Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification

Rongrong Liu,[1] Shiyi Cheng,[2] Lei Tian,[2] Ji Yi[3,4,*]

[1]*Department of Biomedical Engineering, Northwestern University, Evanston, IL, 60208, USA*
[2]*Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA*
[3]*Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston MA, 02118, USA*
[4]*Department of Biomedical Engineering, Boston University, Boston MA, 02215, USA*
**Corresponding author: jiyi@bu.edu*

Abstract: Measurement of blood oxygen saturation (sO_2) by optical imaging oximetry provides invaluable insight into local tissue functions and metabolism. Despite different embodiments and modalities, all label-free optical imaging oximetry utilize the same principle of sO_2-dependent spectral contrast from hemoglobin. Traditional approaches for quantifying sO_2 often rely on analytical models that are fitted by the spectral measurements. These approaches in practice suffer from uncertainties due to biological variability, tissue geometry, light scattering, systemic spectral bias, and variations in experimental conditions. Here, we propose a new data-driven approach, termed deep spectral learning (DSL) for oximetry to be highly robust to experimental variations, and more importantly to provide uncertainty quantification for each sO_2 prediction. To demonstrate the robustness and generalizability of DSL, we analyze data from two visible light optical coherence tomography (vis-OCT) setups across two separate *in vivo* experiments in rat retina. Predictions made by DSL are highly adaptive to experimental variabilities as well as the depth-dependent backscattering spectra. Two neural-network-based models are tested and compared with the traditional least-squares fitting (LSF) method. The DSL-predicted sO_2 shows significantly lower mean-square errors than the LSF. For the first time, we have demonstrated *en face* maps of retinal oximetry along with pixel-wise confidence assessment. Our DSL overcomes several limitations in the traditional approaches and provides a more flexible, robust, and reliable deep learning approach for *in vivo* non-invasive label-free optical oximetry.

1. Introduction.

Microvascular systems deliver oxygen to support cellular metabolism and maintain biological functions. Within the local microenvironment of blood vessels, oxygen unloads from hemoglobin and diffuses freely from red blood cells (RBC) to tissues following the gradient of oxygen partial pressure (pO_2), which determines the oxygen saturation (sO_2) of hemoglobin. The measurement of microvascular sO_2 thus can provide invaluable insight into local tissue metabolism, inflammation, and oxygen-related pathologies (e.g. cancers, diabetic milieu and complications, and cardiovascular diseases) [1–5].

In recent years, several non-invasive and label-free optical imaging oximetry techniques have been developed to measure microvascular sO_2. Despite their differences, the fundamental mechanism is the same that is based on the sO_2-dependent spectral contrast from hemoglobin [6]. The spectral measurement is then related to sO_2 through a complex physical model incorporating tissue geometry, heterogeneous tissue scattering, light attenuation and propagation, and imaging optical instruments. This model is often simplified and analytically formulated under different approximations and assumptions. Examples include spatial frequency domain imaging [7,8] in the diffusive regime under the P3 approximation, multi-wavelength imaging [9–12]and visible light optical coherence tomography (vis-OCT) [13–26] in the ballistic regime based on the Beer's law combined with the first Born approximation, photoacoustic microscopy/tomography assuming a uniform laser fluence inside the tissue [27,28], and photothermal imaging assuming a linear relation between the blood absorption and the change of the optical signal [29–32]. The sO_2 estimation thus requires solving an ill-posed inverse problem that is inevitably subject to model inaccuracies, noise, systemic spectral bias, and experimental conditions. One widely used inversion method is the spectral least-squares fitting (LSF) that estimates the sO_2 by matching the spectral data with the analytical model. However, in practice multiple sources of spectral errors exist that are impossible to be fully parameterized in a simple analytical form, which in turn compromises the sO_2 estimation accuracy, repeatability, and

cross-comparison between different devices, test subjects, and time. Therefore, it is imperative to develop a more robust model to enable more accurate quantification of microvascular sO_2 for label-free optical imaging oximetry.

In this work, we develop a new data-driven deep spectral learning (DSL) method to enable highly robust and reliable sO_2 estimation. By training a neural network to directly relate the spectral measurements to the corresponding independent sO_2 labels, DSL bypasses the need for a rigid parametric model, similar to existing deep learning methods for solving optical inverse problems[33–37]. We show that DSL can be trained to be highly robust to multiple sources of variabilities in the experiments, including different devices, imaging protocols, speeds, and other possible longitudinal variations.

A crucial feature of our DSL method is uncertainty quantification. Due to biological variations and tissue heterogeneity, an assessment of the reliability of each sO_2 measurement is crucial in clinical applications and for guarding against vulnerabilities in making overly confident predictions when imaging rare cases[38]. Existing model-based methods generate a single value of sO_2 for each spectral measurement, i.e. a point estimate. The accuracy and uncertainty of the point estimate can only be assessed by taking repeated measurements against the ground-truth in a well-controlled experiment. This uncertainty estimation presents a clear limitation for many biomedical applications in which the ground-truth is often inaccessible *in vivo,* and the statistical analysis can only be performed retrospectively on those repeated measurements [13,19,22,32,39]. Instead of assessing the variabilities in the data retrospectively, we develop our DSL model based on an uncertainty learning framework[37] to encapsulate the statistics in the learned model, essentially shifting the burden of repeated measurements in the model-based methods to the training phase of DSL. After the training, the DSL model predicts both sO_2 and its tandem standard deviation, assessing the uncertainty for each sO_2 prediction (i.e. a statistical distribution describing all the possible sO_2 levels of each prediction given the measurements). Most importantly, we show that the DSL predicted statistics closely match those obtained from ensemble calculations. This means that the confidence level calculated from the DSL prediction can be used as a surrogate estimate to the true accuracy of the estimate – making DSL highly reliable.

We demonstrate DSL using two sets of vis-OCT spectral measurements for oximetry on rat retina from [13,19]. Two DSL models are investigated, including a 1D fully connected neural network (FNN) and a 1D convolutional neural network (CNN). Our results show that both DSL models are significantly outperforming the LSF, both in terms of the estimation accuracy and the robustness to experimental variations. We further conduct quantitative statistical analysis based on the uncertainty learning to establish the confidence level of the two proposed models, and further justify the reliability of DSL. Next, imaging oximetry is demonstrated on *en face* sO_2 maps of rat retina along with the corresponding uncertainty maps, providing a visualization of the DSL predictions. This allows us to assess the accuracy of the prediction based on the underlying physiological conditions.

