



Article Improvement of the Performance of Scattering Suppression and Absorbing Structure Depth Estimation on Transillumination Image by Deep Learning

Ngoc An Dang Nguyen ^{1,2,†}, Hoang Nhut Huynh ^{1,2,†} and Trung Nghia Tran ^{1,2,*}

- ¹ Laboratory of Laser Technology, Faculty of Applied Science, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City 72506, Vietnam; dnnan.sdh19@hcmut.edu.vn (N.A.D.N.); hhnhut@hcmut.edu.vn (H.N.H.)
- ² Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc, Ho Chi Minh City 71308, Vietnam
- * Correspondence: ttnghia@hcmut.edu.vn
- These authors contributed equally to this work.

Abstract: The development of optical sensors, especially with regard to the improved resolution of cameras, has made optical techniques more applicable in medicine and live animal research. Research efforts focus on image signal acquisition, scattering de-blur for acquired images, and the development of image reconstruction algorithms. Rapidly evolving artificial intelligence has enabled the development of techniques for de-blurring and estimating the depth of light-absorbing structures in biological tissues. Although the feasibility of applying deep learning to overcome these problems has been demonstrated in previous studies, limitations still exist in terms of de-blurring capabilities on complex structures and the heterogeneity of turbid medium, as well as the limit of accurate estimation of the depth of absorptive structures in biological tissues (shallower than 15.0 mm). These problems are related to the absorption structure's complexity, the biological tissue's heterogeneity, the training data, and the neural network model itself. This study thoroughly explores how to generate training and testing datasets on different deep learning models to find the model with the best performance. The results of the de-blurred image show that the Attention Res-UNet model has the best de-blurring ability, with a correlation of more than 89% between the de-blurred image and the original structure image. This result comes from adding the Attention gate and the Residual block to the common U-net model structure. The results of the depth estimation show that the DenseNet169 model shows the ability to estimate depth with high accuracy beyond the limit of 20.0 mm. The results of this study once again confirm the feasibility of applying deep learning in transmission image processing to reconstruct clear images and obtain information on the absorbing structure inside biological tissue. This allows the development of subsequent transillumination imaging studies in biological tissues with greater heterogeneity and structural complexity.

Keywords: point spread function (PSF); de-blurring; scattering suppression; depth estimation; Attention Res-Unet; DenseNet169; absorbing structure; turbid medium

1. Introduction

Optical imaging is crucial in biomedical research and diagnostics, bridging pre-clinical and clinical applications. The potential of light, especially near-infrared light, for imaging blood vessels on the skin surface and abnormal breast detection has been recognized in studies focusing on bio-metric and medical applications [1–4]. The prospects for developing non-invasive imaging devices based on near-infrared light are promising, offering advantages such as the absence of ionizing radiation, cost-effectiveness compared to existing methods, and suitability for further studies. However, transillumination images face strong scattering challenges. Previous research focused on the suppression of scattering and the



Citation: Dang Nguyen, N.A.; Huynh, H.N.; Tran, T.N. Improvement of the Performance of Scattering Suppression and Absorbing Structure Depth Estimation on Transillumination Image by Deep Learning. *Appl. Sci.* 2023, *13*, 10047. https://doi.org/ 10.3390/app131810047

Academic Editor: Cosimo Nardi

Received: 6 August 2023 Revised: 30 August 2023 Accepted: 1 September 2023 Published: 6 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). restoration of clear images from blurred images [5–12]. Optical computed tomography (OCT) utilizing near-infrared light has been proposed and has shown satisfactory results in small animal imaging [7]. Deep learning (CNN) and stacking methods were proposed that were used to estimate the depth and de-blurring transillumination images of a turbid medium [8–12]. The effectiveness of previous studies is limited to a depth of absorbing structure shallower than 15.0 mm [10,12].

This study's models are based on novel machine learning mechanisms that combine different types of neural networks and sparse coding techniques to achieve high-quality image super-resolution [13]. The proposed models are also capable of handling multimodal and cross-domain image processing tasks, such as enhancing images from different sources or modalities, transferring styles or attributes between images, or generating realistic images from sketches or text descriptions [14]. The proposed models are inspired by some of the recent advances in machine learning algorithms and mechanisms for image processing, as well as some of the applications of image processing techniques for machine learning [15]. This study proposes new deep learning models to improve absorbing structures' de-blurring and depth estimation. The following sections of this paper will provide detailed information about the training dataset, the model employed for the de-blurring and depth estimation of absorbing structures, the performance parameters of the training process, and results and discussions concerning the de-blurring and depth estimation of absorbing structures.

2. Materials and Methods

2.1. Data Preparation

The deep learning model requires numerous training pairs for optimal accuracy and performance. The de-blurring process involved working with a carefully curated dataset of blurred and original clear images. The corresponding depth labels associated with the blurry images were used to train the depth estimation model. However, data collection presented practical challenges in acquiring significant training pairs. The depth-dependent point spread function (PSF), which characterizes light scattering in biological tissue at different depths, was implemented to convolve the original clear images, generating the desired blurred images to overcome this limitation.

Figure 1 shows the difference between fluorescent and transillumination images with the assumption that the light diffused well in the absorbing object plan. In fluorescent imaging, the light point source is placed inside the scattering medium, as shown in Figure 1a, and the light distribution on the observing surface (dashed orange line) can be mathematically represented by Equation (1) [6]:

$$PSF(d,\rho) = C\left(\mu'_{s} + \mu_{a} + \left[\kappa_{d} + \frac{1}{\sqrt{\rho^{2} + d^{2}}}\right]\frac{d}{\sqrt{\rho^{2} + d^{2}}}\right)\left(\frac{\exp[-\kappa_{d}\sqrt{\rho^{2} + d^{2}}]}{\sqrt{\rho^{2} + d^{2}}}\right)$$
(1)

where $k_d^2 = 3\mu_a(\mu'_s + \mu_a)$. *C*, μ'_s , μ_a , and *d* represent the constants with respect to ρ and *d*, the reduced scattering coefficient, the absorption coefficient, and the depth of the light source, respectively.

