Article

Deep Learning of Diffuse Optical Tomography Based on Time-Domain Radiative Transfer Equation

Yuichi Takamizu [1,*] , Masayuki Umemura [1], Hidenobu Yajima [1], Makito Abe [1] and Yoko Hoshi [2]

1   Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba 305-8577, Japan
2   Preeminent Medical Photonics Education and Research Center, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan
*   Correspondence: takamizu@ccs.tsukuba.ac.jp

Abstract: Near infrared diffuse optical tomography (DOT) is a potential tool for diagnosing cancer by image reconstruction of tissue optical properties. A variety of image reconstruction methods for DOT have been attempted, in general, based on the diffusion equation (DE). However, the image quality is still insufficient to clinical use, which is mainly attributed to the fact that the DE is invalid in some regions, such as low-scattering regions, and the inverse problem is inherently ill-posed. In contrast, the radiative transfer equation (RTE) accurately describes light propagation in biological tissue and also the DOT by deep learning is recently thought to be an alternative approach to the inverse problem. Distribution of time of flight (DTOF) of photons estimated by the time-domain RTE lends itself to deep learning along a temporal sequence. In this study, we propose a new DOT image reconstruction algorithm based on a long-short-term memory and the time-domain RTE. In simulation studies, using this algorithm, we succeeded in detection of an absorbing inclusion with a diameter of 5 mm, an absorber mimicking cancer, which was embedded in a two-dimensional square model (4 cm × 4 cm) with an optically homogeneous background. Multiple absorbers and a bigger absorber embedded in this model were also detected. We also demonstrate that, if simulation data by beam injection from multiple directions are employed as a training set, the accuracy of detection is improved especially for multiple absorbers.

Keywords: diffuse optical tomography; time-domain radiative transfer equation; deep learning

1. Introduction

Diffuse optical tomography (DOT) using near infrared light (700–900 nm) is one of the most sophisticated optical imaging techniques for biological tissue. This technique is a promising imaging modality for cancer detection owing to its sensitivity to the hemoglobin oxygenation level. Diffuse optical tomography has mainly been developed with three measurement methods: continuous wave (CW), time-domain (TD), and frequency-domain (FD) measurements. Continuous wave and FD measurements provide information about only intensity and that about intensity and phase, respectively. In contrast, distribution of time of flight (DTOF) of photons, which is the histogram of arrival time of photons, is obtained from TD measurement. Thus, TD measurement provides more information needed for image reconstruction compared to CW and FD measurements.

To reconstruct optical properties of biological tissue in DOT, two mathematical problems must be solved, the forward problem and the inverse problem [1]. The forward problem is to follow the propagation of scattered light in biological tissue with given optical properties and thereby predict the scattered light measurements, while the inverse problem is to reconstruct tissue optical properties from scattered light measurements using the forward model. The image reconstruction for DOT is a nonlinear, ill-posed inverse problem, which suffers from the lack of data diversity and from instabilities to noise. Hence, the feasibility of DOT depends upon how precisely the forward problem is calculated and how stably the inverse problem is solved.

The propagation of light in turbid media with absorption is governed by the radiative transfer equation (RTE) [2]. Numerical schemes to directly solve the RTE in biological tissue have been proposed [3,4]. However, they are computationally expensive, since RTE in three-dimensional space results in a six-dimensional problem in photon phase space. In addition, for the solution of the inverse problem to converge, the forward problem must be solved multiple times for a single reconstruction. Therefore, the RTE calculations have been a bottleneck in algorithms when solving the inverse problem. Thus far, light propagation has often been approximated by a diffusion equation (DE) using the P1 approximation of the RTE. The diffusion approximation is a simplification of the RTE for optically thick media in which multiple scattering is dominant [5]. Based on frequency-domain as well as time-domain DE, various approaches for image reconstruction in DOT have been attempted [6].

However, actual tissue systems are in the so-called mesoscopic scattering regime, in which light undergoes multiple scattering but the scattered light is not perfectly diffusive. For mesoscopic scattering, the diffusion approximation breaks down especially near sources and boundaries. To circumvent the shortcomings of the diffusion approximation, hybrid schemes that combine radiative transfer with diffusion have been implemented [7,8]. Hybrid schemes can reduce the computational cost to solve light propagation dramatically, but the transition between RTE and DE cannot be determined a priori. Therefore, full RTE calculations are desirable. As for the frequency-domain RTE, several solvers of the inverse problem have been proposed [9,10]. Since, however, image reconstruction based on time-domain RTE is still immature, we have recently developed a time-domain RTE solver, TRINITY (Time-dependent Radiation Transfer in Near-infrared Tomography) [11], which is based on our former steady-state RTE solver, ART [12].

Deep learning brings a new possibility for image reconstructions. Thus far, most deep learning schemes for DOT have been based on DE [13–15], although a deep learning scheme based on the steady RTE has also been developed [16]. However, no attempt has yet been made for the time-domain RTE. In this study, we construct a novel DOT algorithm based on deep learning and a time-domain RTE, in which DTOFs are used as training data. As for the deep learning, we utilize an LSTM (Long Short-Term Memory) method, which is an extension of the artificial recurrent neural network (RNN) architecture to process not only single data points but entire sequences of temporal data. We apply this algorithm to detect highly absorbing areas in a two-dimensional mathematical model of biological tissue with the optically homogeneous background. The approach similar to ours has been reported by the other research group, whereas this study employed an FD DOT system [15].

In this paper, firstly we describe the details of the algorithm and image reconstruction of a single absorber with datasets obtained from a single beam injection. Then, we present image reconstruction of multiple absorbers with a subtraction method and improvement of image quality by using datasets obtained from multiple beam injections.

2. Methodology

2.1. Model

Throughout this paper, we work with a two-dimensional model for solving the RTE and performing image reconstruction. We consider a target tissue 4 cm × 4 cm in size, which we divide into 28 domains to specify the positions of absorbers, as shown in Figure 1. We model the absorber as a round absorber with a diameter of 5 mm. This setup is based on an experiment using a phantom composed of polyurethane with titanium oxide (scatterer) and carbon black (absorber).