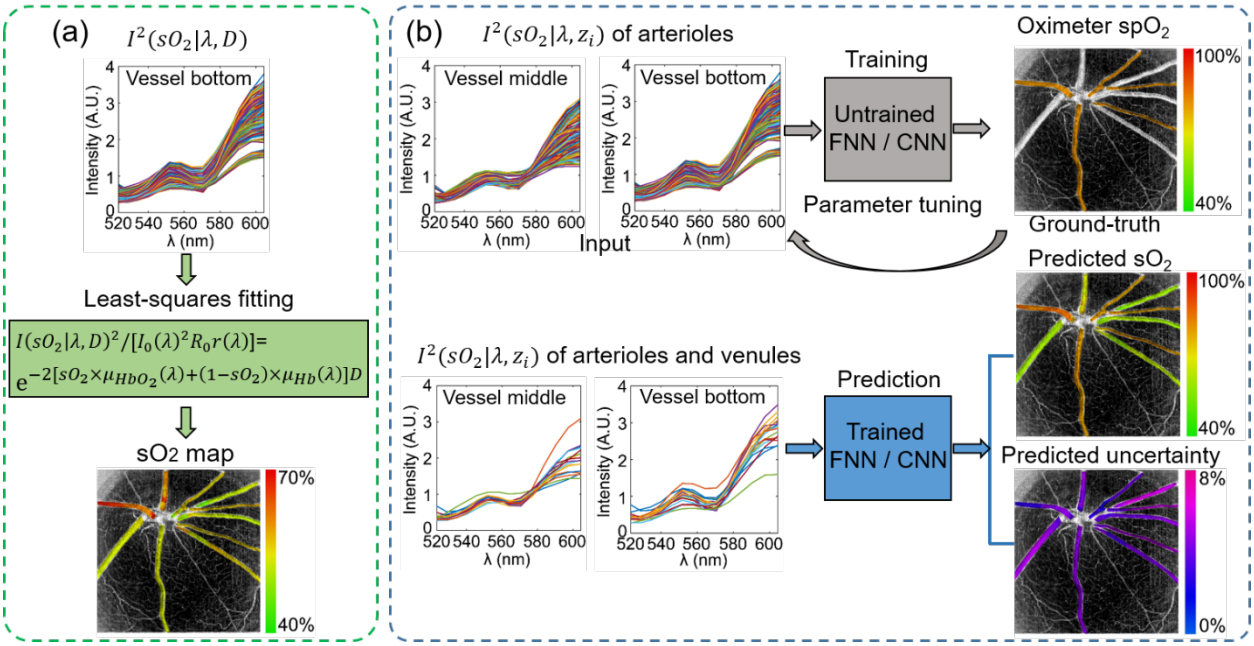

Fig. 1. Methods for calculating retinal blood vessel oxygen saturation (sO_2) by (a) the traditional LSF, and (b) our neural-network-based DSL with uncertainty quantification.

2. Principle of Least Square Fitting

Vis-OCT uses ballistic photon and coherence gating to localize the optical signal within a tissue volume. At the bottom of the vessel wall, light double-passing through the vessel lumen gives rise to the detectable spectral contrast that can be analytically formulated based on the Beer's law [13],

$$I(sO_2|\lambda, z) = I_0(\lambda)\sqrt{R_0 r(\lambda)}e^{-[sO_2\times\mu_{HbO_2}(\lambda)+(1-sO_2)\times\mu_{\mathrm{Hb}}(\lambda)]z} \quad \textbf{(1)}$$

where $I_0(\lambda)$ is the light source's spectrum; R_0 is the reflectance of the reference arm and assumed to be a constant; $r(\lambda)$ (dimensionless) is the reflectance at the vessel wall, whose scattering spectrum can be modeled by a power law under the first Born approximation $r(\lambda)=A\lambda^{-\alpha}$ with A being a dimensionless constant and α modeling the decaying scattering spectrum from the vessel wall; The optical attenuation coefficient μ (mm^{-1}) combines the absorption (μ_a) and scattering coefficients (μ_s) of the whole blood, which are both wavelength- and sO_2-dependent:

$$\mu = \mu_a + W\mu_s \quad \textbf{(2)}$$

where W is a scaling factor for the scattering coefficient and is empirically set to 0.2 [6,21,40]. The subscripts Hb and HbO_2 denote the contribution from the deoxygenated and oxygenated blood, respectively. z denotes the light penetrating depth.

To estimate sO_2, the traditional approach applies the least-squares procedure that fits the vis-OCT spectral measurement to the analytical model by minimizing the total squares of the error, as illustrated in Fig. 1(a):

$$\min_{sO_2,A,\alpha} \sum_\lambda \|\log(I_m(\lambda)) - \log(I(sO_2, A, \alpha|\lambda,\ z = D))\|^2 \quad \textbf{(3)}$$

where I_m is typically taken as the vis-OCT spectral measurements extracted from the bottom of the vessel in order to maximize the spectroscopic contrast.

In this analytical model, two free parameters (A, α) in addition to the unknown sO_2 level are introduced in order to more accurately capture realistic biophysical interactions. However, in practice these two parameters cannot

fully capture all the experimental variabilities. While other models may reduce the free parameters to avoid overfitting [14,17,19,23–25], this approach in general is nonetheless rigid and over-simplified to the actual experiments.

3. Principle of Deep Spectral Learning

In DSL, instead of using a rigid analytical model, we train a neural network to link the spectral measurements and the independently measured sO_2 labels, as illustrated in Fig. 1(b). By doing so, DSL bypasses the need for parametric tuning and model simplification and approximation needed in fitting the analytical models. Furthermore, by removing the restrictions imposed by the analytical model, DSL allows utilizing multiple sets of spectral measurements taken at different depths and enables a more holistic spectral-sO_2 analysis. Specifically, we demonstrate high-quality predictions using concatenated spectra data from both the bottom and the center of the vessels in vis-OCT. Because the pulse oximeter measures the systemic arterial sO_2, the same as the retinal arterioles, we use the retinal arterial spectra as the training input paired with the independently measured pulse oximeter sO_2 (spO_2) as the ground-truth label. After training, the network makes predictions for both arterials and veins.

In addition to sO_2 prediction, a major important feature of our data-driven DSL method is to quantify the uncertainty for *each* prediction. To do so, we specially design the loss function for training the neural network to properly capture the underlying statistics of the data. The commonly used loss function, such as the mean squared error (MSE), assumes a homogeneous Gaussian distribution of the errors in the predictions. This severely limits its ability to adapt to different types of spectral data variations (e.g. spectral signal outliers, non-uniform noise, and unevenly sampled data) that are inevitably *inhomogeneous*. To account for this, we design a customized loss function derived from a *heterogeneous* Gaussian distribution model. Using the training data set (I_i, $[spO_2]_i$), $i = 1,2,\ldots, N$, where I_i and $[spO_2]_i$ are the ith vis-OCT spectral measurement and the ground-truth pulse oximeter spO_2, respectively, our loss function $L_G(w)$ is

$$L_G(w) = \sum_{i=1}^{N} \frac{([sO_2]_i(w) - [spO_2]_i)^2}{\sigma_i^2(w)} + \log\left(\sigma_i^2(w)\right) \quad \textbf{(4)}$$

where $[sO_2]_i$ and σ_i denote respectively the neural network predicted mean and standard deviation of the underlying Gaussian distribution for the ith training data pair; w is the learned neural network weights. The main idea of this loss function assumes that the prediction made on each data follows a distinct Gaussian distribution, and the network is trained to predict the underlying mean and standard deviation [41]. Accordingly, the standard deviation σ quantifies the uncertainty for each sO_2 prediction.