In transillumination imaging, the light source is placed outside the scattering medium, as shown in Figure 1b, and the light distribution on the observing surface (black line) is a collection of the distribution of the light-missing points. The depth-dependent PSF is derived from a light source, so we cannot apply it directly for transillumination imaging. To make the PSF applicable, we invert the distribution of the light-missing point (black line) to have the distribution the same as the distribution of light in fluorescent imaging (dashed orange line), as shown in Figure 1b.



Figure 1. The intensity distribution at the medium surface in fluorescent (**a**) and transillumination (**b**) imaging.

The effectiveness of this approach was rigorously evaluated through comprehensive simulations and experimental validations [7,10,12]. Convolution images of the original structures with depth-dependent point spread functions were used at different depths to generate the data in this study, as described in Equation (2) and Figure 2.

$$y = h \otimes x \tag{2}$$

where \otimes denotes convolution operation.



Figure 2. Blurred image generation process.

In this study, the original structure images are images obtained from the image of 12 randomized structures in a transparent medium. These pre-blurred images are specifically designed to emulate the intricate structural characteristics of blood vessels beneath the skin. The blurred images were generated by convoluting the original structures with the PSF given by Equation (1) at different depths with the parameter values of $\mu'_s = 1.0 \text{ mm}^{-1}$ and $\mu_a = 0.00536 \text{ mm}^{-1}$. These parameters were used in all simulations described here-inafter. These coefficients are pivotal in understanding light–tissue interactions, particularly in optical imaging and spectroscopy. μ'_s signifies the extent of light scattering as it traverses tissue, indicating the likelihood of scattering per unit path length. This is crucial because of diverse tissue structures that cause light to scatter in various directions. μ_a gauges the extent of light absorbed by tissue during propagation, closely related to light-absorbing constituents such as hemoglobin, lipids, and water. Different tissues and substances possess

varying absorption characteristics at different wavelengths. In the given context, specific μ'_s and μ_a values might be set to simulate tissue optical properties in a model. These values depend on the type of tissue, the wavelength of light, and the experimental conditions. Setting these values likely aims to ensure that simulations closely mimic real tissue optical behavior under specific circumstances. Although the specific rationale hinges on precise values, the choice of these parameters is vital to accurately model light–tissue interactions and align simulation outcomes with experimental data.

In studies involving optical imaging and simulations of light propagation in biological tissues, setting the values of μ'_s (reduced scattering coefficient) and μ_a (absorption coefficient) is a critical step. Researchers usually consider the following factors when determining μ'_s and μ'_a values:

- Empirical data: Experimental measurements of optical properties in specific tissue types at various wavelengths can serve as a foundation for determining the appropriate values. These measurements can come from the literature or new measurements conducted by the researchers themselves.
- Literature references: Previous studies often report ranges or specific values of µ_s and µ_a for similar tissue types. Researchers can use these references as a starting point and adjust the values based on their experimental setup.
- Theoretical models: There are established theoretical models that relate optical properties to tissue composition and structure. Researchers can leverage these models to estimate μ_s and μ_a based on the known components and concentrations in the tissue.
- Tissue variation: Different tissues exhibit different optical properties as a result of variations in cellular composition, structure, and pigmentation. Consequently, the specific tissue under investigation must be carefully considered when selecting μ_s and μ_a values.
- Wavelength Dependence: Optical properties can vary with the wavelengths of light. Researchers may choose μ_s and μ_a values that align with the wavelength range used in their experimental setup.
- Validation: Validating the chosen values involves comparing the simulation results with actual experimental observations. If the simulated outcomes closely match the experimental data, this provides confidence in the suitability of the parameter values.
- Sensitivity analysis: Researchers may conduct sensitivity analyses to assess how changes in μ_s and μ_a impact simulation results. This analysis helps to determine reasonable ranges for these parameters.

In essence, substantiating the selections of μ_s and μ_a values typically entails a combination of empirical data, theoretical frameworks, references from the literature, and validation against experimental results. The specific strategy can be flexible to existing resources, unique tissue attributes, and the specific objectives of the experiment.

For the de-blurring study, a comprehensive dataset consisting of 8000 pairs of clear and blurred images was generated by convoluting 10 of 12 original structures with the PSF given by Equation (1) at depths ranging from 0.1 to 20.0 mm (interval 0.1 mm) and then rotating at four different angles, as illustrated in Table 1. The remaining 2 of the 12 original structures were used to generate data for testing. During the training process for de-blurring, the generated dataset was used to train the models with a batch size of 8. The learning rate was set to 10^{-4} , and the models were trained for 100 epochs.

For the depth estimation study, the corresponding depth labels associated with the blurry images were used. A dataset of 70,400 images was generated that depicts the absorbing structures within the scattering medium at different depths. The blurred images in this dataset were generated by convoluting 11 of the 12 original structures with the PSF given by Equation (1) at depths ranging from 0.5 to 20.0 mm (interval 0.5 mm) and then rotating at 160 different angles, as illustrated in Table 1. The remaining original structure was used to generate data to test the performance of the convolution neural network models. During the depth estimation training process, the generated dataset was used to

train the models with a batch size of 32. The learning rate was set to 10^{-4} , and the models were trained for 20 epochs.

Table 1. Dataset for training, validation, and testing of scattering de-blurring and depth estimation.

Model	Training	Validation	Testing	Total
De-blurred	5600	1600	800	8000
Estimate depth	56,320	14,080	7040	70,400

Training was carried out on a high-performance workstation that features two Intel[®] Xeon[®] CPUs E5-2683 v4 with 64 GB of RAM. In addition, an NVIDIA Quadro K2200 graphics processing unit was used to accelerate the computational tasks involved in the training process. The specific training parameters used, including batch size, learning rate, and number of epochs, are provided in Table 2.