In this paper, we present a two-dimensional model. We compare our RTE calculation and image reconstruction with results from the setup of the experiment, which is a three-dimensional cylinder model. The source of incident beam and detectors are located at the half of the cylinder in height. The details of the experiment and the 3D model will be separately reported. The optical properties of the phantom are characterized by

$$\mu_a = 0.21/\text{cm} \quad \mu_s = 22.45/\text{cm} \tag{1}$$

where μ_a and μ_s are the absorption and scattering coefficient, respectively. Within this background material, an absorbing pole with the following coefficients

$$\mu_a = 0.64/\text{cm} \quad \mu_s = 22.63/\text{cm} \tag{2}$$

is inserted. Scattering dominates absorption in the inserted pole as well as in the background. Absorption features appear in light emerging from the pole at the detectors, since the absorption coefficient is larger by a factor of three in the pole. This is a principle for detecting cancer positions in biological tissue. In this paper, we show the results for the cases in which the absorption coefficient of the absorber is 3 times larger than the surrounding tissue. However, we have also tested how smaller differences in absorption coefficient can be identified by deep learning. It is shown that our classification method can work even for cases in which the absorption coefficient is 1.5 times larger.

The sources of the incident beams and detectors are located at eight points as shown in Figure 1, where the positions of the sources are labelled S1–S8 and positions of detectors are labelled D1–D8.
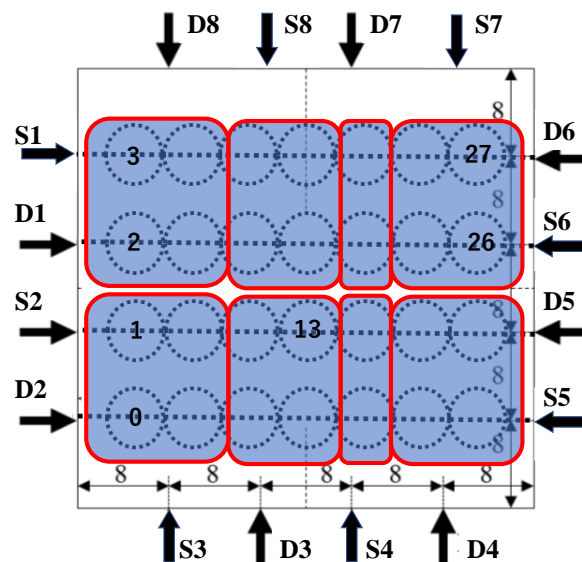
Figure 1. The configuration of target tissue. The dimensions are 4 cm × 4 cm. The tissue is divided into 28 domains, each of which can possess a round absorber with a diameter of 5 mm. Eight incident beam directions (labelled S1–S8) and eight detector positions (D1–D8) are shown. We divide the whole area into eight groups, each of which has four or two domains.

We solve the following time-dependent radiative transfer equation in two-dimensional space to follow scattered and absorbed light,

$$\frac{1}{c}\frac{\partial I}{\partial t} + \boldsymbol{n} \cdot \nabla I = -\mu I + \eta \tag{3}$$

with

$$\mu = \mu_a + \mu_s \eta = \mu_s \oint \phi(\boldsymbol{n}, \boldsymbol{n}') I(\boldsymbol{n}') d\Omega', \tag{4}$$

where I is the specific intensity of light, η is the emissivity by scattered photons, and ϕ is a phase function. The emissivity term represents radiative transfer from one direction $\boldsymbol{n}'$ to another direction $\boldsymbol{n}$. We employ the Henyey–Greenstein function for ϕ, that is,

$$\phi(\boldsymbol{n}, \boldsymbol{n}') = \frac{1}{2\pi} \frac{1 - g^2}{1 + g^2 - 2g(\boldsymbol{n} \cdot \boldsymbol{n}')} \tag{5}$$

where g is the scattering anisotropy parameter. The case for $g = 1$ or $g = 0$ means perfectly forward scattering and isotropic scattering, respectively. We take $g = 0.62$ for the simulations. This value is consistent with the phantom experiment.

To solve Equation (3), we have developed a new solver, TRINITY (Time-dependent Radiation Transfer in Near-infrared Tomography)(Yajima et al. [11]), which is based on our former steady-state RTE solver, ART (Authentic Radiative Transfer) [12]. In this method, the distribution of light rays is constructed independently from the the grids that possess source functions. The most important advantage is the simultaneous reduction of computational cost and numerical diffusion. We set up 64^2 cell grids for the radiative transfer calculations.

Using the radiative transfer calculations, we define the absorption measure as

$$A(t) = [I_{\mathrm{abs}}(t) - I_{\mathrm{noabs}}(t)] / I_{\mathrm{noabs}}(t) \tag{6}$$

where I_{noabs} is the intensity of outgoing light toward a detector when no absorber is included, and I_{abs} is the intensity when one or multiple absorbers are embedded. The quantity of this intensity has been integrated over angles. In Figure 2, the temporal profiles of $A(t)$ are shown for absorbers embedded in various domains. The absolute difference, $I_{\mathrm{abs}}(t) - I_{\mathrm{noabs}}(t)$, is used for DOT. However, $I_{\mathrm{abs}}(t)$ changes by orders of magnitude, depending on the detector positions. Therefore, the relative values given by (6) are more suitable for the deep learning. The temporal profiles of $A(t)$ provide training data for LSTM deep learning. For a given incident laser beam, we obtain datasets for temporal profiles of $A(t)$ at the positions of 8 detectors.
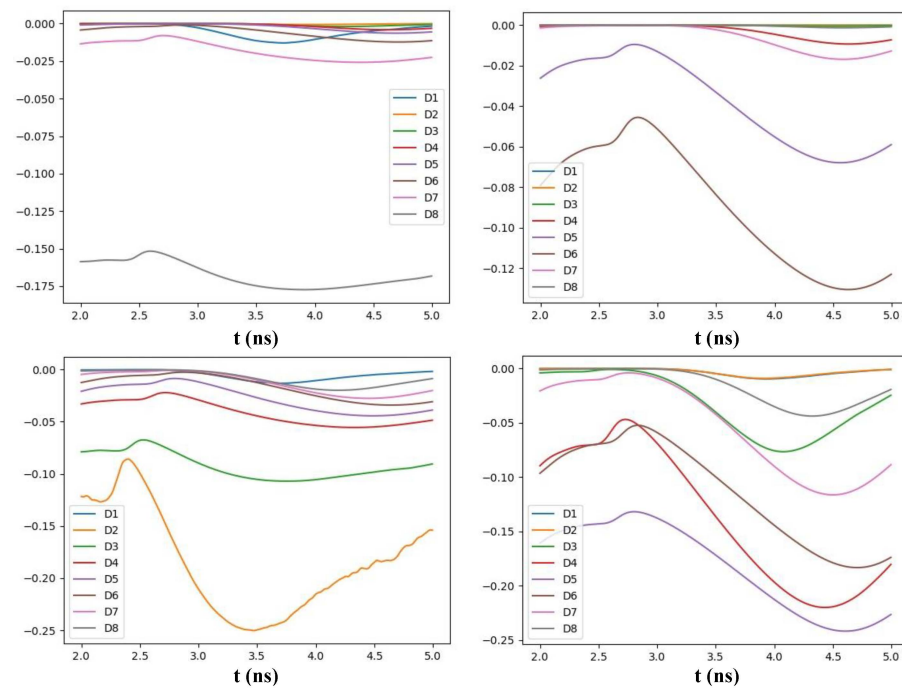