We investigate two neural network models, including an FNN, and a CNN model, whose network architectures are shown in Fig. 2(a) and 2(b), respectively. The detailed descriptions of the data preprocessing and network architecture are included in the method section.

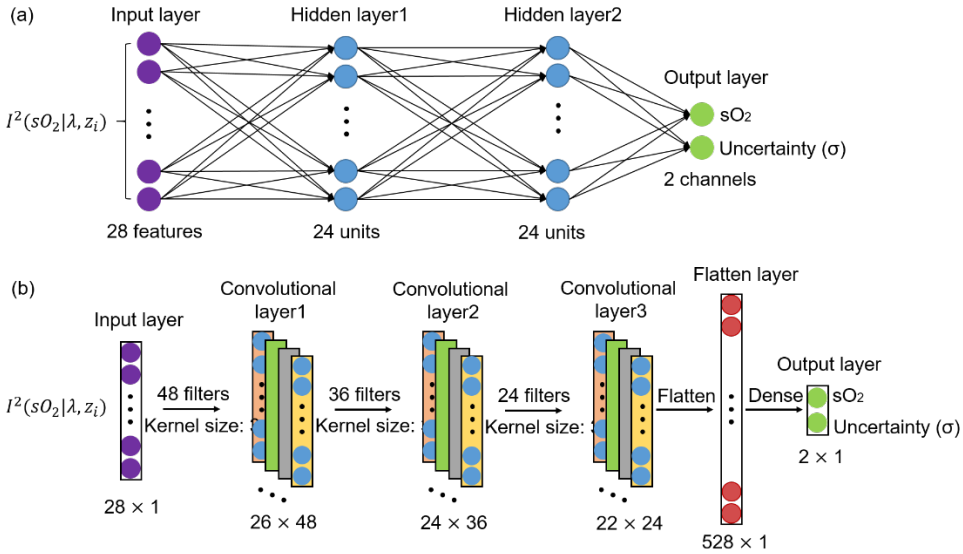

Fig. 2. Structures of the FNN model (a) and the CNN model (b) for sO_2 prediction with uncertainty quantified by the predicted standard deviation.

4. Results and Discussion

A. Data source

To evaluate the effectiveness of the DSL approach, we compiled two datasets from the previous literatures on vis-OCT for retinal oximetry [13,19]. Specifically, the data in Fig. 3(a) are from Ref. [13]and Fig. 3(b) from Ref. [19]. Both data sets used similar experimental protocols that the oxygen content in the ventilation gas were adjusted to induce systemic hypoxia or hyperoxia, and vis-OCT measurements were taken at each ventilation condition. In Ref. [13], a step-wise hypoxia challenge was given to four rats, reducing oxygen content gradually from 21% to 9%. In Ref. [19], eight rats were measured by five ventilation conditions, sequencing from normoxia, hyperoxia (100% O_2), 5% carbon dioxide (21% O_2, 74% N_2, and 5% CO_2), hypoxia (10% O_2, 90% N_2), and finally to normoxia. At each ventilation condition, the systemic arterial spO_2 reading was taken by a pulse oximeter attached to a rear leg of the rats. The spO_2 readings are used as the ground-truth label for the major retinal arterioles for neural network training. Depth-dependent backscattering spectra of rat retinal arterioles in vis-OCT were extracted at each ventilation conditions as spectral input to the neural network. The extracted arteriole spectra with the spO_2 labels were then split into training and testing sets. In Ref. [13], data from three of the four rats were used as the training sets with the remaining one as the testing set. In Ref. [19], data from seven rats were used as the training sets and the remaining one is the testing set. Fig. 3(a) and 3(b) summarize the number of spectra extracted and the corresponding spO_2 labels in descending order, respectively. Fig. 3(c) shows the histogram of all the training and testing spectra with their corresponding spO_2 labels.

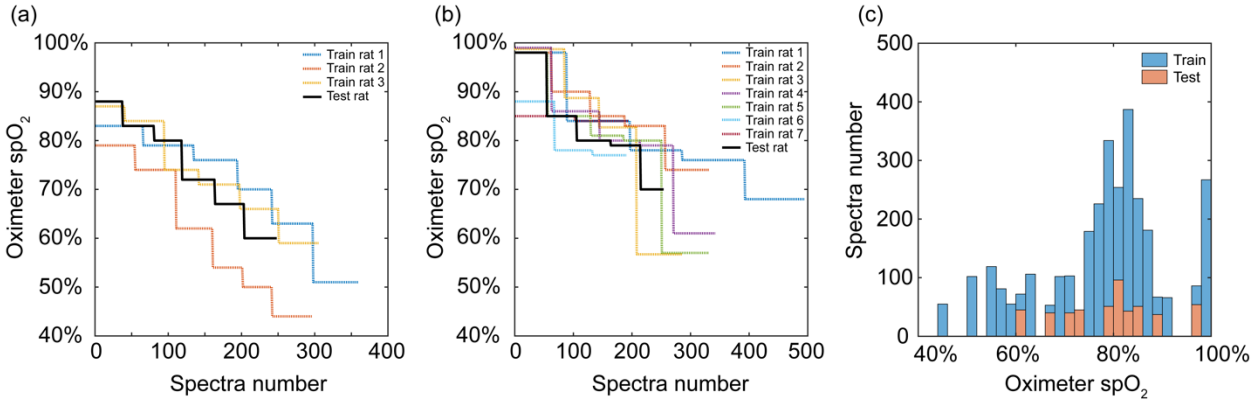

Fig. 3. Oximeter spO_2 readings (ground-truth labels) for training and testing. (a) The readings of rat retinal arterioles from normoxia to hypoxia from Ref. [13]. (b) The readings of rat retinal arterioles from hyperoxia, 5% CO_2, normoxia to hypoxia from Ref. [19]. (c) The histograms of all the readings.

B. Prediction of arterial sO_2

Once the networks are trained, the sO_2 prediction from the testing data by FNN and CNN are plotted in Fig. 4(a) and 4(b), respectively, along with the ground-truth oximeter spO_2 readings and the LSF estimates for comparison. The first 248 predicted sO_2 values are the testing data from Ref. [13] and the rest 254 are from Ref. [19]. Using the data from Ref. [13], the sO_2 estimation using both DSL models (FNN, CNN) and the LSF agree well with the change in the spO_2 readings. Much lower variations are observed in the DSL predictions than those from LSF, owing to DSL's improved robustness to noise and other random signal fluctuations. When the same LSF parameters and inverse calculation procedure for the data from Ref. [13] are applied to the data from Ref. [19], large deviations from the oximeter spO_2 readings are observed in the estimates. In comparison, the sO_2 predicted by both DSL models consistently agree with the spO_2 readings, demonstrating the DSL's robustness to experimental variations present in these two data sets, such as difference in SNR, imaging protocol and speed. The absolute errors calculated from the sO_2 predictions and the corresponding spO_2 readings from the FNN and CNN are plotted in Fig. 4(c) and 4(d), as compared to those from the LSF. Errors from both DSL models are significantly lower than the LSF model, especially for the testing data from Ref. [19]. To quantitatively compare the three different models, we calculate the mean square errors (MSE) of the FNN and CNN models as 0.3149% and 0.3384%, respectively, both of which are much lower than 3.870% by the LSF. An important feature of our DSL models is its ability to quantify uncertainty via the tandem standard deviation (σ) for each sO_2 prediction (Fig. 4(e) and 4(f)). Overall, both FNN and CNN predict σ around 5-6%, and the variation of σ justifies the use of heterogeneous Gaussian distribution model in our customized loss function. We also see that the prediction for Ref. [19] has lower value and variation on σ than Ref. [13], presumably due to more animal numbers and larger training dataset (Fig. 3). Out of 2779 training data, 1930 of them are from Ref. [19].