Table 2. Parameters for de-blurred and depth estimation models.

Parameters	De-Blurred	Depth Estimation
μ'_s	$1.0~\mathrm{mm}^{-1}$	$1.0~\mathrm{mm}^{-1}$
μ_a	0.00536 mm^{-1}	$0.00536 \ { m mm}^{-1}$
d_{\min} - d_{\max}	0.1–20.0 mm	0.5–20.0 mm
Step depth	0.1 mm	0.5 mm
Batch size	8	32
Learning rate	10^{-4}	10^{-4}
Epoch	100	20
Loss function	Dice-coef loss	Categorical Cross-entropy
Optimizer	Adam	Adam
Input shape	256 imes 256 imes 1	224 imes 224 imes 1

2.2. Image De-Blurring

Transillumination imaging techniques for visualizing absorbing structures within the body often encounter blurring. While both scattering and absorption contribute to image blurring, scattering plays a dominant role. Many research efforts have been made to overcome this challenge. In the previous studies of the group, the suppression of the scattering effects on the transillumination image was carried out by deconvolution with the depth-dependent PSF, and the deep learning scatter blurring method has also proved to be feasible and efficient [5–7,10,12]. However, these methods still have some limitations in implementation, such as the imperfection of the deconvolution technique, the long computation time, the computational hardware requirements, and the limitation of effective de-blurring shallower than 15.0 mm. In the previous study, we employed fully convolutional networks (FCN) based on the U-net with skip connections. The training process for the scattering de-blurring model is visually represented in Figure 3, offering a clear visualization of the methodology employed. The results show that we can obtain a clear image of the absorbing structure as deep as several to 10.0 mm in a turbid medium.

To address this challenge, the Attention U-Net model and the Attention Res-UNet model were incorporated for the de-blurring process [16,17]. The attention gate is a mechanism that selectively emphasizes specific regions of interest while suppressing the activation of irrelevant regions on a given input feature map *X*. To achieve this, the attention gate takes advantage of a gating signal $G \in \mathbb{R}^{C' \times H \times W}$, which is obtained at a coarser scale and incorporates contextual information. When additive attention is employed, the attention gate calculates the gating coefficient. Initially, both the input *X* and the gating signal *G* undergo linear mapping to a $\mathbb{R}^{F \times H \times W}$ dimensional space. Subsequently, the output is compressed in the channel domain to generate a spatial attention weight map $S \in \mathbb{R}^{1 \times H \times W}$, as shown in Figure 4. The entire process can be formulated as described in Equation (3)

and Equation (4) [18], where ϕ , ϕ_x , and ϕ_g are linear transformations implemented as convolution 1×1 . S

Υ

$$=\sigma(\varphi(\delta(\phi_x(X) + \phi_g(G))))$$
(3)

$$=SX$$
 (4)



Figure 3. De-blurring using a deep learning model: (a) training process with pair of images before and after blurring, and (b) image de-blurring process.



Figure 4. The diagram of the attention gate.

Residual blocks, which are skip-connection blocks, are designed to learn residual functions by referring to the layer input instead of learning unreferenced functions. These blocks were originally introduced as a component of the Res-Net architecture. In a formal sense, denoting the desired underlying mapping as H(x), the stacked non-linear layers aim to approximate an additional mapping that captures the difference between the current output and the input, denoted as Equation (5). By explicitly modeling the residual mapping, the network can effectively learn residual functions and enhance optimization.

$$F(x) := H(x) - x \tag{5}$$

The original mapping is reformulated as F(x) + x, where F(x) + x represents a residual component, thus giving rise to the term "residual block H(x)". The rationale behind this approach lies in the observation that optimizing the residual mapping is often more feasible than optimizing the original, unreferenced mapping. In certain cases, minimizing the residual to approach zero can be simpler than fitting an identity mapping using a series of non-linear layers. The network is better equipped to learn mappings that resemble identity transformations by incorporating skip connections. The proposed framework encompasses a novel deep learning architecture known as Res-UNet-a and a novel loss function based on the Dice loss. Res-UNet-a combines a U-Net encoder/decoder backbone with residual connections, Atrous convolutions, pyramid scene parsing pooling, and multitasking inference, thus enhancing its capabilities for various image analysis tasks, as shown in Figure 5.



Figure 5. A diagram of the residual block.

2.3. Depth Estimation

In transillumination imaging, the extent of blurring depends on the depth of the absorbing structure within the scattering medium, with an increase in depth resulting in a progressively more blurred image. To estimate the depth of the absorbing structure, a convolutional neural network (CNN) model is trained using generated blurred images. We used Res-Net-based convolutional neural networks (CNN) in the previous study. The training process for the depth estimation model is visually represented in Figure 6, which provides a clear visualization of the methodology used. The results show that we can effectively estimate the depth of the absorbing structure as deep as several to 10.0 mm in a turbid medium. Four pre-trained models, namely Res-Net50, VGG-16, VGG-19, and Dense-Net169 [19,20], were used for the depth estimation challenge. The images were paired with their respective depth labels during the training phase. An estimate of the depth of the absorbing structure was obtained by entering a blurred image into the CNN model. This process aligns with the fundamental classification task within deep learning. To ensure consistency, the training performance settings described in Table 2 were applied in different models, taking into account computational constraints and system compatibility. Figure 6 illustrates the estimation procedure for the depth of the absorbing structure using the CNN model.



Figure 6. Estimating depth of absorbing structure with deep learning model: (**a**) training process and (**b**) depth estimating process.

3. Metrics

The Dice coefficient is widely used to assess the agreement at the pixel level between a predicted segmentation and its corresponding ground truth. It quantifies the similarity by calculating twice the area of overlap divided by the sum of the total number of pixels in both images. Equation (6) expresses the Dice coefficient as [21]:

$$\text{Dice-coef} = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{6}$$

where *X* and *Y* represent the predicted set of pixels and the ground truth, respectively.

