent with the  $0.072 \text{ mm}^{-1}$  in the digital mouse [59].

## 5. DISCUSSION AND CONCLUSION

To extend OAT to high-resolution quantitative imaging, QOAT was developed to reconstruct  $\mu_a$  values that are accurately related to the physiological characteristics of tissues. However, conventional QOAT relies on complex and time-consuming iterative calculations that require extensive computational resources, and the final fidelity and spatial resolution depend strongly on experience-based selection of optimization parameters. DL may

be a more efficient alternative but requires appropriate data-label pairs for accurate reconstruction of tissue imaging. To avoid this requirement, we have developed a dual-path QOAT network with unsupervised data translation from simulation to experiment domains. This QOAT-Net can reconstruct  $\mu_a$  with relative errors below 10% in less than one second, approaching the real-time imaging requirements for many biomedical and preclinical applications. The comparison between QOAT-Net and the iterative method demonstrates that QOAT-Net can solve the optical inverse problem without the reference sample and the trade-off between accuracy and spatial resolution, which are typical of optimization algorithms. To the best of our knowledge, this is the first demonstration that DL-based QOAT can accurately estimate high-resolution images of  $\mu_a$  in deep tissues.

The strong performance of QOAT-Net is due in large measure to the three loss functions of absorption coefficient, photon fluence, and initial pressure in the dual-path network. These three loss functions cover all three variables in the OAT mathematical model, allowing QOAT-Net to function more accurately. The strong performance can also be attributed to the unsupervised simulation-to-experiment data translation, in which the SEED-Net is trained with simulation data and typical experimental results for the specific QOAT system being used, effectively generating experimental data-label pairs to substitute for actual manual labeling.

The framework we suggest here may be feasible for image reconstruction using various OAT systems and types of tissues. Combining this approach with a multispectral optoacoustic tomography system may further reconstruct the physiological and pathological parameters of deep tissue. In this way, the framework may compensate for the lack of labeled datasets that hinders DL in various types of imaging, such as diffuse optical tomography and fluorescence molecular tomography. Our approach can cover a wider range of biomedical imaging tasks by introducing more complete data models.

**Funding.** National Natural Science Foundation of China (81771880, 82171989, 681871393, 62075156, 61971303, 61901342); Natural Science Foundation of Tianjin Municipal Science and Technology Commission (19JCQNJC12800); Key Fund of Shenzhen Natural Science Foundation (grant No. JCYJ20200109150212515); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. (694968) (PREMSOT).

**Acknowledgment.** The authors would like to thank NVIDIA Corporation for donation of the Titan Xp GPU used for this research. We thank Dr. Yihan Wang, from School of Life Science and Technology, Xidian University, for his help with performance comparison of different methods. We thank Dr. Robert J. Wilson for consultation on the manuscript. J. L. and B. S. conceived the project. J. L., C. W., and T. C. implemented the image reconstruction, processed algorithms, and analyzed the data. T. C., C. W., and S. L. contributed to experiments. T. L. and F. G. provided modeling tools. J. L., B. S., C. W., and T. C. wrote the paper. F. G. and V. N. supervised the research and edited the paper. All the authors discussed the results and contributed to the writing of the paper.

**Disclosures.** The authors declare no financial or commercial conflicts of interest.

**Data availability.** The main data supporting the results in this study are available within the main text and Supplement 1. The raw and training datasets generated during the study are too large to be publicly shared, yet they are available for research purposes from the corresponding authors on reasonable request.

**Code availability.** We are glad to share codes involved in our work with other colleagues, since the presented results may be additionally promising for other emerging approaches. Codes, models, all testing data, and software instructions in the paper may be downloaded from “[74]”. Part of the data will be selectively



released due to copyright issues but can be also available from the corresponding author upon reasonable request.

