


Article

One-Step Enhancer: Deblurring and Denoising of OCT Images

Shunlei Li ^{1,2}, Muhammad Adeel Azam ^{1,2}, Ajay Gunalan ^{1,2} and Leonardo S. Mattos ^{1,*} ¹ Department of Advanced Robotics, Istituto Italiano di Tecnologia, 16163 Genoa, Italy² Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, 16145 Genoa, Italy

* Correspondence: leonardo.demattos@iit.it

Abstract: Optical coherence tomography (OCT) is a rapidly evolving imaging technology that combines a broadband and low-coherence light source with interferometry and signal processing to produce high-resolution images of living tissues. However, the speckle noise introduced by the low-coherence interferometry and the blur from device motions significantly degrade the quality of OCT images. Convolutional neural networks (CNNs) are a potential solution to deal with these issues and enhance OCT image quality. However, training such networks based on traditional supervised learning methods is impractical due to the lack of clean ground truth images. Consequently, this research proposes an unsupervised learning method for OCT image enhancement, termed one-step enhancer (OSE). Specifically, OSE performs denoising and deblurring based on a single step process. A generative adversarial network (GAN) is used for this. Encoders disentangle the raw images into a content domain, blur domain and noise domain to extract features. Then, the generator can generate clean images from the extracted features. To regularize the distribution range of retrieved blur characteristics, KL divergence loss is employed. Meanwhile, noise patches are enforced to promote more accurate disentanglement. These strategies considerably increase the effectiveness of GAN training for OCT image enhancement when used jointly. Both quantitative and qualitative visual findings demonstrate that the proposed method is effective for OCT image denoising and deblurring. These results are significant not only to provide an enhanced visual experience for clinicians but also to supply good quality data for OCT-guide operations. The enhanced images are needed, e.g., for the development of robust, reliable and accurate autonomous OCT-guided surgical robotic systems.

Keywords: optical coherence tomography; image enhancement; generative adversarial network; unsupervised learning

**Citation:** Li, S.; Azam, M.A.;Gunalan, A.; Mattos, L.S. One-Step Enhancer: Deblurring and Denoising of OCT Images. *Appl. Sci.* **2022**, *12*, 10092. <https://doi.org/10.3390/app121910092>

Academic Editor: Lucian Mihai Itu

Received: 23 August 2022

Accepted: 5 October 2022

Published: 7 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

1. Introduction

Optical coherence tomography (OCT) is an imaging technology able to produce high-resolution images of living tissues. Most OCT devices used in clinical studies have a resolution of approximately 10 μm and a depth of penetration up to 2 mm in soft tissues [1]. However, OCT image quality is significantly degraded by speckle noise introduced by the low-coherence interferometry used in the imaging process and by blur arising from relative motions between the device and the tissue [2]. This has a strong impact on subsequent analysis and makes clinical application challenging. Therefore, efficient OCT image enhancement methods are urgently needed [3].

By improving the light source, hardware-based approaches reduce the noise of the detector and scanner to some extent, but the speckle in the imaging system cannot be eliminated. Software-based approaches such as non-local means or block-matching and 3D filtering (BM3D) can provide good results, but need laborious efforts of parameter tuning for different noise levels [4]. Block matching and 4D collaborative filtering (BM4D) expands BM3D to three-dimensional picture volumes [5]. Sliding window, adaptive statistical-based filters, and patch correlation-based filters are the three main classes of digital filters used to

denoise images [6]. However, these methods have limitations that reduce their potential for clinical applications, such as a long processing time and excessive smoothness [7].

Recently, convolutional neural networks (CNNs) have started to be considered as a potential solution for such image enhancement tasks. For example, Kai Zhang et al. proposed a feed-forward denoising convolutional neural network (DnCNN) able to handle Gaussian denoising with unknown noise levels based on a residual learning strategy [8]. In addition, Rico-Jimenez et al. proposed a self-fusion network that was pre-trained to fuse 3 frames to achieve near-real-time processing frame rates [9]. However, supervised learning methods such as these are laborious in terms of training data acquisition, requiring well-paired training images (images with noise and blur and corresponding clean images). Furthermore, the use of standard CNNs may lead to loss of details due to averaging processes [10]. These characteristics make standard CNNs impractical for OCT image enhancement. To overcome these limitations, Chunhao Tian et al. proposed a generative adversarial network (GAN) for the problem of restoring low-resolution OCT fundus images to their high-resolution counterparts [11]. In addition, several other methods based on GAN have been proposed for unpaired image enhancement, such as CycleGAN [12], SNR-GAN [10], and SiameseGAN [13].

Another interesting unsupervised learning strategy for OCT image enhancement is disentangled representation. This strategy divides each feature into narrowly specified variables and encodes them as distinct dimensions. Recently, it has been used for image-to-image translation, such as in BicycleGAN [14] and cross-cycleGAN [15]. In addition, DRGAN implemented this unsupervised learning method for reducing speckle with disentangled representation [16]. However, even though these GAN-based models provided promising results for OCT image despeckling, the problem of blurriness of OCT images still needs to be solved.