Moreover, the loss of the Dice coefficient is employed as a measure of dissimilarity between the predicted and ground-truth segmentation. It is computed by subtracting the Dice coefficient from 1. Equation (7) presents the formulation of the loss of the Dice coefficient [21].

Dice-coef loss =
$$1 - \frac{2 \times |X \cap Y|}{|X| + |Y|}$$
(7)

The Intersection over Union (IoU) is commonly utilized as an evaluation metric for object detection accuracy in the dataset by calculating the ratio of the overlap and the union areas between the predicted and ground-truth regions. Equation (8) [22] represents the IoU formula as:

$$IoU = \frac{Area \text{ of overlap}}{Area \text{ of union}}$$
(8)

When developing a depth estimation model, accuracy is used as a classification metric to measure the proportion of correctly predicted instances where the predicted depth exceeds the actual depth. It provides insight into the model's performance on the dataset. The accuracy is computed by dividing the sum of the True Negatives (*TN*) and True Positives (*TP*) by the total number of samples. Equation (9), illustrates the accuracy formula [23]:

$$Accuracy = \frac{TP + TN}{\text{total sample}}$$
(9)

4. Results and Discussion

4.1. De-Blurring Image

The significant impact of the attention gate on the scattering de-blurring process is demonstrated by the results presented in Figure 7. The images obtained with the Attention Unet model (D) show higher clarity and fidelity than those obtained with the standard U-Net architecture (C), as the attention mechanism effectively suppresses the scattering influence and improves the image reconstruction of the absorbing structure. Quantitative evaluation, as indicated by the Intersection Over Union (IoU) index, further supports the superiority of the gating attention approach. The IoU index of 0.908 achieved when using gating attention exceeds the IoU index of 0.831 obtained with the standard U-Net architecture. This substantial improvement demonstrates the ability of the attention gate to capture the relevant features better and reduce the impact of scattering, leading to more accurate and precise de-blurring results.



Figure 7. Scattering de-blurring with and without attention gate: (**A**) transillumination image through turbid medium, (**B**) image taken through clear water, (**C**) output image from U-Net model, and (**D**) output image from Attention Unet model.

The effectiveness of the attention gate can be attributed to its ability to selectively focus on informative regions and suppress the interference caused by scattering. By assigning different attention weights to different parts of the image, the attention gate improves the model's capability to accurately capture and reconstruct the absorbing structure image, even at greater depths. These findings highlight the potential of the attention gate in improving the scattering de-blurring process. Incorporating the gating attention mechanism into the U-Net architecture can significantly enhance the quality and reliability of de-blurred images, particularly in scenarios with high levels of scattering. Further exploration and optimization of the attention gate in various imaging applications hold promise for advancing the image reconstruction and de-blurring field.

The effectiveness of the residual block in the scattering de-blurring process is illustrated by the results presented in Figure 8. The images obtained from the Res-UNet model show a remarkable improvement in the de-blurring outcome compared to those obtained from the U-Net model, as evidenced by the higher IoU index of 0.885. This indicates a more accurate reconstruction of the original absorber image at a depth of 15 mm, even in the presence of scattering and blurring effects. In contrast, the standard U-Net architecture yields a slightly lower IoU index of 0.831, indicating a relatively inferior de-blurring performance. The superior performance of the Residual U-Net model can be attributed to the ability of residual blocks to facilitate the propagation of gradient information effectively. By allowing for the direct flow of information through skip connections, residual blocks enable the model to capture and restore important features of the absorbing image more efficiently. Consequently, the Residual U-Net model exceeds the standard U-Net architecture in mitigating the negative impact of scattering and achieves more accurate de-blurring results. These findings demonstrate the significance of incorporating residual blocks into deep learning models for scattering de-blurring tasks. The Residual U-net model proves to be a promising approach for addressing challenges associated with image blurring in the presence of scattering media. Further investigations and optimizations can be conducted to enhance the performance of the Residual U-Net model and explore its

potential applications in various imaging tasks, such as medical diagnostics and image analysis in turbid environments.



Figure 8. Scattering de-blurring with and without residual block: (**A**) transillumination image through turbid medium, (**B**) image taken through clear water, (**C**) output image from U-Net model, and (**D**) output image from Residual Unet model.

The primary objective of this study was to obtain de-blurred images by effectively compensating for scattering effects. To accomplish this, the models employed in this approach included Attention U-Net and Attention Res-UNet. These models were trained using a carefully curated input and output image pairs dataset. A combination of PSF convolutions at different depths was used to train the Attention U-Net and Attention Res-UNet networks for image de-blurring. This involved pairing the original images with their corresponding blurred counterparts. This approach enabled the networks to learn the intricate relationships between different depths and their corresponding blurred representations, facilitating accurate image de-blurring.

In the statistics table of the Dice coefficient for the two models, Attention Unet and Attention Res-UNet, as shown in Table 3, we can observe crucial information on the performance of these models. The Attention Unet model achieved a minimum Dice coefficient of 0.931 and a maximum of 0.999487, with a mean of 0.996319 and a median of 0.999195. The variability in the performance of this model is represented by a standard deviation of 0.009583. Similarly, the Attention Res-UNet model exhibits comparable parameters, with a minimum Dice coefficient of 0.930391 and a maximum of 0.999492. The mean and median values for this model are 0.996443 and 0.999223, respectively. The performance variability of the Attention Res-UNet model is gauged by a standard deviation of 0.009603. Overall, both models demonstrate consistent performance with minimal variation across the Dice coefficient values. This underscores the efficiency and general applicability of the models in de-blurring absorption structures within a dispersed medium. Figure 9 provides a visual representation of the process.

Table 3. Performance Comparison of Attention Unet and Attention Res-UNet Models based on Dice

 Coefficient Statistics.