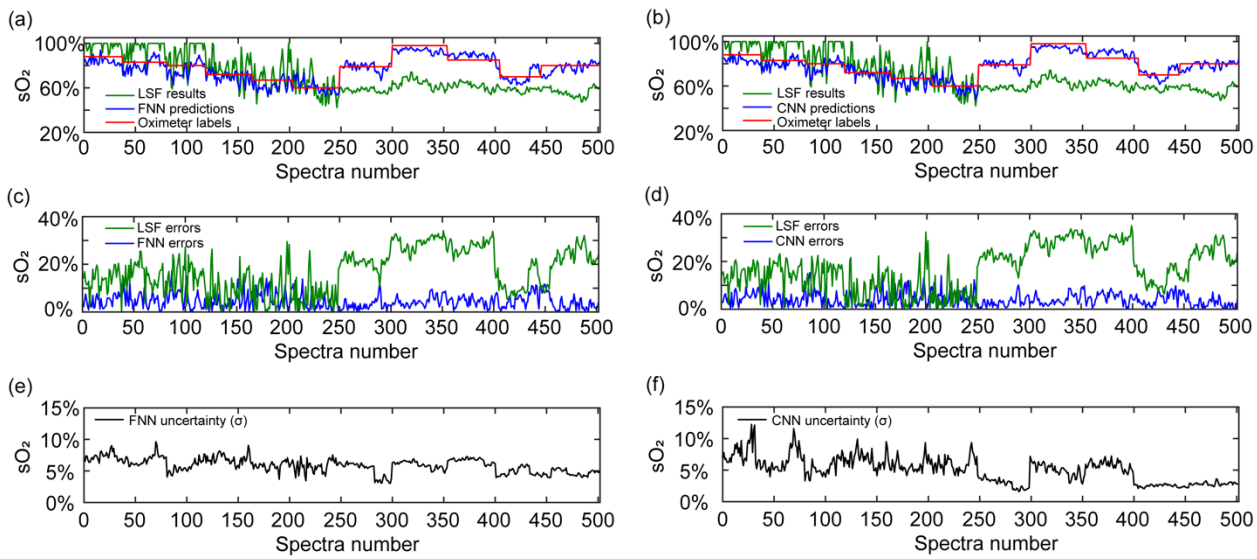

Fig. 4. Rat retinal arterioles sO_2 of the testing data at different external oxygenation status by the FNN model, the CNN model, and the LSF. The predicted sO_2 by the FNN (a) and the CNN (b), compared with the oximeter spO_2 readings and the LSF calculations. The predicted uncertainties of sO_2 by the FNN (c) and the CNN model (d) measured by the standard deviations (σ). The errors of the predicted sO_2 by the FNN (e) and the CNN model (f) as compared with the LSF results.

C. Evaluation of the quantified uncertainty

Our uncertainty quantification assumes that the predicted sO_2 follows a heterogeneous Gaussian distribution given different inputs. To validate our uncertainty metrics, we retrospectively calculated the actual probability that the ground-truth (spO_2) falls within a certain confidence interval of the predicted sO_2, and summarize the results using the reliability diagram [34,37,42]. To construct a reliability diagram, we gather a sub-set of predictions with a specific standard deviation σ_0. We then calculate the probability from this sub-set of data that satisfy the criterion of $|\mathrm{sO}_2 - \mathrm{spO}_2| < \eta\sigma_0$, where sO_2 and spO_2 are the prediction and the corresponding ground-truth:

$$P(\sigma_0, \eta) = \frac{1}{|S_{\sigma_0}|} \sum_{i \in S_{\sigma_0}} \mathrm{I}_{\{|\mathrm{sO}_2 - \mathrm{spO}_2| < \eta\sigma_0\}} \quad \textbf{(5)}$$

where η is a variable that defines the confidence interval, and S denotes the sub-set of the prediction with the specified standard deviation. In practice, we relax σ_0 to σ_0±1% in order to include sufficient data for ensuring reliable statistical calculations. Intuitively, the probability will approach 1 when η increases, *i.e.* a larger error tolerance. At the same time, η also corresponds to a theoretical confidence by the Normal distribution. The reliability diagram essentially plots the actual probability against η, or theoretical confidence. Ideally, the actual probability will equal to the theoretical confidence -- falling on the diagonal line in the graph.

The reliability diagrams for both models are shown in Fig. 5. To cover over 90% of the total 502 predictions of the testing data in the reliability diagram, we set σ_0 = 5% and 7% for FNN model and σ_0 = 3%, 5% and 7% for the CNN model, respectively. For both models, the sO_2 predictions with uncertainty falling in the 7%±1% range are slightly conservative, with $P(\sigma_0, \eta)$ higher than the predicted confidence; the others with uncertainty falling in 5%±1% and 3%±1% are slightly overconfident, with $P(\sigma_0, \eta)$ lower than the confidence. Visualization 1(a) – 1(f) in Supplementary Material illustrate the sub-sets of data when σ_0 = 3%, 5% and 7% for both models. Statistical results with linear fits are also shown in Fig. 5(a) and 5(b) for the FNN and the CNN models, respectively. Table 1 provide a summary of the fitting parameters, slope and constant, and how many of the total 502 test data were counted for the statistical analysis within each uncertainty range.