This paper presents a novel solution for simultaneous denoising and deblurring of OCT images without requiring a well-paired training dataset. This is achieved with a deep learning GAN architecture that exploits disentangled representation, as shown in Figure 1. After training, the encoder for content and the generator for a clean image can enhance the original image quality. More specifically, the proposed method learns to disentangle noise, blur and content from raw OCT images and then uses this knowledge to generate enhanced images. In order to accommodate for little content information, Kullback—Leibler (KL) divergence [17] loss is used to regularize the distribution range of extracted blur attributes. As shown in Figure 2, the content encoders learn to extract content features from unpaired clean and raw images, while the blur and noise encoders capture blur and noise information from low-quality raw images.

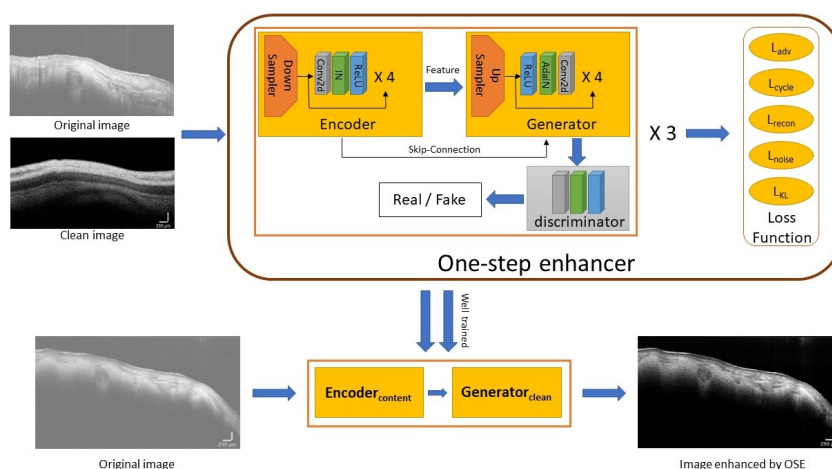


Figure 1. Workflow of proposed OSE image enhancement method: unpaired raw and clean OCT images are used as the input of an unsupervised learning strategy based on disentangled representation and GAN. This process allows the one-step enhancer (OSE) to learn to extract content from low-quality OCT images and generate clean images.

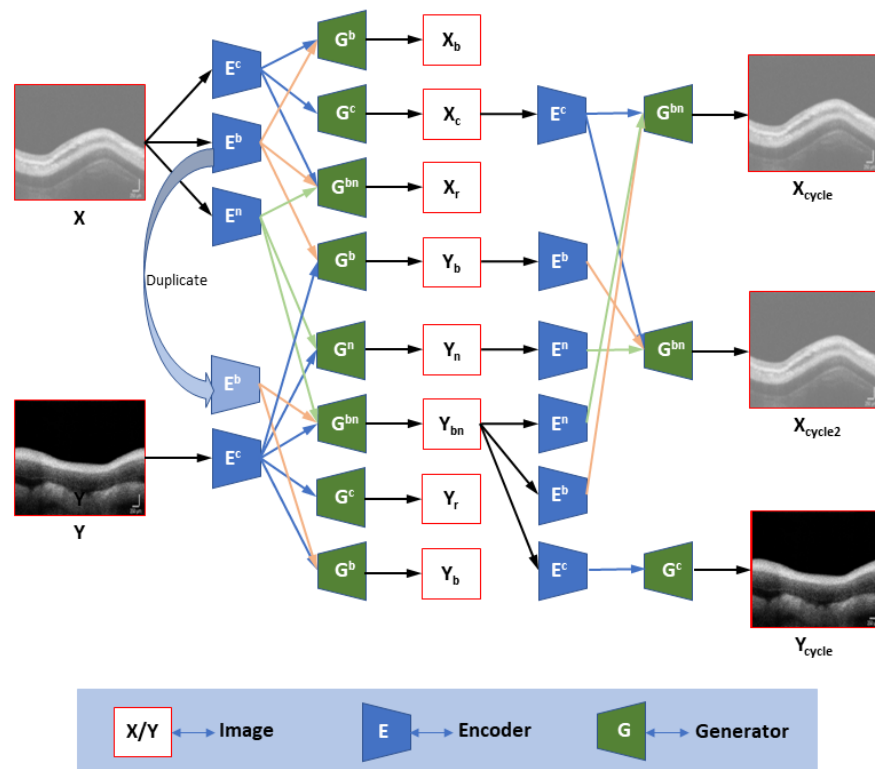


Figure 2. Framework of proposed image enhancement method. X and Y are inputs, where subscripts b, c, r, n, bn and cycle are blurred, clean, reconstructed, noisy, blurred and noisy, and cycled, respectively. The encoder and generator’s superscripts c, b, n, and bn is content, blur, noise, and blur-noise, respectively.

The next sections of this paper are organized as follows: in Section 2, we explain related work, including GAN-based deblurring, GAN-based denoising, and disentangled representation. In Section 3, we describe our proposed method, including the problem formulation, definition of loss functions, method implementation and assessment method. In Section 4, we present experiments and results. Finally, conclusions are presented in Section 5.