Model	Minimum	Maximum	Mean	Median	Std
Attention Unet	0.931056	0.999487	0.996319	0.999195	0.009583
Attention Res-UNet	0.930391	0.999492	0.996443	0.999223	0.009603

Figure 10 illustrates a representative example of input and output images obtained by scatter blurring at various depths, specifically 0.1, 5.0, 10.0, and 20.0 mm. In particular, the corresponding correlation indices for these depths were reported as 0.9360, 0.9167, 0.9130, and 0.9059, respectively. These correlation indices served as valuable quantitative indicators, providing insights into the level of agreement between the predicted de-blurred output and the ground truth images.



Figure 9. Training and validation for de-bluring process: (**A**,**B**) Attention U-Net, (**C**,**D**) Attention Res-UNet.



Figure 10. Representative images demonstrating the de-blurring process at various depths: (**A**) 0.1 mm, (**B**) 5 mm, (**C**) 10 mm, and (**D**) 20 mm.

Figure 11 illustrates the original and restored images of the absorbing structure after applying the Attention U-Net and Attention Res-UNet models. Then, the correlation coefficient is calculated. As the depth of the absorbing structure increases, the blurring effect becomes more pronounced, leading to a rapid decline in the quality of the blurred image. Furthermore, the reduction in training images significantly affects the correlation coefficient.



Figure 11. Correlation analysis between original and deblurred images with 256×256 input image size.

The test results for the Attention U-Net model indicated that the correlation coefficient exhibited a high value ranging from 0.9149 to 0.9013 for depths between 0.1 and 5.0 mm. Beyond 5.0 mm, the correlation coefficient gradually decreased, reaching 0.8921 at a depth of 10.0 mm. Subsequently, for depths ranging from 10.1 to 20.0 mm, the correlation coefficient rapidly decreased from 0.8918 to 0.8801 at a depth of 20.0 mm. In particular, the rate of decrease in the correlation coefficient became more pronounced once the depth exceeded 10.1 mm.

Similarly, the Attention Res-UNet model yielded test results indicating a high correlation coefficient ranging from 0.9308 to 0.9069 for depths between 0.1 and 5.0 mm. Beyond 5.0 mm, the correlation coefficient gradually decreased, reaching 0.8845 at a depth of 14.0 mm. In particular, for depths greater than 10.0 mm, the rate of decrease in the correlation coefficient increased. Finally, for depths ranging from 14.1 to 20.0 mm, the correlation coefficient exhibited a rapid decline from 0.8942 to 0.8876 at a depth of 20.0 mm. Once again, the rate of decrease in the correlation coefficient method are in the correlation coefficient decreased rapidly for depths exceeding 14.1 mm.

In the subsequent experiment, the size of the training input image was modified from 256×256 pixels to 112×112 pixels while keeping the other training parameters in Table 2 unchanged. The results obtained from this adjustment are depicted in Figure 12.



Figure 12. Correlation analysis between original and deblurred images with 112×112 input image size.

The test results of the Attention U-Net model revealed that within the depth range of 0.1 to 5.0 mm, the correlation coefficient initially reached a high value and gradually decreased from 0.9186 to 0.8722. The highest correlation coefficient was achieved at a depth of 0.1 mm, registering a value of 0.9186. As the depth increased from 5.1 to 10.0 mm, the correlation coefficient experienced a gradual decrease, reaching 0.8184 at a depth of 10.0 mm. Subsequently, for depths ranging from 10.1 mm to 20.0 mm, the correlation coefficient exhibited a rapid drop from 0.8172 to 0.6927 at a depth of 20.0 mm. Remarkably, once the depth surpassed 7.0 mm, the rate of decline in the correlation coefficient with respect to depth became more pronounced.

Similarly, the Attention Res-UNet model yielded noteworthy test results. At depths ranging from 0.1 to 5.0 mm, the correlation coefficient reached a high value and gradually decreased from 0.9337 to 0.9023. The highest correlation coefficient was observed at a depth of 0.1 mm, yielding a value of 0.9337. For depths extending from 5.1 to 14.0 mm, the correlation coefficient exhibited a gradual decrease from 0.9080 to 0.8736 at a depth of 14.0 mm. In particular, depths greater than 10.0 mm experienced an accelerated decline in the correlation coefficient. Finally, within the depth range of 14.1 to 20.0 mm, the correlation coefficient decreased rapidly from 0.8717 to 0.8539 at a depth of 20.0 mm. Once the depth surpassed 11.6 mm, the rate of decrease in the correlation coefficient with respect to depth decreased rapidly.

For the Attention U-Net model, it was observed that at depths ranging from 0.1 to 0.5 mm, employing an input size of 112×112 pixels yielded better performance, with a difference in the correlation coefficient ranging from 0.6% to 0.9%. On the contrary, at depths ranging from 0.6 to 5.0 mm, adopting an input size of 256 × 256 pixels achieved superior performance, exhibiting a difference in the correlation coefficient ranging from 0.02% to 3.08%. In particular, for depths ranging from 5.1 to 20.0 mm, the difference in performance between the two input sizes increased rapidly, ranging from 3.35% to 20.91%.

For the Attention Res-Unet model, it was observed that at depths ranging from 0.1 to 0.9 mm, employing an input size of 112 × 112 pixels resulted in improved performance, with a difference in the correlation coefficient ranging from 0.4% to 0.6%. On the other hand, at depths ranging from 1.0 to 5.0 mm, adopting an input size of 256 × 256 pixels yielded better performance, exhibiting a difference in the correlation coefficient index ranging from 0.15% to 0.75%. Furthermore, for depths ranging from 5.1 to 20.0 mm, the difference in performance between the two input sizes increased rapidly, ranging from 0.7% to 3.55%.