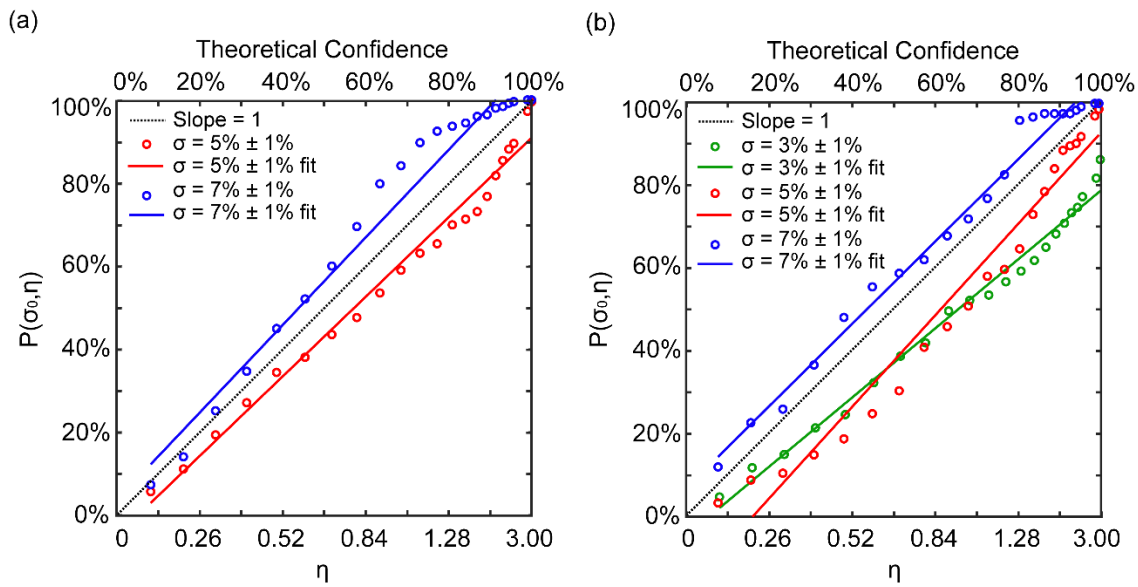

Fig. 5. Statistical analysis of the sO_2 predictions with the quantified uncertainty. (a) The Linear fit of the empirical accuracy to the confidence for the FNN model when the uncertainty measured by the standard deviation (σ) is 5%± 1% and 7%± 1%. (b) The linear fits of the empirical accuracy to the confidence for the CNN model when the standard deviation is 3%± 1%, 5%±1%, and 7%±1%.

Table 1: Parameters of linear fit of $P(\sigma_0, \eta)$ to theoretical confidence for sO_2 predictions with different uncertainties (σ) by the FNN and the CNN models.

Models and uncertainties	Slope	Constant	Datasets
The FNN model (σ: 5% $\pm$ 1%)	0.9600	-0.04886	219
The FNN model (σ: 7% $\pm$ 1%)	1.057	0.03658	252
The CNN model (σ: 3% $\pm$ 1%)	0.8335	-0.04820	156
The CNN model (σ: 5% $\pm$ 1%)	1.104	-0.1797	181
The CNN model (σ: 7% $\pm$ 1%)	0.9929	0.06807	122

D. *En face* sO_2 maps with uncertainty quantification

After model test and uncertainty analysis, we used the testing data from Ref. [12] and applied the FNN and CNN models for retinal imaging oximetry in comparison to LSF (Fig. 6), at three different ventilation conditions. The oximetry by both FNN and CNN clearly reflect the sO_2 changes of all vessels from hypoxia, normoxia, to hyperoxia, with the predicted sO_2 of arterioles matching with the oximeter spO_2 readings well. The sO_2 contrast between arterioles and venules are also clearly visualized in hypoxia and normoxia. In comparison, although the sO_2 by the LSF reflects a similar sO_2 change form hypoxia to hyperoxia and the arteriovenous contrast, all estimated arterioles sO_2 are much lower than the spO_2 readings due to the experimental variations present in the two experiments. There are higher variances in the sO_2 results within each individual blood vessel by LSF, particularly in Fig. 6(i). These results clearly indicate the superior robustness and resilience to variations in experimental conditions and within vessels by DSL. Importantly, our DSL models enable direct visualization of the "*en face*" uncertainty maps of the sO_2 predictions in Fig. 7. On average, the CNN model has lower uncertainty with more confident sO_2 predictions than the FNN model (Fig. 7(a)-7(f)). At hypoxia, it appears the sO_2 estimation is problematic at the periphery on the right side due to lower SNR, where sO_2 is inconsistent within the vessel, as shown in Fig. 6(a), 6(d) and 6(g). At those regions, the CNN model gave correct uncertainty estimation, recapitulating higher values of σ (Fig. 7(d)). The uncertainty increases in hyperoxia, presumably due to less training data (Fig. 3c).

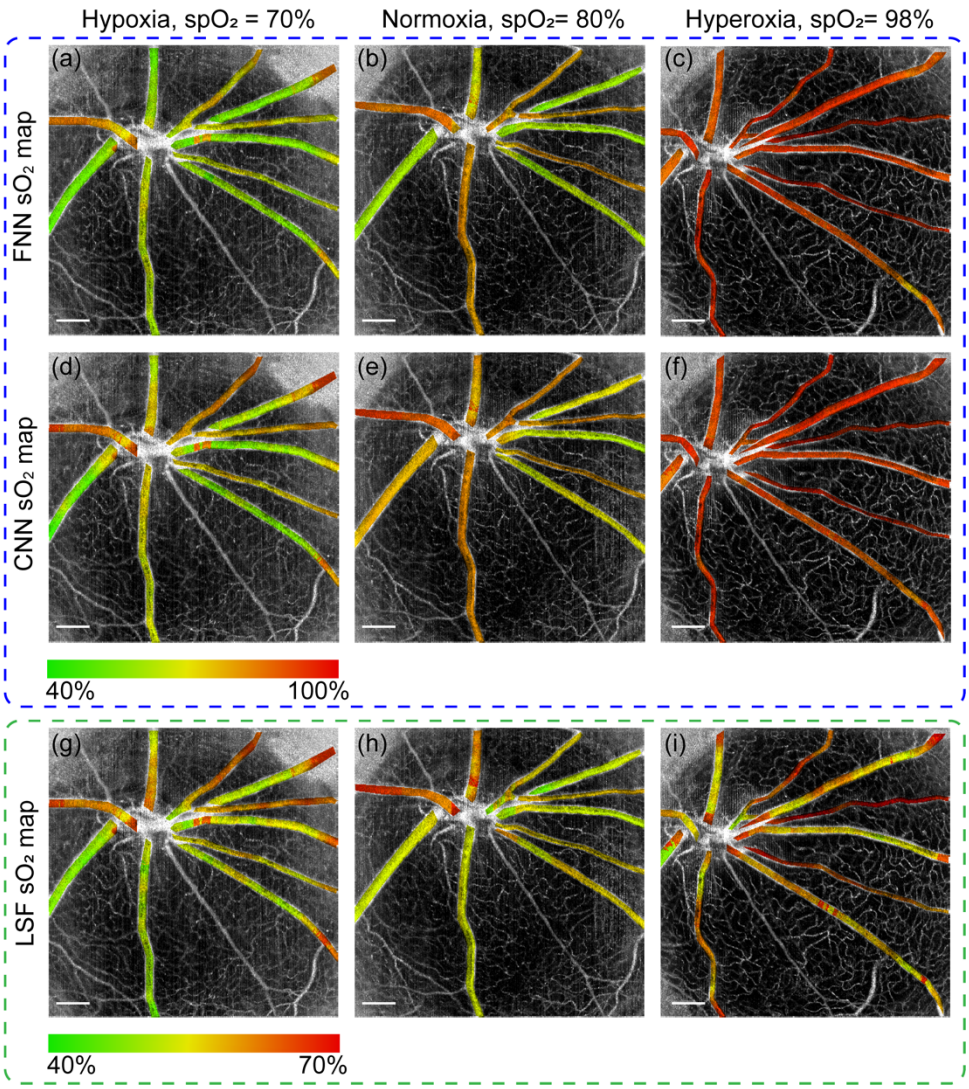

Fig. 6. The *en face* sO_2 maps of the test rat retinal arterioles and venules at hypoxia in (a), (b), and (c), normoxia in (d), (e), and (f), and hyperoxia in (g), (h), and (i) status. The predicted sO_2 by the FNN model are in (a), (d), and (g), and by the CNN model are in (b), (e), and (h). The calculated sO_2 by the LSM are in (c), (f), and (i). The oximeter spO_2 readings at hypoxia, normoxia, and hyperoxia status are 70%, 80%, and 98%, respectively. Scale bar: 500 μm.