Figure 2. Temporal profiles of absorption measure $A = I_{\mathrm{abs}}(t) / I_{\mathrm{noabs}}(t) - 1$ on 8 detectors (D1–D8). The horizontal axis is time in units of nanoseconds (ns). The upper left panel shows the profiles for an absorber located at the #3 position, upper right panel for an absorber at #26 position, lower left panel for an absorber at #0 position, and lower right panel for an absorber at the #13 position.

2.2. Multi-Step Classification Method

It is difficult to classify all 28 positions directly at once. Thus, we adopt a two-step classification. In the first step, all datasets are classified into 8 groups of domains, and each group is classified into domains in the second step. We find that this multi-classification method is more effective for absorber detection than a single-step method. In Figure 1,

the grouping for the two-step classification is illustrated. The small red boxes are the 8 groups, and one group is composed of either 4 or 2 domains. Although such domain decomposition is not unique, we have found that this particular decomposition works for our deep learning scheme.

The training and test data are the temporal profiles of absorption measure $A(t)$ at the eight detector positions, which are obtained through radiative transfer simulations for an absorber located at a given domain. In addition, we add random noises on the simulation data. We assume a Gaussian type of noise which forms $\exp(-x^2/2\sigma^2)$ with a standard deviation of $\sigma = 0.001$ or 0.01. This is suitable for classification by LSTM deep learning. Temporal profiles of the absorption measure with noises are plotted in Figure 3. We generate 2500 datasets in total, which include 89 datasets for each absorber position.
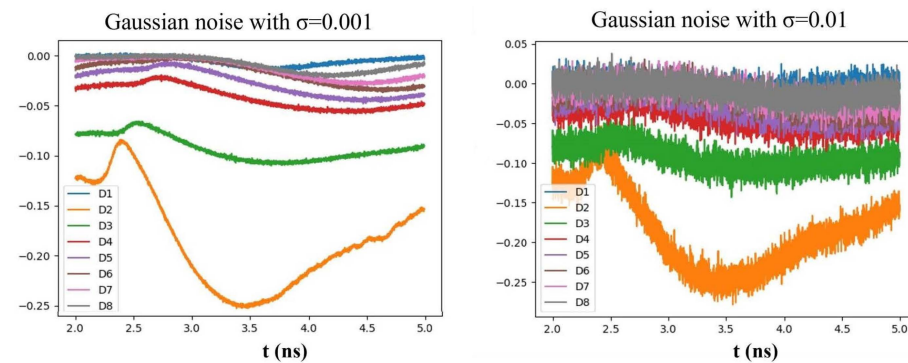
Figure 3. Datasets with noises: the **left panel** shows training data with Gaussian noises whose standard deviation is $\sigma = 0.001$, and the **right panel** is test data with Gaussian noises of $\sigma = 0.01$.

2.3. LSTM Deep Learning Method

In order to classify absorber positions, we employ a deep learning method. For temporal data, Long Short Term Memory (LSTM) learning is an effective tool that is an extension of the artificial recurrent neural network (RNN) architecture to analyze not only single data points but also entire sequences of temporal data. In Figure 4, an example of a LSTM networks that we employ is shown. LSTM networks have improved recurrent neural networks by using gates to selectively retain and forget information, which are relevant and not relevant, respectively. Lower sensitivity to time gaps makes LSTM networks robust for analysis of temporal data compared with a simple recurrent network.

LSTM has four neural network layers. Figure 4 shows the explicit interaction of such layers. The input data are x_t at time step t. We put the absorption data $A(t)$ in x_t and the output data are y_t. Although RNN uses just one *tanh* layer, some output data are reused as input data in the next time step in LSTM. In Figure 4, boxes and circles represent layers and pointwise operations, respectively. The symbol σ is the sigmoid function, which determines whether data can be transferred to the next gate. The symbol *tanh* is the hyperbolic tangent function. It is used to restrict data in the range $(-1, 1)$. The actual calculations are conducted as follows:

$$D_t = \sigma\left(W_D \cdot y_{t-1} + U_D \cdot x_t + b_D\right) \tag{7}$$

$$E_t = \sigma\left(W_E \cdot y_{t-1} + U_E \cdot x_t + b_E\right) \tag{8}$$

$$F_t = tanh\left(W_F \cdot y_{t-1} + U_F \cdot x_t + b_F\right) \tag{9}$$

$$G_t = \sigma\left(W_G \cdot y_{t-1} + U_G \cdot x_t + b_G\right) \tag{10}$$

$$C_t = E_t \times F_t + D_t \times C_{t-1} \tag{11}$$

$$y_t = G_t \times tanh(C_t) \tag{12}$$

Here, *W*, *U*, and *b* are weight matrix and bias in the neural network, respectively. Though we have to set the value of *b* as a parameter, the parameters, *W* and *U*, are automatically determined by each learning step in order to make a good classification. This is the key point for deep learning. The *sigmoid* and *tanh* activation functions are used for each component of any 8-dimensional vector such as y_{t-1} and x_t. The symbol $\times$ denotes an element-wise product. LSTM uses three types of gates: input, output, and forget gates, which are shown by different colors of circles in Figure 4. This allows the network to retain or forget some information.
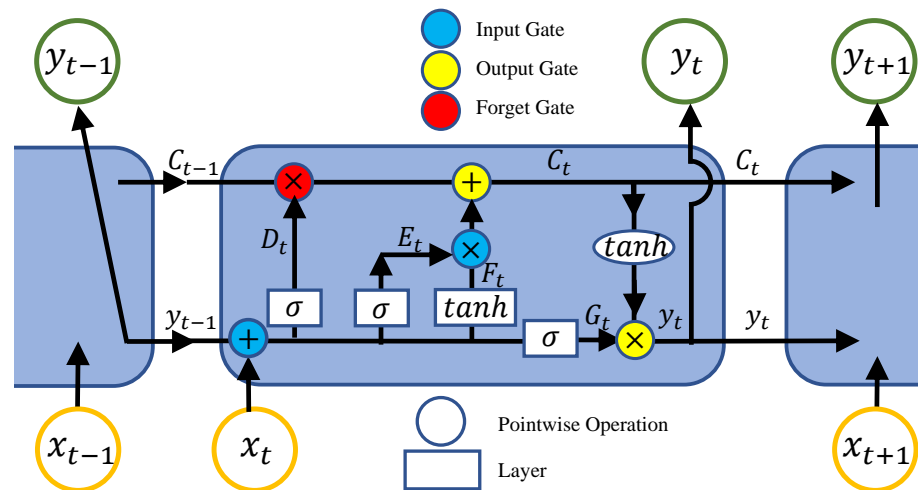
Figure 4. A schematic view of our deep learning scheme based on LSTM. Boxes and circles represent layers and pointwise operations, respectively. Blue circles are input gates and yellow circles are output gates. In addition, a forget gate shown by red circles is embedded, which allows the network to retain or forget some information. See text for the details of the algorithm. For example, the text σ means a sigmoid function and tanh means a tangent hyperbolic function, respectively, which are written in Equations (7)–(12).

We implement a deep learning scheme composed of two LSTM layers plus a final dense layer utilizing TensorFlow 2.1.0, a free and open-source library for deep learning. We set the bias function *b* zero as a simple setup. In order to classify multiple domains, we use a categorical cross entropy as a loss function and a softmax activation function. The cross entropy loss function is defined as

$$-\sum_{j=1}^{M} d_j \log(p_j) \tag{13}$$

where *M* is the number of domains, d_j is a binary indicator which assumes the value 1 if domain label *j* is the correct classification, and is otherwise 0, and p_j is the predicted probability observing from the LSTM network domain *j*. We calculate a separate loss for each domain label and sum over all domains. The softmax activation function is defined as

$$p_j(x_j) = \frac{e^{x_j}}{\sum_{i=1}^{M} e^{x_i}} \tag{14}$$

where x_i is element of the input data vector *x*. We use the softmax activation function to solve a classification problem.

We conduct deep learning based on the the algorithm shown in Figure 5. What we are doing in deep learning this time is classification learning. The output data are categorized by positions. The learning steps are taken as three methods as follows:

- Multi-step classification: First, we divide the 28 positions into 8 groups and identify an appropriate group using a deep learning model for the groups. Then, an absorber in the group is identified with a different deep learning model;
- Data subtraction method, where we subtract the data of a firstly detected absorber and then reanalyze the remaining data. In the case of an absorber larger than the size used for the training data, we put the data into the deep learning model twice, where the machine predicts two domains to express the big absorber;
- Multi-beam injection method: we use time-domain data for beams with different injection places. The accuracy of detection is improved especially for multi absorbers, that is, improving accuracy for the prediction.

All detailed explanations will be shown in the later section.
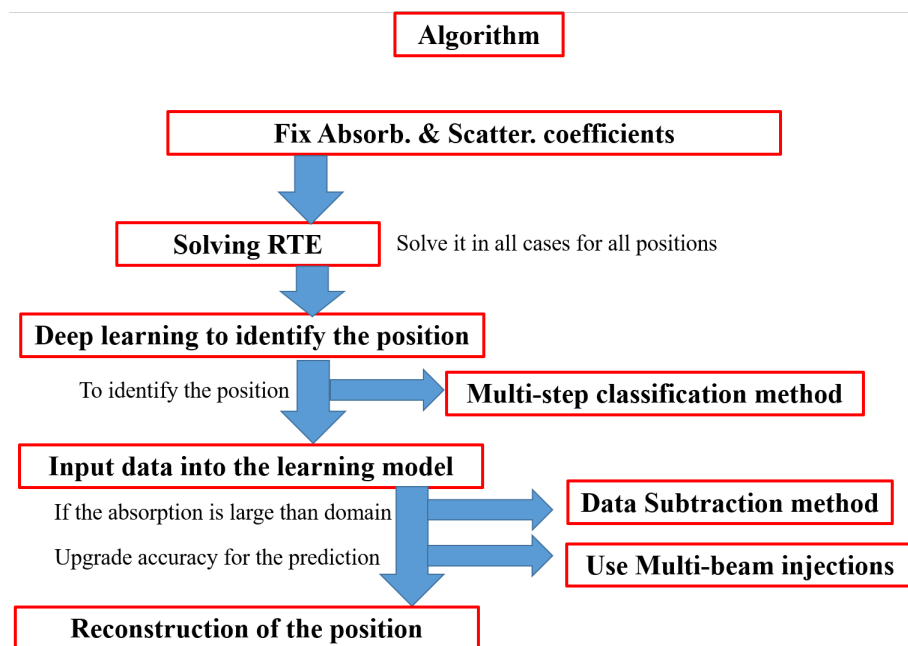
Figure 5. The algorithm of the present deep learning based on the LSTM method.
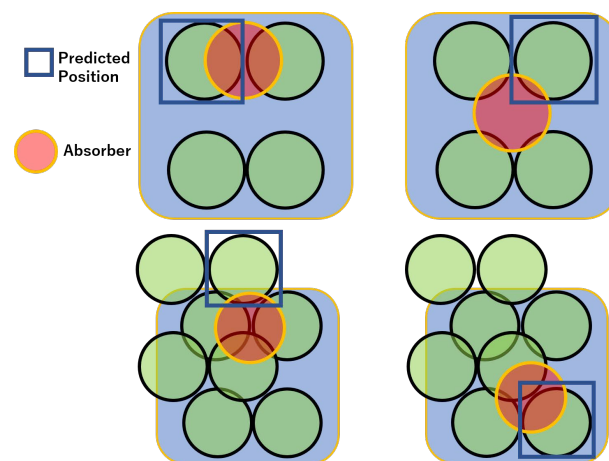
3. Results on Single Absorber Detection