The results show the impact of resizing the input training image from 256×256 pixels to 112×112 pixels on the correlation coefficient at different depths. These findings demonstrate the importance of optimizing the input image size to achieve optimal performance in terms of the correlation coefficient at different depths, as shown in Figure 13. The observed trends can be ascribed to the interplay of scattering phenomena and the depth of the absorbing structure. With increasing depth, the scattering effects intensified, leading to diminished correlation coefficients. Moreover, the selection of the input image size exerted a notable influence on performance, primarily by affecting the model's ability to capture intricate features amidst scattering influences. In particular, the optimal input size exhibited variability depending on depth, thus facilitating improved adaptability to varying degrees of scattering. These discernments underscore the imperative of factoring in depth and input size while addressing scattering-induced de-blurring tasks, thereby providing valuable insights for optimizing model efficacy across diverse scenarios. Further studies could delve into the intricate dynamics connecting depth, scattering effects, and input size, thereby advancing the potential for refining the applicability and precision of de-blurring models.

The validity of the diffusion approximation is based on the condition that the thickness of the scattering medium is significantly greater than the average free-path length of $1/\mu'_s$. Consequently, caution must be exercised when applying Equation (1) in cases where $\sqrt{\rho^2 + d^2}$ is not greater than $1/\mu'_s$. As shown in Figure 1, the observing plan is considered to be significantly larger than the light distribution on the surface. Therefore, it is better to generate an appropriate wide image for training to ensure a result with a deep-absorbing structure. The light distribution on the surface has a Gaussian distribution shape. The image size in a dimension should be more significant than three times the standard deviation of the light distribution on the surface of the medium when calculating the deepest light point source distribution by Equation (1) in the turbid medium, as shown in Figure 13.



Figure 13. Optimizing input image size across various depths.

Depth = 20.0 mm

4.2. Depth Estimation

Depth estimation is an essential task for analyzing the properties of absorption structures. This section presents a deep learning approach to estimate the depths of absorption structures from their images. For this purpose, a dataset of 7040 images of absorption structures, each labeled with one of 40 depth values ranging from 0.1 mm to 20 mm, is used. Four state-of-the-art deep learning models, namely ResNet50, VGG16, VGG19, and DenseNet169, are trained and evaluated in this dataset. Accuracy is the evaluation metric that measures the percentage of images whose depth labels are correctly predicted by the models. Table 4 shows the training and validation accuracy of each model after 20 epochs.

Table 4. The training and validation accuracy of each model after 20 epochs.

Model	Training Accuracy	Validation Accuracy
ResNet50	0.4312	0.3921
VGG16	0.5124	0.4678
VGG19	0.4894	0.4500
DenseNet169	0.7323	0.6250

Table 4 shows that DenseNet169 achieves the highest accuracy in both the training and validation sets, followed by VGG16, VGG19, and ResNet50. All models perform better than in previous experiments with a smaller dataset, indicating the positive impact of dataset size and diversity on model performance. However, the accuracy of all models is still low, indicating the difficulty of the depth estimation task. To further analyze the behavior of the models, the accuracy curves of each model were plotted during training and validation, as shown in Figure 14.



Figure 14. Accuracy evaluation of various models: ResNet50, VGG16, VGG19, and DenseNet169.

In Figure 14, the conspicuous features include low accuracy values and pronounced fluctuations, indicating a struggle of the models to glean effective insights from the original 7040-image dataset. This conundrum can be attributed to the inherent complexity of the dataset, characterized by an extensive array of depth classes (40) coupled with a limited count of images per class (fewer than 176 images). Consequently, the models struggled to discern nuanced differentiators across various depth levels, impairing their capacity for comprehensive learning. The fluctuations in accuracy, evident in the jagged trajectory after each epoch, underscored the models' susceptibility to data fluctuations, amplifying the instability quotient.

To improve the performance of the models, data augmentation techniques were applied to increase the size and diversity of the dataset. Specifically, angle rotation was used to generate new images from existing ones by randomized 160 different angles and rotating them at angles between 0 and 360 degrees. This resulted in an augmented dataset of 70,400 images (7040×10) with the same depth labels as before.

The decision to employ 160 different angles for image rotation during the data generation process in this study is purposeful and aligned with the goal of improving the robustness and generalization capabilities of the trained convolutional neural network (CNN) models. This technique, commonly referred to as data augmentation, serves to simulate varying viewpoints and orientations of the same scene or object, thereby aiding the model in comprehending and identifying features from diverse angles. In the context of estimating depth from blurred images of absorbing structures within a scattering medium, the rationale behind incorporating image rotation at numerous angles can be succinctly summarized:

- Increased variability: By generating images from multiple angles, the dataset gains greater diversity. This variability acts as a defense against overfitting, ensuring that the model learns broader transferable features instead of memorizing specific training samples.
- Robustness to orientation: Real-world scenarios involve objects with varying orientations. Training the model on images spanning different orientations enhances its resilience to changes in object rotation.
- Feature extraction: Image rotation encourages the model to learn invariant features. It
 requires the model to emphasize features consistent across orientations, thus aiding in
 the extraction of pertinent and informative features for accurate depth estimation.
- Generalization: Exposure to an extensive array of angles equips the model with the ability to generalize its insights to novel orientations during inference.

In essence, the choice of 160 different angles probably stems from a balance between creating a suitably diverse dataset and managing the computational demands of training. This numerical selection may have emerged through iterative experimentation and validation, ensuring that the model benefits from enhanced diversity while maintaining a manageable training process.

Using the augmented data set, the same models (ResNet50, VGG16, VGG19, and DenseNet169) underwent rigorous training and evaluation. This assessment used a multi-faceted set of evaluation metrics, encompassing accuracy, precision, recall, and F1 score, as shown in Table 5. These metrics serve as vital indicators of the efficacy of the model in distinct facets of the depth estimation task. These models were subjected to rigorous training spanning 100 epochs, with a batch size of 20 and a learning rate set at 0.001. This comprehensive evaluation regimen ensured meticulous scrutiny of the models' competence from various vantage points.

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	0.9212	0.9221	0.9212	0.9216
VGG16	0.9324	0.9338	0.9324	0.9331
VGG19	0.9294	0.9300	0.9294	0.9297
DenseNet169	0.9523	0.9535	0.9523	0.9529

Table 5. Evaluation metrics of different models after 100 epochs on the augmented dataset.