2. Related Work

2.1. GAN-Based Speckle Removal

OCT images are known to suffer from speckle noise, which are artifacts produced mostly by the coherent nature of the image formation process. Recently, various GAN-based models have been developed to remove such noise from OCT images based on knowledge extracted from unpaired training data. These include SNR-GAN [10], ARM-SRGAN [18], nonlocal-GAN [19], and DRGAN [16].

SNR-GAN was proposed by Yan Guo et al. to establish an end-to-end structure-aware noise reduction GAN that uses cycle GAN to translate data between noisy and clean domains [20]. In order to preserve subtle features during denoising, they used regional structural similarity index (SSIM) loss of image patches instead of the entire image. This method enabled promising improvements in terms of signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR) and SSIM index with a processing speed of 0.3 s per image.

The ARM-SRGAN is a GAN-based method developed for fast and reliable generation of super-resolution (SR) images without relying on a paired training dataset of low- and super-resolution images [18].

The nonlocal-GAN method, unlike cycle-GAN based methods that include two generators and two discriminators, is based on only one generator and one discriminator [19].

The discriminator can learn the features of noise in noisy OCT images and then direct the denoising generator without reference images.

Finally, DRGAN was proposed by Yongqiang Huang et al. as an unsupervised denoising method that disentangles the noisy image into content and noise spaces by using corresponding encoders. It then predicts the denoised OCT image based on the extracted content features [16]. According to the published results, DRGAN outperforms the methods mentioned above in noise reduction and detail preservation.

2.2. GAN-Based Deblurring

Boyu Lu et al. proposed a method for unsupervised deblurring via disentangled representations with a single image [21]. To properly encode blur information into the deblurring framework, the model disentangles the content and blur characteristics from blurred images.

3. Proposed Method

3.1. Problem Formulation

Overall, the learning process for image enhancement based on unpaired data is realized using disentanglement to decode the image features and GAN to generate clean images. For implementing this, the proposed method consists of three parts: (1) encoders for content (E^c) and features (E^b, E^n and E^{bn} for blur, noise and blur-noise); (2) generators of blurred, noisy, blurred-noisy, and clean images (G^b, G^n, G^{bn}, G^c); and (3) discriminators for blurred, noisy, blurred-noisy, and clean image discrimination (D^b, D^n, D^{bn}, D^c).

An overview of the proposed architecture is shown in Figure 2. Given an input blur-noise data X and unpaired clean data Y , the content encoder E^c extracts content information from corresponding samples, and E^b, E^n estimate the feature information in X . Then G^b, G^n , and G^{bn} take features and content information to generate corresponding images, and G^c generates a clean image. Finally, the discriminators distinguish between the real and generated images.

Since clean images should only contain content components, a well-trained content encoder E^c should allow the generation of the desired enhanced images. This is achieved by exploiting information from the blur and noise domains. For the blur domain, considering the content information of E^c and the blur features of E^b , the generated blur images guide the encoder E^c towards extracting content information from blurred images. Similarly, generating and then distinguishing noisy images from clean ones guides E^c towards extracting content from noisy images. In addition, we enforce the last layers of the encoders for content, blur and noise to share weights, which contributes to guiding E^c towards learning how to effectively extract content information from raw images.

Specifically, E^c encodes inputs X and Y as content features F_x^c and F_y^c , respectively. The blur feature F_x^b and noise feature F_x^n are encoded from X by E^b and E^n . Then, as shown in Equations (1) and (2), the reconstructed Y_r is generated from F_y^c using G^c , and the reconstructed X_r is generated from F_x^c, F_x^b, F_x^n using G^{bn} .

$$Y_r = G^c(F_y^c) \quad (1)$$

$$X_r = G^{bn}(F_x^c, F_x^b, F_x^n) \quad (2)$$

Generators are used to generate new images based on the features described above according to Equations (3)–(7).

$$X_b = G^b(F_x^c, F_x^b) \quad (3)$$

$$X_c = G^c(F_x^c) \quad (4)$$

$$Y_b = G^b(F_y^c, F_x^b) \quad (5)$$

$$Y_n = G^n(F_y^c, F_x^n) \quad (6)$$

$$Y_{bn} = G^b(F_y^c, F_x^b, F_x^n) \tag{7}$$

Disentanglement is then used to handle unpaired inputs and generate new images from a corresponding domain. Features are obtained from generated images: $F^c - X_c, F^b - Y_b, F^n - Y_n, F^c - Y_{bn}, F^b - Y_{bn}, F^n - Y_{bn}$. Finally, cycled inputs are obtained as follows:

$$X_{cycle} = G^{bn}(F^c - X_c, F^b - Y_{bn}, F^n - Y_{bn}) \tag{8}$$

$$X_{cycle2} = G^{bn}(F^c - X_c, F^b - Y_b, F^n - Y_n) \tag{9}$$

$$Y_{cycle} = G^{bn}(F^c - Y_{bn}) \tag{10}$$

After training the model and addressing disentanglement, E^c can extract content features from low-quality images, and then clean images can be obtained using G^c .

3.2. Loss Function

The overall loss function includes five subfunctions: domain adversarial loss (L_{adv}), cycle consistency loss (L_{cycle}), reconstruction loss (L_{recon}), noise patch loss (L_{noise}) and KL divergence loss (L_{KL}). Their interconnections with the processing framework is illustrated in Figure 3.