The TD data in our simulations consist of signals at eight detectors. The multi-position of detectors leads to time difference in a picosecond scale. Light propagation simulations by RTE provide precise information on the time difference at detectors. We aim to develop a new method to predict the position of the absorber from the experimental data.

Here, we consider the case for a laser beam injected from S2. Figure 2 shows the resultant temporal profiles of the absorption measure $A(t)$, depending on the position of the absorber. For example, the upper left panel shows temporal profiles of $A(t)$ for an absorber at #3. Since the absorber is very close to detector D8, the curve labeled D8 shows significant absorption. Similarly, the lower left panel (absorber at #0) and the upper right panel (absorber at #26) exhibit strong absorption at the detectors in the vicinity of the absorber. The lower right panel (absorber at #13) shows the result for an absorber near the center of the whole region. In this case, all detectors register significant absorption, and the temporal profiles depend strongly upon the detector position. These features of the absorption measure $A(t)$ are used for training data in the deep learning method used in this work as described below.

The first classification is conducted to determine a target group out of eight groups. In this classification, the number of learning epochs is about 40 and the number of training data and the number of test data are 2500 and 500, respectively. The final loss value is 0.03, and the resultant accuracy to determine the target group reaches 99%. As the the second step, the sub-classification is achieved to determine a domain containing an absorber in

the group. For groups composed of four domains, the number of learning epochs is 10 and the number of training data and test data are 1000 and 200, respectively. For groups composed of two domains, the number of learning epochs is 10 and the number of training data and test data are 600 and 100, respectively. As a result, we find that the final accuracy rate to predict the domain reaches 99.9%. The accuracy rate depends on the noise level. If we employ Gaussian noises with $\sigma = 0.01$ that is ten times larger than the fiducial value ($\sigma = 0.001$), then the accuracy rate is further improved.

We also study the detection of an absorber shifted from a domain. Figure 6 shows the results for the detection of a shifted absorber. The absorber is shown by a red circle, and the predicted position is indicated by a square. In Figure 6, the upper left panel shows that the model can predict a position near the correct position. The upper right panel shows the results for an absorber at the center of four domains. Again, a near position is predicted. Next, we consider the refinement of domains. In the lower panels of Figure 6, the results for eight domains are shown. The accuracy for the prediction of a position can be improved. These results demonstrate that, by the refinement of domains, we can predict the position of absorber more precisely.

Figure 6. Results of position prediction for absorbers with offsets. Red circles are absorbers, and blue squares are predicted positions.

4. Applications to Multiple Absorbers

In the above, we have used just the data for a single absorber to predict the position. However, if there are multiple absorbers, the temporal profiles of $A(t)$ emerge as the superposition of information of those absorbers. Hence, it is hard to predict all positions simply using the above method. In this case, it is effective to subtract the data of the firstly detected absorber and then reanalyze the remaining data in order to predict the position of another absorber. Here, we demonstrate the effectiveness of such data subtraction method to detect two absorbers or an absorber bigger than a domain.

4.1. Two Absorbers

First, we obtain temporal profiles of $A(t)$ for absorbers located at two different domains. We then apply our method to the data. Once a position of an absorber is predicted, we subtract the profile data of the absorber from the original data. Then, we again apply our detection method to the remaining data after subtraction. Figure 7 shows the results for two absorbers, where orange and green circles are absorbers and predicted positions are indicated by blue squares. These results show that the subtraction method works successfully to predict positions of two absorbers. In most cases, correct positions or close positions are predicted, while in the case shown in the lower right panel, prediction is not so accurate. This can be improved by multi-beam injection as shown below. After we tested 60 cases for two absorbers, we found that the data subtraction method is effective for detecting multiple absorbers at a high accuracy rate of 88%.
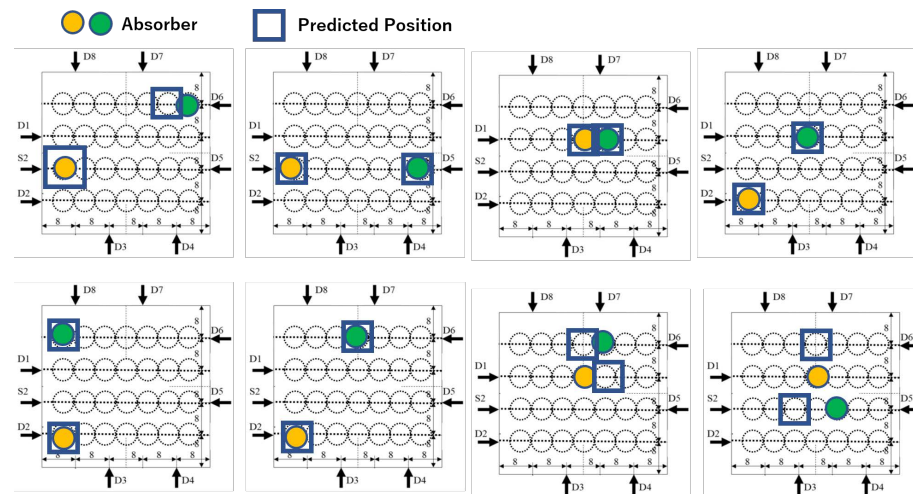
Figure 7. Example results for two absorbers. Orange and green circles are absorbers, and predicted positions are indicated by blue squares.