From Table 5, it can be observed that:

- All models attained substantial scores across evaluation metrics, indicating proficient performance in depth estimation.
- DenseNet169 secured the highest values in all metrics, followed by VGG16, VGG19, and ResNet50.
- The models demonstrated consistent alignment between accuracy, precision, recall, and F1 score, reflecting balanced performance in positive and negative classes.
- In particular, the application of angle rotation as an enhancement technique yielded notable improvements in the evaluation metrics compared to the previous experiment with the original dataset.

The progression of the training and testing process over 100 epochs is visually captured in the collection of four graphs shown in Figure 15. This visualization offers valuable insights: ResNet50 illustrates a gradual and consistent increase in accuracy across epochs, albeit with a modest final value. On the contrary, VGG16 and VGG19 exhibit swift accuracy improvements in the initial epochs, followed by a more gradual enhancement rate. In particular, DenseNet169 demonstrates a consistent and rapid accuracy advancement throughout the epochs, culminating in a substantial final accuracy value. It is important to note that all models exhibit diminished accuracy fluctuations after each epoch compared to the earlier experiment, indicating an improved level of learning stability.



Figure 15. Accuracy curves of (A) ResNet50, (B) VGG16, (C) VGG19, and (D) DenseNet169 models.

In terms of training accuracy, a rapid increase was observed from epochs 1 to 10, rising from 0.4412 to 0.9055. Subsequently, the training accuracy continued to improve, but the rate of increase decreased with each epoch. Over the next ten epochs, the training accuracy increased by only 0.06, reaching 0.9645 by the 20th training session. The use of the expanded training dataset of 70,400 images contributed to the improved accuracy of the DenseNet169 model. The model achieved an accuracy of more than 65%, indicating the importance of this research and the generated dataset to estimate the depth of absorption structures in near-infrared images. The slow increase in accuracy from the 10th training session onward can be attributed to the challenge of extracting specific features for each class in the classification model, which comprises 40 classes that represent different depths. Moreover, the increasing blurring of images of absorbing structures at depths above 16.0 mm poses difficulties in distinguishing the blurred images. Furthermore, Figure 16 illustrates the results of the correlation analysis between the depth estimated by the CNN model and the depth given during testing. As the depth increased, the estimation error also increased. The experiment involved 8000 images at 20 depths ranging from 1.0 mm to 20.0 mm, with 40 images per depth for testing. The correlation coefficient was $R^2 = 0.9911$, demonstrating the feasibility of the CNN-DenseNet-169 model in estimating the depths from images of absorbing structures.



Figure 16. Correlation Analysis of Given and Estimated Depths.

Figure 17 illustrates the workflow of the proposed method. First, the original image was convolved with a point spread function (PSF) to simulate the blurring effect caused by light scattering and absorption in biological tissue. This process yielded a blurred image of the absorption structures. Second, the blurred image underwent de-blurring through a fully convolutional network (FCN) model, which could have been either the Attention UNet or the Attention Res-UNet, in order to recover the original image. Lastly, the blurred image was subjected to decoding using a convolutional neural network (CNN) model to estimate the depth of the absorption structures. In further studies, these results will be optimized to reconstruct the 3D structure of biological tissue from a 2D image.



Figure 17. Correlation analysis of given and estimated depths.

5. Conclusions

The de-blurring and estimation depths of the absorbing structure in transillumination images taken through a turbid medium such as biological tissue have attracted significant interest among researchers and experts in biomedical optics in recent years. This study addresses the challenge of de-blurring and depth estimation in transillumination images by utilizing the dependent point spread function (PSF) derived for the light source within a scattering medium. The neural network (NN) technique is employed to find the deep learning models capable of de-blurring the image and estimating the depth of the absorption structure inside a turbid medium. The effectiveness of deep learning for de-blurring transillumination images and also depth estimation has been successfully demonstrated for depths ranging from 0.1 to 10.0 mm in previous studies. Although previous attempts have been made to enhance blurred images, the technique proposed in this study offers another solution.

The attention gate and the residual block were proposed to de-blur the image. Attention Unet and Residual Unet models were examined compared to the Unet model. Attention UNet and Residual Unet models yielded better performance than the Unet model. Attention Res-Unet then examined the performance compared to Attention Unet. Both the Attention U-Net and Attention Res-UNet models achieved correlation coefficients exceeding 88% even at a depth of 20.0 mm, affirming the applicability of deep learning models to de-blur transillumination images. Finally, Attention Res-Unet shows better performance in terms of the correlation between a de-blurred image and the original given image. The impact of image size on the result was also investigated. To ensure a better result with a deep-absorbing structure, the image size in a dimension should be more significant than three times the standard deviation of the light distribution on the medium's surface when calculating the turbid medium's deepest light point source distribution.

This study examined four different models, ResNet-50, VGG-16, VGG-19, and DenseNet-169, to estimate the depth of the absorbing structure. DenseNet-169 demonstrates superior performance among these models, achieving an accuracy rate greater than 65%. This research and the generated dataset prove valuable in accurately estimating the depth of the absorption structure from transillumination images. The evaluation of 1600 test images at 40 different depths ranging from 0.5 mm to 20.0 mm yielded a correlation coefficient of $R^2 = 0.9911$, which affirms the feasibility of the DenseNet169 model in estimating the depth of the absorbing structure.

It should be noted that this proposed technique requires a substantial amount of training data and computational power. However, these challenges can be addressed through the appropriate selection of PSFs and advances in computing capabilities. Consequently, this study confirms the feasibility of deep learning in clarifying blurred images and estimating the depth of absorption structures using PSF and CNN models based on training data.