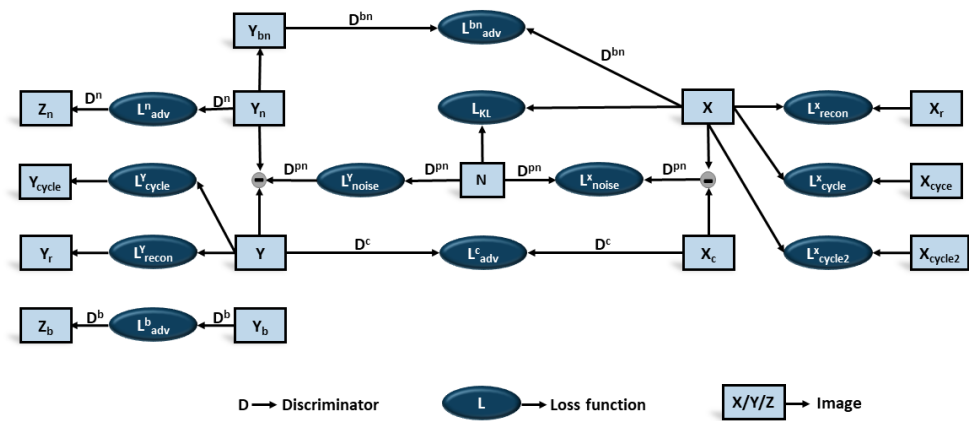


Figure 3. Diagram illustrating the inputs and outputs of loss functions. The inputs include X: original images, Y: clean images, Z: real blurred/noisy images.

(1) Domain adversarial loss: L_{adv} pushes the discriminators to pick the best encoders and generators to minimize the adversarial loss functions, which include content information loss L_{adv}^c , blur feature loss L_{adv}^b , noise feature loss L_{adv}^n , and blur-noise feature loss L_{adv}^{bn} . The domain adversarial loss is defined as:

$$L_{adv} = \underset{E,G}{\operatorname{argmin}} \max_D (L_{adv}^c + L_{adv}^b + L_{adv}^n + L_{adv}^{bn}) \tag{11}$$

where E stands for the encoder, G for the generator, and D for the discriminator. The four adversarial loss functions are defined below, where Z_b and Z_n are real blurred and noisy images, and \mathbb{E} is the expectation operator.

$$L_{adv}^c = \mathbb{E}[\log D^c(Y)] + \mathbb{E}[1 - \log D^c(X_c)] \tag{12}$$

$$L_{adv}^b = \mathbb{E}[\log D^b(Z_b)] + \mathbb{E}[1 - \log D^b(Y_b)] \tag{13}$$

$$L_{adv}^n = \mathbb{E}[\log D^n(Z_n)] + \mathbb{E}[1 - \log D^n(Y_n)] \tag{14}$$

$$L_{adv}^{bn} = \mathbb{E}[\log D^{bn}(X)] + \mathbb{E}[1 - \log D^{bn}(Y^{bn})] \tag{15}$$

(2) Cycle consistency loss: inspired by CycleGAN [20], cycle consistency loss is introduced to guarantee that the enhanced image X_c represents a proper reconstruction

of the raw sample image, and that Y^{bn} can be translated back to the original clean image domain. The cycle consistency loss further limits the space of the generated samples and preserves the content of original images.

$$L_{cycle} = \mathbb{E}[\|X - X_{cycle}\|_1] + \mathbb{E}[\|X - X_{cycle2}\|_1] + \mathbb{E}[\|Y - Y_{cycle}\|_1] \tag{16}$$

where $\|\cdot\|_1$ represents the l_1 -norm.

(3) Reconstruction loss: the reconstruction loss is applied to facilitate $X = X_r$ and $Y = Y_r$. Consequently, G^c and G^{bn} can reconstruct the inputs to generate a clean counterpart of X and a blur-noise counterpart of Y .

$$L_{recon} = \mathbb{E}[\|X - X_r\|_1] + \mathbb{E}[\|Y - Y_r\|_1] \tag{17}$$

(4) Noise patch loss: to overcome the obstacle of multiple types of noise, we leverage noise patches from the background of raw images and use a discriminator D^{pn} to distinguish between real noise and generated noise as follows:

$$L_{noise}^X = \mathbb{E}[\log D^{pn}(N)] + \mathbb{E}[\log D^{pn}(X - X_b)] \tag{18}$$

$$L_{noise}^Y = \mathbb{E}[\log D^{pn}(N)] + \mathbb{E}[\log D^{pn}(Y - Y_b)] \tag{19}$$

According to Equations (18) and (19), the noise patch loss is given by:

$$L_{noise} = \underset{E,G}{\operatorname{argmin}} \max_D (L_{noise}^X + L_{noise}^Y) \tag{20}$$

(5) KL divergence loss: to guarantee that the blur encoder $E^b(X)$ only encodes blur components, Y_b is generated from $E^c(Y)$ and $E^b(X)$ in Equation (5). This discourages $E^b(X)$ from encoding content information from X . Furthermore, KL divergence loss is used to regularize the blur feature distribution $F_b = E^b(X)$ to bring it closer to a normal distribution $p(F) \sim N(0, 1)$. KL divergence is minimized to obtain the KL loss as described in [22]:

$$L_{KL} = \frac{1}{2} \sum_{i=1}^N (\mu_i^2 + \sigma_i^2 - \log(\sigma_i^2) - 1) \tag{21}$$

where μ and σ are the mean and standard deviation of F_b , and N is its dimension. The KL divergence loss can reduce the gap between the prior $p(F)$ and the learned distributions. This further suppresses any content information contained in F_b .