**Supplemental document.** See Supplement 1 for supporting content.

## REFERENCES

1. A. Taruttis and V. Ntziachristos, "Advances in real-time multispectral photoacoustic imaging and its applications," *Nat. Photonics* **9**, 219–227 (2015).
2. J. Li, A. Chekkoury, J. Prakash, S. Glasl, P. Vetschera, B. Koberstein-Schwarz, I. Olefir, V. Gujrati, M. Omar, and V. Ntziachristos, "Spatial heterogeneity of oxygenation and haemodynamics in breast cancer resolved in vivo by conical multispectral photoacoustic mesoscopy," *Light Sci. Appl.* **9**, 57 (2020).
3. K. Haedicke, L. Agemy, M. Omar, A. Berezhnoi, S. Roberts, C. Longo-Machado, M. Skubal, K. Nagar, H.-T. Hsu, K. Kim, T. Reiner, J. Coleman, V. Ntziachristos, A. Scherz, and J. Grimm, "High-resolution photoacoustic imaging of tissue responses to vascular-targeted therapies," *Nat. Biomed. Eng.* **4**, 286–297 (2020).
4. L. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," *Nat. Methods* **13**, 627–638 (2016).
5. M. Omar, J. Aguirre, and V. Ntziachristos, "Photoacoustic mesoscopy for biomedicine," *Nat. Biomed. Eng.* **3**, 354–370 (2019).
6. X. Deán-Ben, T. Fehm, S. J. Ford, S. Gottschalk, and D. Razansky, "Spiral volumetric photoacoustic tomography visualizes multi-scale dynamics in mice," *Light Sci. Appl.* **6**, e16247 (2017).
7. X. Deán-Ben, G. Sela, A. Lauri, M. Kneipp, V. Ntziachristos, G. G. Westmeyer, S. Shoham, and D. Razansky, "Functional photoacoustic neuro-tomography for scalable whole-brain monitoring of calcium indicators," *Light Sci. Appl.* **5**, e16201 (2016).
8. M. P. Fronheiser, S. A. Ermilov, H.-P. Brecht, A. Conjusteau, R. Su, K. Mehta, and A. A. Oraevsky, "Real-time photoacoustic monitoring and three-dimensional mapping of a human arm vasculature," *J. Biomed. Opt.* **15**, 021305 (2010).
9. R. Li, P. Wang, L. Lan, F. P. Lloyd, C. J. Goergen, S. Chen, and J.-X. Cheng, "Assessing breast tumor margin by multispectral photoacoustic tomography," *Biomed. Opt. Express* **6**, 1273–1281 (2015).
10. E. Z. Zhang, J. G. Laufer, R. B. Pedley, and P. C. Beard, "In vivo high-resolution 3D photoacoustic imaging of superficial vascular anatomy," *Phys. Med. Biol.* **54**, 1035–1046 (2009).
11. T. P. Matthews and M. A. Anastasio, "Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements," *Inverse Probl.* **33**, 124002 (2017).
12. G. Xu, J. R. Rajian, G. Girish, M. J. Kaplan, J. B. Fowlkes, P. L. Carson, and X. Wang, "Photoacoustic and ultrasound dual-modality imaging of human peripheral joints," *J. Biomed. Opt.* **18**, 010502 (2012).
13. J. Weber, P. Beard, and S. Bohndiek, "Contrast agents for molecular photoacoustic imaging," *Nat. Methods* **13**, 639–650 (2016).
14. L. Yiu, L. Nie, and X. Chen, "Photoacoustic molecular imaging: from multiscale biomedical applications towards early-stage theranostics," *Trends Biotechnol.* **34**, 420–433 (2016).
15. W. Lu, Q. Huang, G. Ku, X. Wen, M. Zhou, D. Guzatov, P. Brecht, R. Su, A. Oraevsky, L. V. Wang, and C. Li, "Photoacoustic imaging of living mouse brain vasculature using hollow gold nanospheres," *Biomaterials* **31**, 2617–2626 (2010).
16. M. Schwarz, A. Buehler, J. Aguirre, and V. Ntziachristos, "Three-dimensional multispectral photoacoustic mesoscopy reveals melanin and blood oxygenation in human skin in vivo," *J. Biophotonics* **9**, 55–60 (2016).
17. B. T. Cox, S. R. Arridge, and P. C. Beard, "Estimating chromophore distributions from multiwavelength photoacoustic images," *J. Opt. Soc. Am. A* **26**, 443–455 (2009).
18. S. Tzoumas, A. Nunes, I. Olefir, S. Stangl, P. Symvoulidis, S. Glasl, C. Bayer, G. Multhoff, and V. Ntziachristos, "Eigen spectra photoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues," *Nat. Commun.* **7**, 12121 (2016).
19. L. Nie and X. Chen, "Structural and functional photoacoustic molecular tomography aided by emerging contrast agents," *Chem. Soc. Rev.* **43**, 7132–7170 (2014).
20. Z. Yuan and H. Jiang, "Quantitative photoacoustic tomography," *Philos. Trans. A* **367**, 3043–3054 (2009).
21. B. T. Cox, J. G. Laufer, P. C. Beard, and S. R. Arridge, "Quantitative spectroscopic photoacoustic imaging: a review," *J. Biomed. Opt.* **17**, 061202 (2012).