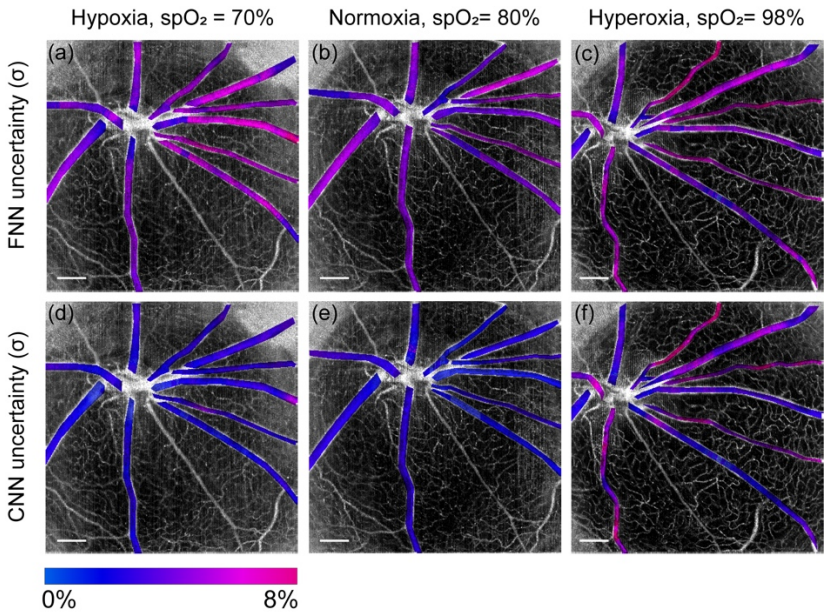

Fig. 7. The *en face* uncertainty (σ) maps for sO_2 predictions of the test rat retinal arterioles and venules by the FNN model in (a), (b), and (c), and the CNN model in (d), (e), and (f). The predicted uncertainty (σ) at hypoxia status are in (a) and (d), at normoxia status are in (b) and (e), and at hyperoxia status are in (c) and (f). The oximeter spO_2 readings at hypoxia, normoxia, and hyperoxia status are 70%, 80%, and 98%, respectively. Scale bar: 500 μm

5. Summary

In this paper, we present a new framework for optical imaging oximetry based on deep spectral learning (DSL). The DSL offers several unique advantages as compared to existing least-square-fitting (LSF) method. First, it bypasses the need for any rigid analytical models and is highly flexible and resilient to experimental variations. We tested DSL on two datasets from two separate vis-OCT experiments, and showed that DSL maintained consistent agreements with the ground-truth spO_2 despite many differences in these two experiments. In contrast, LSF with the identical parametric setting generated significant biases between the two datasets. Second, without the restriction of any rigid models, DSL allows more flexible and efficient use of the data. Here, we demonstrate the effectiveness of using the spectra from both the middle and bottom of the vessels, since both carry sO_2-dependent spectral contrast. Most importantly, DSL not only provides the point estimate of sO_2 but also quantifies the tandem uncertainty of the prediction. Quantifying the statistical uncertainty for each measurement is not possible using the traditional LSF approach, however is valuable in assessing the fidelity of each measurements, in particular in clinical applications. We validate the uncertainty quantification by using the reliability diagram, and for the first time to our knowledge, have constructed uncertainty map of *in vivo* imaging oximetry showing the estimation confidence by DSL. More generally, our DSL framework presents an attractive data-driven approach for other inverse scattering spectral analysis beyond oximetry.

6. Methods

A. Vis OCT experiments

The vis-OCT systems in Ref. [13] and Ref. [19]had the same spectral range from 520 nm to 630 nm, with the same lateral and axial resolutions estimated as 15 μm and 1.7 μm, respectively. The scanning protocol in Ref. [13] used a raster scan over a 20° square retinal area covering a field of view (FOV) of 2.51mm×2.51mm, with 256×256 pixels in each direction at a 25 kHz A-line rate. The exposure time for the spectrometer camera was 37μs. The entire vis-OCT image stack took 3.3 seconds to acquire. The scanning protocol in Ref. [19] was for optical coherence tomography angiography, scanning a 40 ° square retinal area covering a FOV of 4.37mm×4.37mm, with 400 pixels in the A-line direction and 512 pixels in the B-scan direction at a 50 kHz A-line rate. The exposure time for the spectrometer camera was 17μs. For the sake of vis-OCT angiography, there

were repetitive (5×) unidirectional B-scans of the same cross section, giving a total of 5×512 B-scans for each acquisition. The entire vis-OCT image stack took 25.6 seconds to acquire.

B. Spectral Extraction and Data Preprocessing

Wavelength-dependent vis-OCT images were first generated by a short-time Fourier transform (STFT) with 14 equally spaced Gaussian spectral windows in the k-space. The wavelength spans from 523.4 – 604.5 nm. The size (FWHM) of the Gaussian window in the k-space is 0.32 μm^{-1}, corresponding to a bandwidth of ~17 nm at 585 nm. After STFT, a spectrum can be obtained at each 3D vis-OCT voxel. Next, we performed segmentation to isolate the spectra within retinal arterioles [21,22]. Retinal blood vessels are first segmented from the *en face* projection of the 3D vis-OCT image by a threshold-based algorithm [22]; next all A-lines within the segmented retinal arterioles are shifted in the axial direction in reference to the retinal surface and randomly shuffled. A rolling averaging of 100 shuffled A-lines in Ref. [13]and 250 shuffled A-lines in Ref. [19] was performed with 50 and 125 rolling step size. Because the major vessels are located on top of the retina, signal within the vessels can be averaged in reference to the retinal surface to generate one spectrum. We located the bottom vessel wall[13,18], and averaged signals ±16.63 μm as the vessel bottom spectrum. We then averaged the signal from ~25 μm above the bottom vessel and within a range of ±8.31μm as the spectrum from vessel center. For DSL, these two spectra were concatenated as a whole spectral input. Finally, each individual spectral input was normalized by its mean to ensure similar scaling of all datasets before neural network training.