Considering the equations above, the overall loss function can be written as:

$$L = \lambda_{adv} L_{adv} + \lambda_{cycle} L_{cycle} + \lambda_{recon} L_{recon} + \lambda_{noise} L_{noise} + \lambda_{KL} L_{KL} \tag{22}$$

where the subscripted λ are the coefficients of each corresponding loss function.

3.3. Implementation and Data

The proposed network architecture has a structure similar to DRGAN [16]. The content encoder is composed of an input convolutional layer, a down sampler and four residual blocks. The noise encoder consists of an input convolutional layer, a down sampler and an adaptive average pooling layer with a 1×1 convolutional layer. The blur encoder has four strided convolution layers and one fully connected layer. For the generator, the architectures are symmetric to the content encoder, but vary for generating images of different domains. We use skip-connections between E^c and G^c , with SPADE [23] and adaptive instance normalization [24], to fuse features from different levels. Then, the discriminator applies a series of convolutional and pooling layers to give a binary judgement.

This model was implemented in PyTorch with a Ubuntu 20.04 operation system and an NVIDIA Quadro RTX 8000 GPU. During training, the Adam optimizer was used, and the learning rate was set to 0.0002 for 80 epochs. According to information in [16,21], the hyper-parameters in our framework were experimentally set to the following values: $\lambda_{adv} = 1$, $\lambda_{cycle} = 10$, $\lambda_{recon} = 10$, $\lambda_{noise} = 1$, and $\lambda_{KL} = 0.01$.

We acquired two datasets for this study. One consisted of 30 low-quality OCT images and 30 clean OCT images from three different pork larynxes. The second dataset contained the same number of images captured from two ex vivo rabbit eyes. These custom datasets were captured using a commercial OCT device (TEL320C1, Thorlabs, Inc., Newton, NJ, USA). The pixel size was $0.40 \times 2.47 \mu\text{m}$ (width \times height), and the size of each image was set to $10,000 \times 1024$ pixels. Therefore, the FOV was 4.00×2.53 mm.

The clean images in our datasets were obtained using the Speckle Averaging function of the ThorImageOCT software (version 5.2), which uses 4 successive A-scans to compute the mean and variance values used in the averaging process.

A test set was formed by randomly selecting 20 images with noise and blur and 20 corresponding clean images from each dataset. The remaining 40 images were randomly combined into pairs to form a training set. Since tissue information is concentrated in the middle part of the OCT images, all images were center-cropped to a pixel size of 900×450 to improve training efficiency.

To obtain noise features for the noise loss function, a window with pixel size of 256×256 was used to extract noisy patches from the background of low-quality images. This window was applied to the images in the training set using a stride of 8 pixels. This process extracted a total of 19,360 patches from low-quality images for the input X , and 19,360 patches from clean images for the input Y .

3.4. Experimental Method and Performance Metrics

An ablation study was performed to evaluate the performance of each component in the proposed image enhancement method. This consisted in evaluating the performances of each module separately: first, the denoise and the deblur modules were independently assessed. Then, the performance of the proposed model, which combines both operations into a single step operation, was evaluated. More specifically, we removed speckle and blur from the original images separately, and then used the proposed model to perform a one-step image enhancement.

In addition, to benchmark the performance of the proposed image enhancer, non-learning (BM3D [25]), supervised learning (DnCNN [8]) and unsupervised learning (DRGAN [16]) methods were implemented. The BM3D software implements the traditional OCT image enhancement method, while the DnCNN and DRGAN models were trained on the same dataset used to train our new model.

Performance evaluation used the same test set described above. In addition, the processing time of each method was evaluated both on CPU and GPU. Finally, a visual assessment of the four image enhancement methods was performed using sample images from the test set.

Two metrics were used for quantitative performance assessment: peak signal-to-noise ratio (PSNR) and structure similarity index measure (SSIM). PSNR is commonly used to measure the quality of reconstruction of lossy image compression codecs. It offers an approximation to human perception of reconstruction quality based on differences between the reconstructed and the reference image. SSIM, on the other hand, measures the similarity between two images. The overall index of SSIM evaluates luminance, contrast and structural differences.

$$PSNR = 10 \log_{10} \left(\frac{(\max(I_g))^2}{\frac{1}{MN} \sum_i \sum_j (I_c(i,j) - I_g(i,j))^2} \right) \quad (23)$$

where I_g and I_c are, respectively, the generated and the averaging clean images. M and N are the size of the image.

$$SSIM = \frac{(2\mu_{I_g}\mu_{I_c} + C_1)(2\sigma_{I_g I_c} + C_2)}{(\mu_{I_g}^2 + \mu_{I_c}^2 + C_1)(\sigma_{I_g}^2 + \sigma_{I_c}^2 + C_2)} \quad (24)$$

where μ_{I_g} , μ_{I_c} , σ_{I_g} , σ_{I_c} and $\sigma_{I_g I_c}$ are the local means, standard deviations and cross-covariance of images I_g and I_c . The constants C_1 and C_2 are set according to the literature [26].