22. B. T. Cox, J. G. Laufer, and P. C. Beard, "The challenges for quantitative photoacoustic imaging," *Proc. SPIE* **7177**, 717713 (2009).
23. B. T. Cox, J. G. Laufer, and P. C. Beard, "Quantitative photoacoustic image reconstruction using fluence dependent chromophores," *Biomed. Opt. Express* **1**, 201–208 (2010).
24. F. M. Brochu, J. Brunker, J. Joseph, M. R. Tomaszewski, S. Morscher, and S. E. Bohndiek, "Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in photoacoustic tomography," *IEEE Trans. Med. Imaging* **36**, 322–331 (2017).
25. Y. Wang, J. Li, T. Lu, L. Zhang, Z. Zhou, H. Zhao, and F. Gao, "Combined diffuse optical tomography and photoacoustic tomography for enhanced functional imaging of small animals: a methodological study on phantoms," *Appl. Opt.* **56**, 303–311 (2017).
26. A. Q. Bauer, R. E. Nothdurft, J. P. Culver, T. N. Erpelding, and L. V. Wang, "Quantitative photoacoustic imaging correcting for heterogeneous light fluence distributions using diffuse optical tomography," *J. Biomed. Opt.* **16**, 096016 (2011).
27. A. Hussain, E. Hondebrink, J. Staley, and W. Steenbergen, "Photoacoustic and acousto-optic tomography for functional and quantitative imaging," *Optica* **5**, 1579–1589 (2018).
28. B. T. Cox, S. R. Arridge, K. P. Köstli, and P. C. Beard, "Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method," *Appl. Opt.* **45**, 1866–1875 (2006).
29. B. T. Cox, S. R. Arridge, K. P. Kostli, and P. C. Beard, "Quantitative photoacoustic imaging: fitting a model of light transport to the initial pressure distribution," *Proc. SPIE* **5697**, 49–55 (2005).
30. T. Tarvainen, B. T. Cox, J. P. Kaipio, and S. R. Arridge, "Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography," *Inverse Probl.* **28**, 84009 (2012).
31. T. Jetzfellner, D. Razansky, A. Rosenthal, R. Schulz, K.-H. Englmeier, and V. Ntziachristos, "Performance of iterative photoacoustic tomography with experimental data," *Appl. Phys. Lett.* **95**, 013703 (2009).
32. Z. Yuan and H. Jiang, "Quantitative photoacoustic tomography: recovery of optical absorption coefficient maps of heterogeneous media," *Appl. Phys. Lett.* **88**, 231101 (2006).
33. Z. Yuan, Q. Wang, and H. Jiang, "Reconstruction of optical absorption coefficient maps of heterogeneous media by photoacoustic tomography coupled with diffusion equation based regularized Newton method," *Opt. Express* **15**, 18076–18081 (2007).
34. L. Yao, Y. Sun, and H. Jiang, "Quantitative photoacoustic tomography based on the radiative transfer equation," *Opt. Lett.* **34**, 1765–1767 (2009).
35. P. Shao, B. Cox, and R. J. Zemp, "Estimating optical absorption, scattering, and Grueneisen distributions with multiple-illumination photoacoustic tomography," *Appl. Opt.* **50**, 3145–3154 (2011).
36. S. Li, B. Montcel, Z. Yuan, W. Liu, and D. Vray, "Multigrid-based reconstruction algorithm for quantitative photoacoustic tomography," *Biomed. Opt. Express* **6**, 2424–2434 (2015).
37. Y. Sun, E. Sobel, and H. Jiang, "Quantitative three-dimensional photoacoustic tomography of the finger joints: an in vivo study," *J. Biomed. Opt.* **14**, 064002 (2009).
38. Y. Wang, J. He, J. Li, T. Lu, Y. Li, W. Ma, L. Zhang, Z. Zhou, H. Zhao, and F. Gao, "Toward whole-body quantitative photoacoustic tomography of small-animals with multi-angle light-sheet illuminations," *Biomed. Opt. Express* **8**, 3778–3795 (2017).
39. D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, and L. Maier-Hein, "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," *Proc. SPIE* **10494**, 104942S (2018).
40. J. Schwab, S. Antholzer, R. Nuster, and M. Haltmeier, "DALnet: high-resolution photoacoustic projection imaging using deep learning," arXiv:1801.06693 (2018).
41. H. Lan, D. Jiang, C. Yang, F. Gao, and F. Gao, "Y-Net: hybrid deep learning image reconstruction for photoacoustic tomography in vivo," *Photoacoustics* **20**, 100197 (2020).
42. A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, "Model-based learning for accelerated, limited view 3-D photoacoustic tomography," *IEEE Trans. Med. Imaging* **37**, 1382–1393 (2018).