The de-blurring and depth estimation results obtained for absorption structures at depths from 0.1 to 20.0 mm are highly satisfactory. These findings indicate the usefulness of the proposed methods for observing subcutaneous structures, identifying tumors and small animal parts, and determining depth distributions up to 20.0 mm. In particular, this technique is based solely on computer vision without complex exposure, ultrasound, or additional substances. Therefore, it presents a novel tool for the diagnosis of dermatological diseases, various tumor-associated diseases, vascular diseases, and tissue metabolism.

The results of this study contribute to the development of depth estimation and deblurring methods using deep learning models. Furthermore, merging two targets identified by a single deep learning model will enable the definition of multiple depths within a single image. To expand the model's de-blurring and depth estimation capabilities, it is crucial to increase the number of samples and pairs of images for the training data and expand the depth range. These insights will facilitate the determination of actual dimensions and image depths within the absorption structure for the development of applications using 2D and 3D absorbing structure images in the near future.

Author Contributions: Conceptualization, T.N.T. and H.N.H.; methodology, T.N.T.; software, H.N.H.; validation, N.A.D.N.; analysis, N.A.D.N.; investigation, T.N.T. and H.N.H.; resources, T.N.T.; data curation, N.A.D.N. and H.N.H.; writing—original draft preparation, H.N.H.; writing—review and editing, N.A.D.N. and T.N.T.; visualization, H.N.H.; supervision, T.N.T.; project administration, T.N.T.; funding acquisition, T.N.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying the results presented in this paper are not publicly available but may be obtained from the authors upon reasonable request.

Acknowledgments: We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pan, C.T.; Francisco, M.D.; Yen, C.K.; Wang, S.Y.; Shiue, Y.L. Vein Pattern Locating Technology for Cannulation: A Review of the Low-Cost Vein Finder Prototypes Utilizing near Infrared (NIR) Light to Improve Peripheral Subcutaneous Vein Selection for Phlebotomy. *Sensors* 2019, 19, 3573. [CrossRef] [PubMed]
- Francisco, M.D.; Chen, W.F.; Pan, C.T.; Lin, M.C.; Wen, Z.H.; Liao, C.F.; Shiue, Y.L. Competitive Real-Time Near Infrared (NIR) Vein Finder Imaging Device to Improve Peripheral Subcutaneous Vein Selection in Venipuncture for Clinical Laboratory Testing. *Micromachines* 2021, 12, 373. [CrossRef] [PubMed]
- Frank, N.G.; David, W.; Samuel, D.; Martin, M.; Akwasi, A. Breast-i Is an Effective and Reliable Adjunct Screening Tool for Detecting Early Tumour Related Angiogenesis of Breast Cancers in Low Resource Sub-Saharan Countries. *Int. J. Breast Cancer* 2018, 2018, 2539056.
- Shiryazdi, S.M.; Kargar, S.; Nasaj, H.T.; Neamatzadeh, H.; Ghasemi, N. The accuracy of Breastlight in detection of breast lesions. Indian J. Cancer 2015, 52, 513–516. [PubMed]
- Tobisawa, N.; Namita, T.; Kato, Y.; Shimizu, K. Injection Assist System with Surface and Transillumination Images. In Proceedings of the 2011 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 13–15 May 2011; pp. 1–4.
- 6. Shimizu, K.; Tochio, K.; Kato, Y. Improvement of transcutaneous fluorescent images with a depth-dependent point-spread function. *Appl. Opt.* **2005**, *44*, 2154–2161. [CrossRef] [PubMed]
- Tran, T.N.; Yamamoto, K.; Namita, T.; Kato, Y.; Shimizu, K. Three-dimensional transillumination image reconstruction for small animal with new scattering suppression technique. *Biomed. Opt. Express* 2014, *5*, 1321–1335. [PubMed]
- Goh, C.M.; Subramaniam, R.; Saad, N.M.; Ali, S.A.; Meriaudeau, F. Subcutaneous veins depth measurement using diffuse reflectance image. *Opt. Express* 2017, 25, 25741–25759. [PubMed]
- 9. Nguyen N.A.D.; Van, T.N.P.; Yamamoto, K.; Nguyen M.Q.; Tran, A.T.; Namita, T.; Shimizu K.; Tran T.N. Depth estimation of the absorbing structure in a slab turbid medium using point spread function. *VNUHCM J. Eng. Technol.* 2020, *3*, SI10–SI21.

- Van, T.N.P.; Tran, T.N.; Inujima, H.; Shimizu, K. Three-dimensional imaging through turbid media using deep learning: NIR transillumination imaging of animal bodies. *Biomed. Opt. Express* 2021, 12, 2873–2887. [CrossRef] [PubMed]
- 11. Shourav, M.K.; Choi, J.; Kim, J.K. Visualization of superficial vein dynamics in dorsal hand by near-infrared imaging in response to elevated local temperature. *J. Biomed. Opt.* **2021**, *26*, 026001. [CrossRef] [PubMed]
- Shimizu, K.; Xian, S.; Guo, J. Reconstructing a Deblurred 3D Structure in a Turbid Medium from a Single Blurred 2D Image—For Near-Infrared Transillumination Imaging of a Human Body. Sensors 2022, 22, 5747. [CrossRef] [PubMed]
- 13. Mak, H.W.L.; Han, R.; Yin, H.H.F. Application of Variational AutoEncoder (VAE) Model and Image Processing Approaches in Game Design. *Sensors* 2023, 23, 3457. [CrossRef] [PubMed]
- 14. Qiao, Q. Image Processing Technology Based on Machine Learning. In *IEEE Consumer Electronics Magazine*; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]
- Patil, A. Image Recognition using Machine Learning. 2021. Available online: https://ssrn.com/abstract=3835625 (accessed on 30 August 2023).
- 16. Oktay, O.; Jo Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Kensaku Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
- Maji, D.; Sigedar, P.; Singh, M. Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors. *Biomed. Signal Process. Control* 2022, 71, 103077. [CrossRef]
- He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA 27–30 June 2016; pp. 770–778.
- 19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- 20. Huang, G.; Liu, Z.; Van DerMaaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [CrossRef]
- Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
- 23. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.