C. Network architectures and training

The first FNN model as illustrated in Fig. 2(a) concatenates two spectra from both the bottom and center of the vessels with a size of 1×28 as input. The output predicts both the mean sO_2 level and the uncertainty (measured by the standard deviation) in two output channels, both of which are single values in a unit of volume percentage of blood being oxygenated. The model has two hidden layers, each having 24 units. We use the Relu activation function in the two inner layers, and the sigmoid activation function in the final layer to normalize the predictions between 0% and 100%.

The second CNN model as illustrated in Fig. 2(b) takes the same input as the FNN model, and also predicts the sO_2 with uncertainty. The model has three convolutional layers, with each of them having a filter size of 3 and filter numbers as 48, 36, 24, respectively, and one flatten layer. We use the Relu activation function in the three convolutional layers, and the sigmoid activation function in the final layer.

All the data processing and network training are implemented in Python using TensorFlow/Keras library. Both models were trained with an initial learning rate of 2.5×10^{-3} and we gradually decreased the rate by $1/(1+\alpha N)$, where N is the epoch number, and α is a decay rate set to be 0.01. The same total epoch number of 2000 with the same batch size of 50 ensured that the learning curve could reach a plateau. We set the validation split ratio as 0.2 and selected the model with the minimum validation loss as the optimal one for following sO_2 prediction.

D. Loss function for uncertainty quantification

In the proposed DSL model denoted by its network weights w, the network makes estimation on both the mean and the standard deviation of sO_2 given the input spectral measurement. Assuming that sO_2 of all retinal arterioles of a rat at each particular oxygenation status satisfy different Gaussian probability distributions,

$$p_w(\mathrm{sO}_2|I_i) = N(\mathrm{sO}_2, \sigma^2) \quad \textbf{(6)}$$

where the mean and the standard deviation of the Gaussian distribution is denoted as sO_2 and σ, and N denotes the normal distribution. The neural network learns a highly-complex nonlinear function. During the training, the network weights, w, are estimated by maximizing the joint likelihood over N training data pairs:

$$w = \underset{w \in F}{\operatorname{argmax}} \operatorname{likelihood}(w) = \arg\max_{w \in F} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2(w)}} e^{-\frac{\left([\mathrm{sO_2}]_i(w) - [\mathrm{spO_2}]_i\right)^2}{2\sigma_i^2(w)}} \quad \textbf{(7)}$$

Equivalently, the customized Gaussian loss function $L_G(w)$ is minimized when training the DSL models in Eq. 4.

D. Reconstruction of *en face* Maps of sO_2 and Uncertainty (σ)

To reconstruct a 2D *en face* map for sO_2 or uncertainty (σ) we applied the same spatial signal averaging procedure as described in section 7. B., but this averaging was done pixel-wisely within each blood vessel. We first segmented the vessel area manually. For each pixel from the 2D *en face* map within the segmented blood vessel, an arteriole or a venule, its depth-dependent spectra were generated by averaging spectral signals from its 100 (in the first literature) or 250 (in the second literature) nearest neighbors based on Euclidian distance. Then, these spectra would be input into the FNN or CNN model to predict a sO_2 with uncertainty for this pixel. Next, the above two steps would be iterated pixel-wisely until all pixels within this particular blood vessel had predicted sO_2 and uncertainty. Finally, the above three steps would be iterated until all arterioles and venules of the rat retina had predicted sO_2 and uncertainty for displaying their 2D *en face* maps. To generate Fig. 6 and Fig. 7, we applied an algorithm based on the HSV (hue, saturation, value) color model, where the predicted sO_2 or uncertainty corresponds to image hue, and the angiography signal intensity corresponds to image value and saturation.

FUNDING

NSF 1846784; NIH R01NS108464, R21EY029412, R01CA224911; Bright Focus Foundation G2017077

ACKNOWLEDGEMENT

We acknowledge Siyu Chen providing the original data for his vis-OCT experiment. We acknowledge Robert Linsenmeier, and Hao F. Zhang for their supports in both vis-OCT studies.

REFERENCE

1. E. Stefánsson, "Ocular Oxygenation and the Treatment of Diabetic Retinopathy," Surv. Ophthalmol. **51**(4), 364–380 (2006).

2. P. Vajkoczy, A. Ullrich, and M. D. Menger, "Intravital Fluorescence Videomicroscopy to Study Tumor Angiogenesis and Microcirculation," Neoplasia **2**(1–2), 53–61 (2002).

3. B. I. Levy, E. L. Schiffrin, J.-J. Mourad, D. Agostini, E. Vicaut, M. E. Safar, and H. A. J. Struijker-Boudier, "Impaired Tissue Perfusion," Circulation **118**(9), 968–976 (2008).

4. A. Edwards-Richards, M. DeFreitas, C. P. Katsoufis, W. Seeherunvong, N. Sasaki, M. Freundlich, G. Zilleruelo, and C. L. Abitbol, "Capillary rarefaction: An early marker of microvascular disease in young hemodialysis patients," Clin. Kidney J. **7**(6), 569–574 (2014).

5. R. N. Kalaria, "Neurodegenerative disease: Diabetes, microvascular pathology and Alzheimer disease," Nat. Rev. Neurol. **5**(6), 305–306 (2009).

6. D. J. Faber, M. C. G. Aalders, E. G. Mik, B. A. Hooper, and T. G. van Leeuwen, "Oxygen saturation dependent absorption and scattering of whole blood," Coherence Domain Opt. Methods Opt. Coherence Tomogr. Biomed. VIII **5316**(2), 78 (2004).

7. J. Q. Nguyen, C. Crouzet, T. Mai, K. Riola, D. Uchitel, L.-H. Liaw, N. Bernal, A. Ponticorvo, B. Choi, and A. J. Durkin, "Spatial frequency domain imaging of burn wounds in a preclinical model of graded burn severity," J. Biomed. Opt. **18**(6), 066010 (2013).

8. S. Gioux, A. Mazhar, B. T. Lee, S. J. Lin, A. M. Tobias, D. J. Cuccia, A. Stockdale, R. Oketokoun, Y. Ashitate, E. Kelly, M. Weinmann, N. J. Durr, L. A. Moffitt, A. J. Durkin, B. J. Tromberg, and J. V. Frangioni, "First-in-

human pilot study of a spatial frequency domain oxygenation imaging system," J. Biomed. Opt. **16**(8), 086015 (2011).

9. F. C. Delori, "Noninvasive technique for oximetry of blood in retinal vessels," Appl. Opt. **27**(6), 1113 (1988).

10. O. Palsson, A. Geirsdottir, S. H. Hardarson, O. B. Olafsdottir, J. V. Kristjansdottir, and E. Stefánsson, "Retinal oximetry images must be standardized: A methodological analysis," Investig. Ophthalmol. Vis. Sci. **53**(4), 1729–1733 (2012).

11. N. P. Blair, J. Wanek, A. E. Felder, C. E. Joslin, J. K. Kresovich, J. I. Lim, F. Y. Chau, Y. Leiderman, and M. Shahidi, "Retinal oximetry and vessel diameter measurements with a commercially available scanning laser ophthalmoscope in diabetic retinopathy," Investig. Ophthalmol. Vis. Sci. **58**(12), 5556–5563 (2017).