43. T. Vu, M. Li, H. Humayun, Y. Zhou, and J. Yao, "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," *Exp. Biol. Med. (Maywood)* **245**, 597–605 (2020).
44. T. Lu, T. Chen, F. Gao, B. Sun, V. Ntziachristos, and J. Li, "LV-GAN: a deep learning approach for limited-view photoacoustic imaging based on hybrid datasets," *J. Biophotonics* **14**, e202000325 (2020).
45. N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning photoacoustic tomography with sparse data," *Nat. Mach. Intell.* **1**, 453–460 (2019).
46. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse Probl. Sci. Eng.* **27**, 987–1005 (2018).
47. S. Guan, A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health Inform.* **24**, 568–576 (2019).
48. T. Tong, W.-H. Huang, K. Wang, Z. He, L. Yin, X. Yang, S. Zhang, and J. Tian, "Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data," *Photoacoustics* **19**, 100190 (2020).
49. S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning," *Sci. Rep.* **10**, 8510 (2020).
50. T. Kirchner, J. Gröhl, and L. Maier-Hein, "Context encoding enables machine learning-based quantitative photoacoustics," *J. Biomed. Opt.* **23**, 056008 (2018).
51. J. Gröhl, T. Kirchner, T. Adler, and L. Maier-Hein, "Confidence estimation for machine learning-based quantitative photoacoustics," *J. Imaging* **4**, 147 (2018).
52. C. Cai, K. Deng, C. Ma, and J. Luo, "End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging," *Opt. Lett.* **43**, 2752–2755 (2018).
53. J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, S.-I. Lee, E. Y. Chae, H. H. Kim, Y. M. Bae, Y.-W. Choi, S. Cho, and J. C. Ye, "Deep learning diffuse optical tomography," *IEEE Trans. Med. Imaging* **39**, 877–887 (2020).
54. T. Chen, T. Lu, S. Song, S. Miao, F. Gao, and J. Li, "A deep learning method based on U-Net for quantitative photoacoustic imaging," *Proc. SPIE* **11240**, 112403V (2020).
55. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 2223–2232.
56. Q. Fang, A. P. Tran, and S. Yan, "Improving model-based fNIRS analysis using mesh-based anatomical and light-transport models," *Neurophotonics* **7**, 015008 (2020).
57. W. F. Cheong and S. A. Prael, "A review of the optical properties of biological tissues," *IEEE J. Quantum Electron.* **26**, 2166–2185 (1990).
58. S. L. Jacques, "Corrigendum: optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**, 5007–5008 (2013).
59. Q. Fang and D. R. Kaeli, "Accelerating mesh-based Monte Carlo method on modern CPU architectures," *Biomed. Opt. Express* **3**, 3223–3230 (2012).
60. M. Allard, D. Cote, L. Davidson, J. Dazai, and R. M. Henkelman, "Combined magnetic resonance and bioluminescence imaging of live mice," *J. Biomed. Opt.* **12**, 034018 (2007).
61. T. Lu, J. Li, T. Chen, S. Li, X. Xu, and F. Gao, "Parallelized Monte Carlo photon transport simulations for arbitrary multi-angle wide-field illumination in photoacoustic imaging," *Front. Phys.* **8**, 283 (2020).
62. Q. Fang, "Monte Carlo eXtreme software," <http://mcx.sourceforge.net>.
63. I. Goodfellow, *Advances in Neural Information Processing Systems*, Z. Ghahramani, ed. (Curran Associates, Inc., 2014), Vol. **27**, pp. 2672–2680.
64. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Advances in Neural Information Processing Systems* (2018), pp. 2483–2493.
65. A. L. Maas, Y. H. Awari, and Y. Andrew, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML* (2013), Vol. **30**, pp. 3.
66. S. Gross and M. Wilber, "Training and investigating residual nets," 2016, <http://torch.ch/blog/2016/02/04/resnets.html>.
67. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 1125–1134.
68. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. (Springer, Cham., 2015), p. 231.
69. D. P. K. Adam and J. Ba, "A method for stochastic optimization," arXiv:1412.6980 (2014).
70. R. Michels, F. Foschum, and A. Kienle, "Optical properties of fat emulsions," *Opt. Express* **16**, 5907–5925 (2008).
71. M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Phys. Rev. E* **71**, 016706 (2005).
72. Y. H. Wang, M. L. Xu, F. Gao, F. Kang, and S. Zhu, "Nonlinear iterative perturbation scheme with simplified spherical harmonics (SP<sub>3</sub>) light propagation model for quantitative photoacoustic tomography," *J. Biophotonics* **14**, e202000446 (2021).
73. J. P. Ritz, A. Roggan, C. Isbert, G. Müller, H. J. Buhr, and C. T. Germer, "Optical properties of native and coagulated porcine liver tissue between 400 and 2400 nm," *Laser Surg. Med.* **29**, 205–212 (2001).
74. J. Li, "QOAT-Net," Github (2021) <https://github.com/jiaolitu/QOAT-Net>.