These results imply that a linear combination of each absorption profile is a good approximation for most cases of two absorbers. When the distance between two absorbers is large, the linearity is expected to hold good. However, the upper middle panel in Figure 6 shows that linearity is a good approximation even for adjacent absorbers, if they are located parallel to the incident beam direction. On the other hand, as seen in the lower right panel, the adjacent absorbers are located perpendicular to the incident beam direction. In this case, the nonlinear effects seem to be significant.

As shown here, the data subtraction method is capable of accurately predicting the correct position or nearest positions. In order to evaluate the probability of a correct detection, it is useful to introduce a scoring scheme. As shown in the left panel of Figure 8, the score for a correct position is 2, while the score for correctly predicting adjacent positions is 1. The right panel of Figure 8 shows an example of a total score of 3. In this case, the detection probability of the two absorbers is $\frac{3}{4} = 75\%$ because the score of correct prediction for both absorbers is 4. Based on such scoring rule, we tested 60 examples for two absorbers and found the average of the detection probability to be 72%.

4.2. Bigger Absorbers

Here, we consider an absorber with double size in each direction of a domain. The temporal profiles of $A(t)$ for this absorber are obtained through radiative transfer calculations. Then, the data subtraction method is applied for the data. Figure 9 shows the results of predictions for a big absorber. The left panel of Figure 9 shows the classification based on the subtraction method for four domains. Two domains adjacent to the absorbers are predicted with this method. The right panel of Figure 9 shows the classification using eight refined domains. The prediction probabilities are 72% for domain #34, 16% for #33, and 10% for #32. Such probability distributions are informative for a big absorber. We also tested the case of a 1.5 times bigger absorber, and the resultant probabilities are 75% for #34, 13% for #33, and 8% for #32. Compared to the two times bigger absorber, probabilities at #32 and #33 are slightly lower. Thus, the probability distributions reflect the size of absorber.
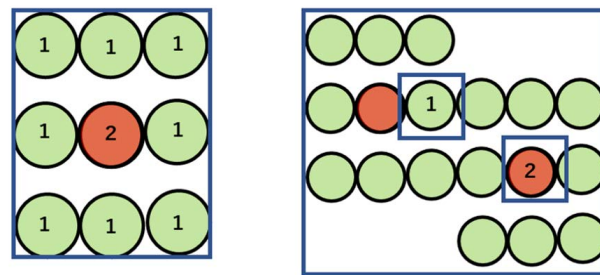
Figure 8. The left panel shows a point at the position for our scoring. If one detects the absorber position, one obtains two points. The right panel shows one example of scoring for two absorbers that leads to three total points, which means the detection probability of 75%.
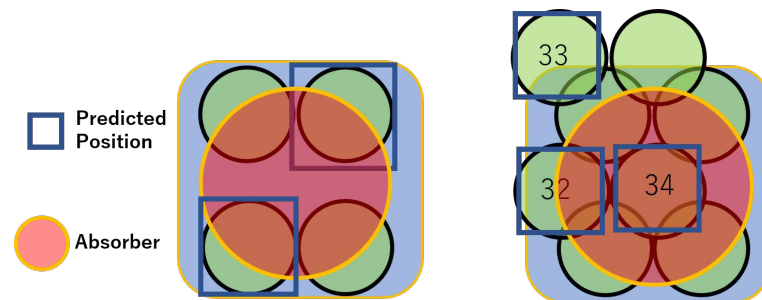

Figure 9. Example results for a big absorber. The left panel shows the classification for four domains, while the right panels shows eight domains. We apply a subtraction method for the case using four domains, but not for the case using eight domains. This figure shows the comparison between the results using a subtraction method and those for small domains without using subtraction. The red circle is the absorber, and predicted positions are indicated by blue squares.
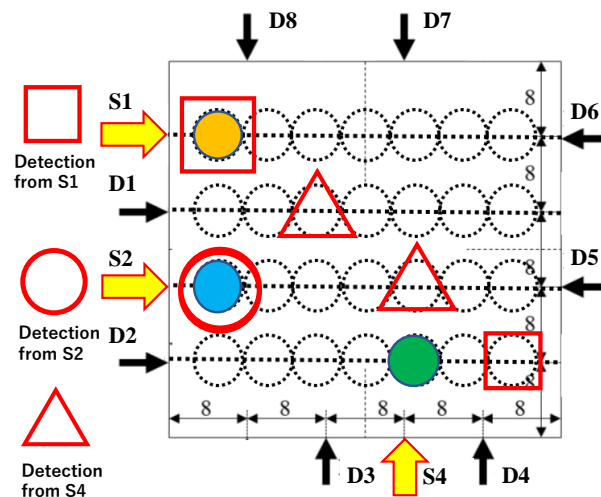
4.3. Multi-Beam Injection Model

Thus far, we have considered the case of only one incident laser beam from S2 shown in Figure 1. Here, we analyze the cases for laser beams injected at multiple positions, that is, S1 to S8 in Figure 1. We can use eight datasets for different incident beams. Applying the data subtraction method, it is found that the detection probability is higher for an absorber located near an incident beam.

Figure 10 shows distributions of detection probabilities. In yellow domains near the beam position, the detection probabilities are high, while probabilities are low in blue domains. If we apply the present method to datasets for three incident beams simultaneously from S1, S2, and S4, the accuracy rate is 99.9% to detect one absorber. In addition, we tested the multi-beam model for three absorbers, using the data subtraction method. The results are shown in Figure 11, where each symbol corresponds to predicted positions for each beam. In this example, the blue absorber is correctly detected, while the orange absorber is correctly detected for one position by detector S1. The total score is 7, 4 for S2, 2 for S1 and 1 for S4. The maximum score is 12. Thus, the accuracy rate is $7/12 = 58\%$. Additional datasets for different beams will improve the detection accuracy rate.
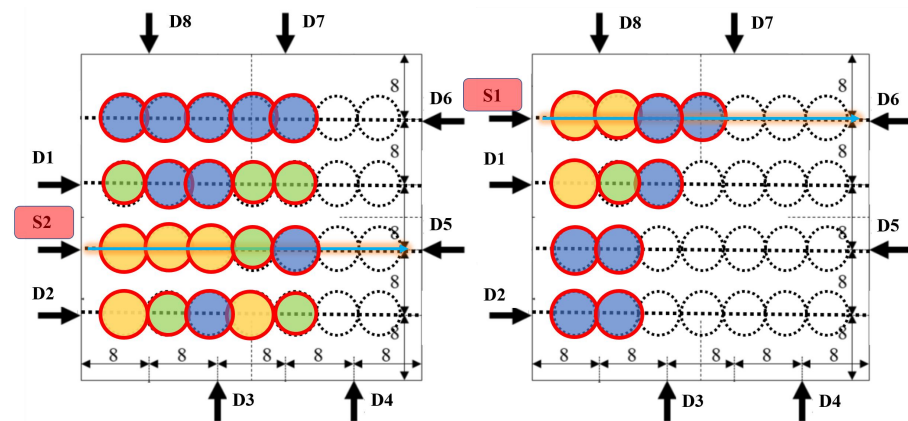
Figure 10. Distributions of detection probabilities for an incident beam from S2 (**left panel**) and S1 (**right panel**): yellow domains highlight most probable positions for detection, while green domains correspond to moderate probabilities and blue domains correspond to low probabilities.