12. J. V. Kristjansdottir, S. H. Hardarson, G. H. Halldorsson, R. A. Karlsson, T. S. Eliasdottir, and E. Stefánsson, "Retinal oximetry with a scanning laser ophthalmoscope," Investig. Ophthalmol. Vis. Sci. **55**(5), 3120–3126 (2014).

13. J. Yi, W. Liu, S. Chen, V. Backman, N. Sheibani, C. M. Sorenson, A. A. Fawzi, R. A. Linsenmeier, and H. F. Zhang, "Visible light optical coherence tomography measures retinal oxygen metabolic response to systemic oxygenation," Light Sci. Appl. **4**(9), 1–10 (2015).

14. X. Shu, L. Beckmann, and H. F. Zhang, "Visible-light optical coherence tomography: a review," J. Biomed. Opt. **22**(12), 1 (2017).

15. F. E. Robles, C. Wilson, G. Grant, and A. Wax, "Molecular imaging true-colour spectroscopic optical coherence tomography: Supplementary," Nat. Photonics **5**(12), 744–747 (2011).

16. J. Yi and X. Li, "Estimation of oxygen saturation from erythrocytes by high-resolution spectroscopic optical coherence tomography.," Opt. Lett. **35**(12), 2094–6 (2010).

17. J. Yi, S. Chen, V. Backman, and H. F. Zhang, "In vivo functional microangiography by visible-light optical coherence tomography," Biomed. Opt. Express **5**(10), 3603 (2014).

18. J. Yi, Q. Wei, W. Liu, V. Backman, and H. F. Zhang, "Visible-light optical coherence tomography for retinal oximetry," Opt. Lett. **38**(11), 1796 (2013).

19. S. Chen, J. Yi, and H. F. Zhang, "Measuring oxygen saturation in retinal and choroidal circulations in rats using visible light optical coherence tomography angiography," Biomed. Opt. Express **6**(8), 2840 (2015).

20. R. Liu, G. Spicer, S. Chen, H. F. Zhang, J. Yi, and V. Backman, "Theoretical model for optical oximetry at the capillary level: exploring hemoglobin oxygen saturation through backscattering of single red blood cells," J. Biomed. Opt. **22**(2), 025002 (2017).

21. R. Liu, J. A. Winkelmann, G. Spicer, Y. Zhu, A. Eid, G. A. Ameer, V. Backman, and J. Yi, "Single capillary oximetry and tissue ultrastructural sensing by dual-band dual-scan inverse spectroscopic optical coherence tomography," Light Sci. Appl. **7**(1), (2018).

22. R. Liu, W. Song, V. Backman, and J. Yi, "Quantitative quality-control metrics for in vivo oximetry in small vessels by visible light optical coherence tomography angiography," Biomed. Opt. Express **10**(2), 465 (2019).

23. S. P. Chong, C. W. Merkle, C. Leahy, and V. J. Srinivasan, "Cerebral metabolic rate of oxygen (CMRO_2) assessed by combined Doppler and spectroscopic OCT," Biomed. Opt. Express **6**(10), 3941 (2015).

24. S. Pi, A. Camino, X. Wei, J. Simonett, W. Cepurna, D. Huang, J. C. Morrison, and Y. Jia, "Rodent retinal circulation organization and oxygen metabolism revealed by visible-light optical coherence tomography," Biomed. Opt. Express **9**(11), 5851 (2018).

25. S. Pi, A. Camino, W. Cepurna, X. Wei, M. Zhang, D. Huang, J. Morrison, and Y. Jia, "Automated spectroscopic retinal oximetry with visible-light optical coherence tomography," Biomed. Opt. Express **9**(5), 2056 (2018).

26. B. T. Soetikno, L. Beckmann, X. Zhang, A. A. Fawzi, and H. F. Zhang, "Visible-light optical coherence tomography oximetry based on circumpapillary scan and graph-search segmentation," Biomed. Opt. Express **9**(8), 3640 (2018).

27. Y. H. Tan, J. A. Davis, K. Fujikawa, N. V. Ganesh, A. V. Demchenko, and K. J. Stine, "Surface area and pore size characteristics of nanoporous gold subjected to thermal, mechanical, or surface modification," J. Mater. Chem. **141**(4), 520–529 (2012).

28. L. Wang, K. Maslov, and L. V. Wang, "Single-cell label-free photoacoustic flowoxigraphy in vivo," Proc. Natl. Acad. Sci. **110**(15), 5759–5764 (2013).

29. R. V. Kuranov, J. Qiu, A. B. McElroy, A. Estrada, A. Salvaggio, J. Kiel, A. K. Dunn, T. Q. Duong, and T. E. Milner, "Depth-resolved blood oxygen saturation measurement by dual-wavelength photothermal (DWP) optical coherence tomography," Biomed. Opt. Express **2**(3), 491 (2011).

30. R. V. Kuranov, S. Kazmi, A. B. McElroy, J. W. Kiel, A. K. Dunn, T. E. Milner, and T. Q. Duong, "In vivo depth-resolved oxygen saturation by dual-wavelength photothermal (DWP) OCT," Opt. Express **19**(24), 23831 (2011).

31. B. Cox, J. G. Laufer, S. R. Arridge, and P. C. Beard, "Quantitative spectroscopic photoacoustic imaging: a review.," J. Biomed. Opt. **17**(6), 061202 (2012).

32. J. Yao, L. Wang, J. M. Yang, K. I. Maslov, T. T. W. Wong, L. Li, C. H. Huang, J. Zou, and L. V. Wang, "High-speed label-free functional photoacoustic microscopy of mouse brain in action," Nat. Methods **12**(5), 407–410 (2015).

33. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**(9), (2017).

34. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach towards scalable imaging through scattering media," Optica **5**(10), (2018).

35. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," Optica **5**(6), (1995).

36. X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-Supervised Automatic Segmentation of Layer and Fluid Region in Retinal Optical Coherence Tomography Images Using Adversarial Learning," IEEE Access **7**, 3046–3061 (2019).

37. Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," Optica **6**(5), 618 (2019).

38. T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P. M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. Decaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene, *Opportunities and Obstacles for Deep Learning in Biology and Medicine* (2018), **15**(141).

39. M. Hammer, W. Vilser, T. Riemer, and D. Schweitzer, "Retinal vessel oximetry-calibration, compensation for vessel diameter and fundus pigmentation, and reproducibility," J. Biomed. Opt. **13**(5), 054015 (2008).

40. N. Bosschaart, G. J. Edelman, M. C. G. Aalders, T. G. Van Leeuwen, and D. J. Faber, "A literature review and novel theoretical approach on the optical properties of whole blood," Lasers Med. Sci. **29**(2), 453–

479 (2014).

41. A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," Neural Inf. Process. Syst. (Nips), (2017).

42. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nat. Methods **15**(12), 1090–1097 (2018).