4. Experimental Results

4.1. Ablation Study Results

Visual results from the ablation study are presented in Figure 4, while the quantitative results are shown in Table 1. A visual analysis of Figure 4 shows that the denoise module is effective in removing noise from the original OCT image (raw image). The data in Table 1 demonstrates improvements of 10.59 in PSNR and 0.24 in SSIM, which confirms this module works properly. However, the problem of unclear tissue layers is not addressed.

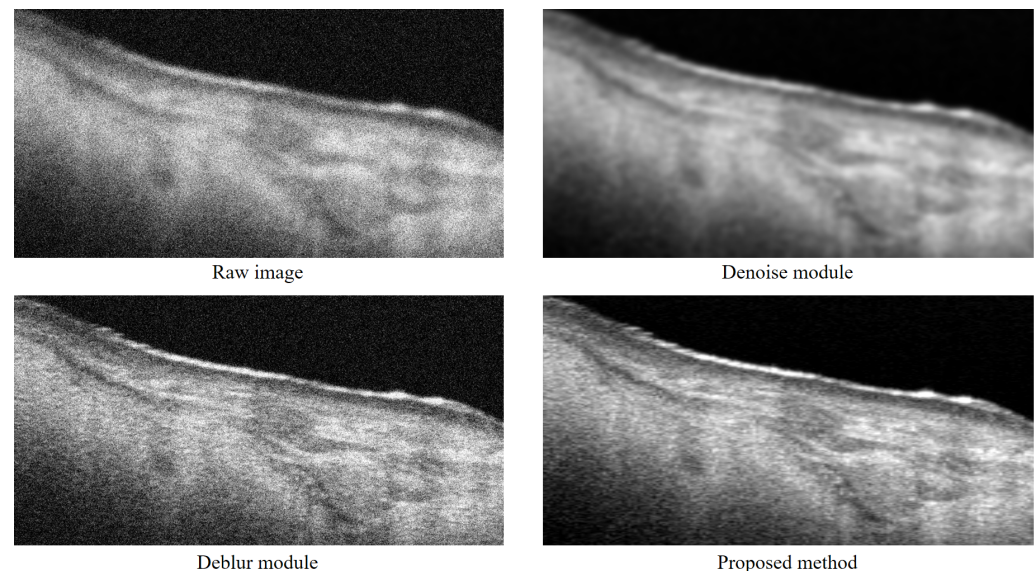


Figure 4. Sample images from the ablation study performed using OCT images from a pork larynx to evaluate the denoise module, the deblur module, and the proposed OSE method.

Table 1. Ablation study results.

Method	Metrics (Mean \pm std)	
	PSNR	SSIM
Original images	8.94 \pm 2.01	0.34 \pm 0.14
Denoise module	19.53 \pm 1.87	0.58 \pm 0.20
Deblur module	17.55 \pm 1.52	0.47 \pm 0.12
OSE	26.71 \pm 2.21	0.81 \pm 0.16

The deblur module, on the other hand, removes blur and makes the layers more visible. This can be observed in Figure 4 by comparing the result of the deblur module with the raw image. In this case, the PSNR improved from 8.94 to 17.55, and the SSIM improved from 0.34 to 0.47. However, noise is still present in the image, and this limits the image enhancement performance.

The proposed method combining both modules provides better enhancement performance than each single module applied separately. A visual inspection of the result in Figure 4 shows that the proposed method was able to effectively enhance the original

raw image. This is corroborated by the data in Table 1, which shows the proposed method achieved top performances in terms of PSNR and SSIM.

4.2. Performance Comparison Results

Figure 5 provides a visual comparison of the image enhancement results achieved with the proposed OSE and with the other state of the art methods: the 3-D block-matching algorithm (BM3D), the supervised learning-based method DnCNN, and the unsupervised learning-based method DRGAN. It can be observed that all methods achieved satisfactory speckle reduction performance, but only OSE was effective in also removing the blurring effects on the image details, as shown in the selected magnified areas.