Figure 11. Example for a multi-beam injection model. three absorbers, and three different incident beams. Each symbol corresponds to detected positions for each beam.

5. Discussion

5.1. Sampling Effect

In the analyses presented in this paper, we employed datasets of all time steps for deep learning. A time step in our simulations is $\Delta t = 0.0006$ ns, while it is $\Delta t = 0.01$ ns in experiments using the phantom. One advantage of LSTM deep learning is to recognize global behaviors of temporal data. Utilizing this advantage, we conducted the LSTM deep learning study using only 1/16 of the simulation snapshots (16 times longer time steps). As a result, we confirmed that the our LSTM model works well even for such coarse-grained data. This allows us to speed up the deep learning dramatically, by adjusting the time step of the simulations to that of phantom experiments.

In addition, in the present analysis, we have assumed noises of a Gaussian type for the training data. We have further investigated a different type of noise, which is a uniform one. As a result, we have confirmed that the training data with uniform noises give results similar to those with the Gaussian noises. Actually, if experiments are conducted many times, the noises in obtained data are expected to approach a Gaussian distribution. Indeed, the experimental data obtained in the phantom show Gaussian-like noises.

Finally, in this paper, we have used 2D simulation data calculated with the radiation transfer solver TRINITY [11]. We have compared the present results to 3D Monte Carlo simulations, and confirmed that the present 2D models show a good agreement with 3D Monte Carlo simulations.

5.2. Comparison with the Conventional DOT Method

Here, we compare the results by our deep learning method with the profile reconstructed by the conventional DOT method [17]. The differences emerge in two ways, that is, in solving the forward problem and the inverse problem, respectively. On the forward problem, the conventional DOT method [17] solves a diffusion equation to trace light propagation. On the other hand, we use RTE calculation that gives us more accurate data. On the inverse problem, one can reconstruct the image by estimating the mean, variance, and skew of DTOF in the conventional DOT method. However, it makes the reconstructed image unclear due to hyperparameter problems. Therefore, here, we use the deep learning method to solve the inverse problem.

The comparison between our deep learning method and the conventional DOT method is shown in Figure 12. The absorption coefficient and position reconstructed by the conventional method is shown by a pink line. This method gives a broader profile, compared to the true value. The absorption coefficients presumed by the deep learning are shown by a blue line. The blue lines show the primary candidate by the deep learning, where the absorption coefficient and position are well presumed compared to the true values. In the present deep learning, it is possible to presume the absorption coefficient and position more accurately.
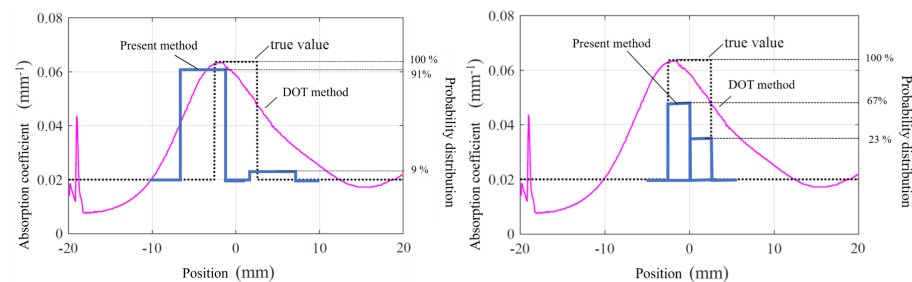
Figure 12. Comparison of the present method with the conventional DOT method. The true absorption coefficient is shown by the dotted line. The pink curve depicts the reconstructed coefficient using the conventional DOT method. The blue line represents the probability distributions of the absorption coefficients predicted by our deep learning method, where the setup is the same as Figure 1. The **left** (**right**) panel is used by resolution as 28 domains (168 domains).

We also compared a higher resolution model using 168 domains. In the right panel of Figure 12, the location of the absorber is identified by a sharp boundary without spatial spread, which comes from a hyperparameter in the conventional DOT method. Thus, our deep learning scheme can provide a more reliable reconstruction method compared to the conventional DOT method. In addition, see [13], which shows a possibility of giving a good reconstructed image since a similar situation can be improved.

We tested the RNN approach to classify the position. The model with RNN shows 20 percent lower accuracy for classification of 28 domains with the same setting. The LSTM approach has an advantage of analyzing data in chronological order. We also tested this RNN approach to predict for the same experimental data above used. The result tells us the wrong prediction that is caused by identifying some data in the wrong position. It is useless to reconstruct the image. The reason why the RNN gives a bad prediction is not good for analyzing long time sequence data, which has long term changes over time. It means that the RNN approach is not suitable for such time sequence. In Ref. [14], they showed the overview of DOT by using deep learning, and there is a good possibility to solve the inverse problem by using deep learning. They tested that propagation of light through the digital phantom and reconstructed image by using deep learning to solve the inverse problem. The result shows a remaining background noise on the reconstructed image. This is related to the fact that their algorithm has hyperparameters for both solving the inverse problem and reconstructing the image. On the other hand, to solve the inverse problem, we try to identify the position by using LSTM deep learning. Its process does

not basically contain a hyperparameter and such parameters can be fixed through every learning step, and it is possible to make the reconstructed image clear when one takes more high resolution. They also showed that the deep learning approach to solve the inverse problem is highly dependent on the type of datasets used. We used the time dependent data to reconstruct the absorber's position. The data we deal with have accurate time dependent information, and it is then good to analyze data by using the LSTM deep learning. The fact that LSTM obtains more information for identifying positions than the RNN leads to a good reconstruction shown as in Figure 12.