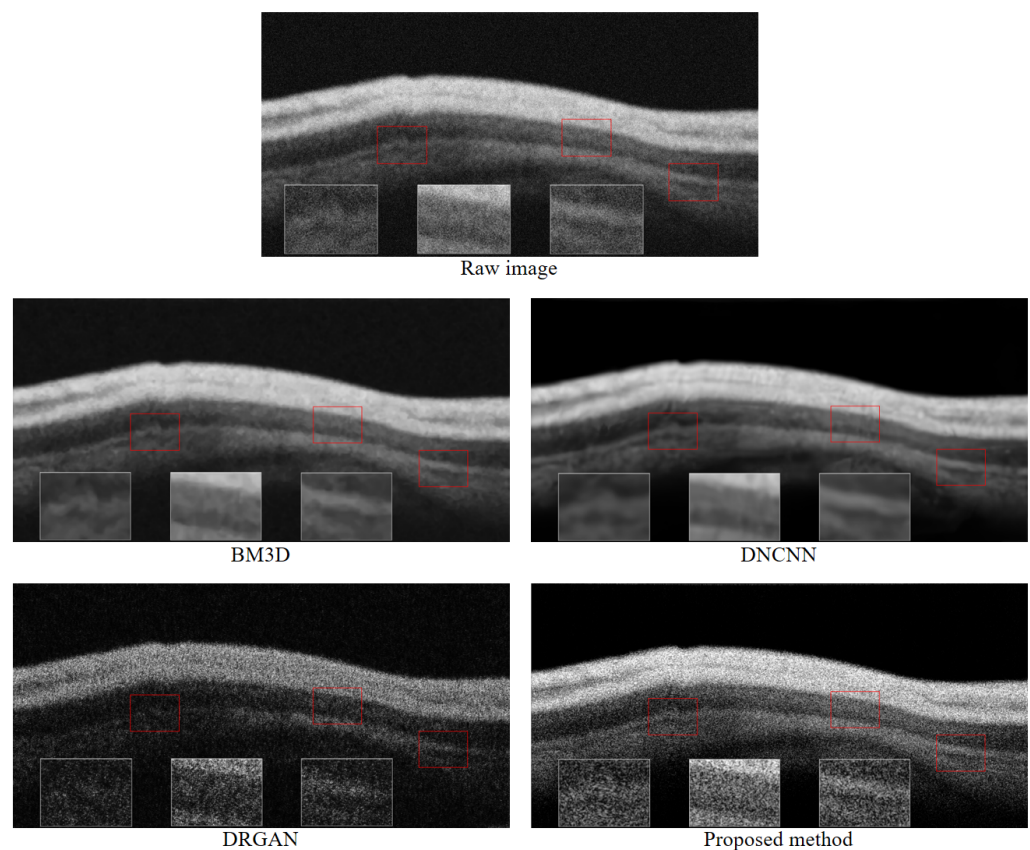


Figure 5. Sample images from the image enhancement performance study comparing the proposed OSE with the state of the art methods BM3D, DnCNN and DRGAN. The image areas marked in red are magnified and shown as inset pictures to facilitate the visual assessment of the different algorithms.

The data in Table 2 summarizes the quantitative performance metrics obtained for the four different enhancement methods. OSE improved PSNR from 8.94 to 26.71 and SSIM from 0.34 to 0.81, outperforming all the other methods in terms of both denoising and deblurring.

Table 3 shows the mean processing time of the methods assessed for $10,000 \times 1024$ pixel images. Considering this data, we can see that although BM3D provides good image enhancement results, it takes much more time than the other methods to process the OCT images. In addition, we can note that DnCNN performed better than DRGAN but was slower. Furthermore, as explained earlier, DnCNN is a supervised learning method that requires a well-paired set of images for training. On the other hand, OSE provided top image enhancement performance with the best processing speed when the computations were performed on a GPU. It achieved a mean processing rate of 8.3 fps on the GPU.

Table 2. Quantitative performance comparison with state-of-the-art methods for OCT image enhancement.

Method	Metrics (Mean \pm std)	
	PSNR	SSIM
Original images	8.94 \pm 2.01	0.34 \pm 0.14
BM3D	24.11 \pm 1.04	0.71 \pm 0.08
DnCNN	23.99 \pm 2.70	0.78 \pm 0.24
DRGAN	16.77 \pm 1.04	0.61 \pm 0.10
OSE	26.71 \pm 2.21	0.81 \pm 0.16

Table 3. Mean processing time for 10,000 \times 1024 pixel OCT images

Method	Mean Processing Time (s)	
	CPU	GPU
BM3D	45.69	-
DnCNN	11.14	0.17
DRGAN	3.77	0.14
OSE	3.86	0.12

5. Conclusions

This paper presented a novel deep learning model for one-step denoising and deblurring of OCT images. This one-step enhancer (OSE) is trained using an unsupervised learning strategy, which allows learning from a mixed dataset of unpaired OCT images. For this, the method uses disentangled representation and generative adversarial network (GAN) to extract content, blur and noise features from raw OCT images, and then learns to generate clean images. The proposed OSE was assessed through both an ablation study and a comparative performance evaluation based on the quantitative metrics PSNR and SSIM. The ablation study demonstrated that the method produced effective denoise and deblur modules, which enabled high performance levels when used in a combined model. The comparative analysis showed the proposed method outperformed state-of-the-art methods for OCT image enhancement, indicating that our one-step enhancer is a valuable alternative for speckle and blur reduction in OCT images.

Author Contributions: Conceptualization, S.L.; Data curation, A.G.; Investigation, S.L. and L.S.M.; Methodology, S.L. and L.S.M.; Software, S.L. and M.A.A.; Supervision, L.S.M.; Writing—original draft, S.L.; Writing—review & editing, M.A.A., A.G. and L.S.M. 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 presented in this study is openly available in IIT's Dataverse repository at <https://doi.org/10.48557/RI3EQG>, reference "OCT dataset of pork larynx and rabbit eyes".