6. Conclusions

In this paper, as a novel deep learning scheme for DOT, we have applied a LSTM deep learning method to temporal absorption profiles of target tissue, which are obtained by directly solving the time-domain RTE.

We have two original contributions in this paper. One is using Trinity code to solve the RTE as accurate time dependent data and the second is using the LSTM deep learning with multi-step classification to solve the inverse problem. The classification of 28 domains increases its accuracy by two-step classification. The time dependent data used in this work have a lot of information, and using LSTM with multi-step classification and a data subtraction method is suitable to classify and identify such data.

On the phantom experiment, we have shown that positions of absorbers can be predicted with high accuracy rates by a multi-step classification method. We have also developed the data for the subtraction method to detect two absorbers or an extended absorber larger than a domain. Our results demonstrate that the application of LSTM deep learning in DOT allows us to detect absorbers without the need to solve an inverse problem. We have shown that positions of absorbers can be predicted with high accuracy rates by a multi-step classification method.

In addition, in applying our method for the data of phantom experiment data, it has been shown that the location of the absorber is identified by a sharp boundary without a spatial spread, which is a weak point in the conventional DOT method.

In this paper, we have shown the first attempt to draw a high absorption coefficient part by combining time-domain RTE and deep learning. In the future work, if we discriminate lower absorption coefficients at multiple wavelengths, then we can convert them into Hb concentrations.

Author Contributions: Methodology, H.Y. and M.A.; writing—review & editing, Y.T.; supervision, M.U. and Y.H. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there are no conflict of interest related to this article.

References

1. Hoshi, Y.; Yamada, Y. Overview of diffuse optical tomography and its clinical applications. *J. Biomed. Opt.* **2016**, *21*, 091312. [CrossRef] [PubMed]
2. Klose, A.D. The forward and inverse problem in tissue optics based on the radiative transfer equation: A brief review. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 1852–1853. [CrossRef] [PubMed]
3. Klose, A.D.; Netz, U.; Hielscher, A.H. Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer. *Med. Phys.* **1999**, *26*, 1698–1707. [CrossRef] [PubMed]

4. González-Rodríguez, P.; Kim, A.D. Comparison of light scattering models for diffuse optical tomography. *Opt. Express* **2009**, *17*, 8756–8774. [CrossRef] [PubMed]
5. Wang, L.V.; Wu, H.-I. *Biomedical Optics*; Wiley: Hoboken, NJ, USA, 2007.
6. Pogue, B.W.; Testorf, M.; McBride, T.; Osterberg, U.; Paulsen, K. Instrumentation and design of a frequency-domain diffuse optical tomography imager for breast cancer detection. *Opt. Express* **1997**, *1*, 391–403. [CrossRef] [PubMed]
7. Kim, H.K.; Hielscher, A.H. A diffusion-transport hybrid method for accelerating optical tomography. *J. Innov. Opt. Health Sci.* **2010**, *3*, 1–13. [CrossRef]
8. Tarvainen, T.; Kolehmainen, V.; Arrdige, S.R.; Kaipio, J.P. Image reconstruction in diffuse optical tomography using the coupled radiative transport-diffusion model. J. Quant. Spectrosc. *Radiat. Transf.* **2011**, *112*, 2600–2608. [CrossRef]
9. Ren, K.; Bal, G.; Hielscher, A.H. Frequency Domain Optical Tomography Based on the Equation of Radiative Transfer. *SIAM J. Sci. Comput.* **2006**, *28*, 1463–1489. [CrossRef]
10. González-Rodríguez, P.; Kim, A.D. Diffuse optical tomography using the one-way radiative transfer equation. *Biomed. Opt. Express* **2015**, *6*, 2006–2021. [CrossRef] [PubMed]
11. Yajima, H.; Abe, M.; Umemura, M.; Takamizu, Y.; Hoshi, Y. TRINITY: A three-dimensional time-dependent radiative transfer code for in-vivo near-infrared imaging. *J. Quant. Spectrosc. Radiat. Transf.* **2022**, *277*, 107948. [CrossRef]
12. Iliev, I.T.; Ciardi, B.; Alvarez, M.A.; Maselli, A.; Ferrara, A.; Gnedin, N.Y.; Mellema, G.; Nakamoto, T.; Norman, M.L.; Razoumov, A.O.; et al. Cosmological radiative transfer codes comparison project - I. The static density field tests. *Mon. Not. R. Astron. Soc.* **2006**, *371*, 1057–1086. [CrossRef]
13. Feng, J.; Sun, Q.; Li, Z.; Sun, Z.; Jia, K. Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography. *J. Biomed. Opt.* **2018**, *24*, 051407. [CrossRef] [PubMed]
14. Balasubramaniam, G.M.; Wiesel, B.; Biton, N.; Kumar, R.; Kupferman, J.; Arnon, S. Tutorial on the Use of Deep Learning in Diffuse Optical Tomography. *Electronics* **2022**, *11*, 305. [CrossRef]
15. Yoo, J.; Sabir, S.; Heo, D.; Kim, K.H.; Wahab, A.; Choi, Y.; Lee, S.I.; Chae, E.Y.; Kim, H.H.; Bae, Y.M.; et al. Deep Learning Diffuse Optical Tomography. *IEEE Trans. Med. Imaging* **2020**, *39*, 877–887. [CrossRef] [PubMed]
16. Fan, Y.; Ying, L. Solving optical tomography with deep learning. *arXiv* **2019**, arXiv:1910.04756.
17. Mimura, T.; Okawa, S.; Kawaguchi, H.; Tanikawa, Y.; Hoshi, Y. Imaging the Human Thyroid Using Three-Dimensional Diffue Optical Tomography. *Appl. Sci.* **2021**, *11*, 1670. [CrossRef]