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

References

- Podoleanu, A.G. Optical coherence tomography. *J. Microsc.* **2012**, *247*, 209–219. [[CrossRef](#)] [[PubMed](#)]
- Drexler, W.; Morgner, U.; Ghanta, R.K.; Kärtner, F.X.; Schuman, J.S.; Fujimoto, J.G. Ultrahigh-resolution ophthalmic optical coherence tomography. *Nat. Med.* **2001**, *7*, 502–507. [[CrossRef](#)] [[PubMed](#)]
- Larin, K.V.; Ghosn, M.G.; Bashkatov, A.N.; Genina, E.A.; Trunina, N.A.; Tuchin, V.V. Optical clearing for OCT image enhancement and in-depth monitoring of molecular diffusion. *IEEE J. Sel. Top. Quantum Electron.* **2011**, *18*, 1244–1259. [[CrossRef](#)]
- Chong, B.; Zhu, Y.K. Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter. *Opt. Commun.* **2013**, *291*, 461–469. [[CrossRef](#)]

5. Maggioni, M.; Katkovnik, V.; Egiazarian, K.; Foi, A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* **2012**, *22*, 119–133. [[CrossRef](#)]
6. Adabi, S.; Turani, Z.; Fatemizadeh, E.; Clayton, A.; Nasiriavanaki, M. Optical coherence tomography technology and quality improvement methods for optical coherence tomography images of skin: A short review. *Biomed. Eng. Comput. Biol.* **2017**, *8*, 1179597217713475. [[CrossRef](#)] [[PubMed](#)]
7. Li, M.; Idoughi, R.; Choudhury, B.; Heidrich, W. Statistical model for OCT image denoising. *Biomed. Opt. Express* **2017**, *8*, 3903–3917. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)] [[PubMed](#)]
9. Rico-Jimenez, J.J.; Hu, D.; Tang, E.M.; Oguz, I.; Tao, Y.K. Real-time OCT image denoising using a self-fusion neural network. *Biomed. Opt. Express* **2022**, *13*, 1398–1409. [[CrossRef](#)] [[PubMed](#)]
10. Guo, Y.; Wang, K.; Yang, S.; Wang, Y.; Gao, P.; Xie, G.; Lv, C.; Lv, B. Structure-aware noise reduction generative adversarial network for optical coherence tomography image. In Proceedings of the International Workshop on Ophthalmic Medical Image Analysis, Shenzhen, China, 17 October 2019; pp. 9–17.
11. Tian, C.; Yang, J.; Li, P.; Zhang, S.; Mi, S. Retinal fundus image superresolution generated by optical coherence tomography based on a realistic mixed attention GAN. *Med. Phys.* **2022**, *49*, 3185–3198. [[CrossRef](#)] [[PubMed](#)]
12. Manakov, I.; Rohm, M.; Kern, C.; Schworm, B.; Kortuem, K.; Tresp, V. Noise as domain shift: Denoising medical images by unpaired image translation. In Proceedings of the Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, International Workshop on Medical Image Learning with Less Labels and Imperfect Data, Shenzhen, China, 17 October 2019; pp. 3–10.
13. Kande, N.A.; Dakhane, R.; Dukkipati, A.; Yalavarthy, P.K. SiameseGAN: A generative model for denoising of spectral domain optical coherence tomography images. *IEEE Trans. Med. Imaging* **2020**, *40*, 180–192. [[CrossRef](#)]
14. Zhu, J.Y.; Zhang, R.; Pathak, D.; Darrell, T.; Efros, A.A.; Wang, O.; Shechtman, E. Toward multimodal image-to-image translation. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
15. Lee, H.Y.; Tseng, H.Y.; Huang, J.B.; Singh, M.; Yang, M.H. Diverse image-to-image translation via disentangled representations. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 35–51.
16. Huang, Y.; Xia, W.; Lu, Z.; Liu, Y.; Chen, H.; Zhou, J.; Fang, L.; Zhang, Y. Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images. *IEEE Trans. Med. Imaging* **2020**, *40*, 2600–2614. [[CrossRef](#)] [[PubMed](#)]
17. Hershey, J.R.; Olsen, P.A. Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP’07, Honolulu, HI, USA, 15–20 April 2007; Volume 4, pp. IV-317–IV-320. [[CrossRef](#)]
18. Das, V.; Dandapat, S.; Bora, P.K. Unsupervised super-resolution of OCT images using generative adversarial network for improved age-related macular degeneration diagnosis. *IEEE Sens. J.* **2020**, *20*, 8746–8756. [[CrossRef](#)]
19. Guo, A.; Fang, L.; Qi, M.; Li, S. Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–12. [[CrossRef](#)]
20. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Lu, B.; Chen, J.C.; Chellappa, R. Unsupervised domain-specific deblurring via disentangled representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10225–10234.
22. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
23. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346.
24. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
25. Mäkinen, Y.; Azzari, L.; Foi, A. Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching. *IEEE Trans. Image Process.* **2020**, *29*, 8339–8354. [[CrossRef](#)] [[PubMed](#)]
26. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]