Optik - International Journal for Light and Electron Optics 284 (2023) 170924

Contents lists available at ScienceDirect

Optik - International Journal for Light and Electron Optics

journal homepage: www.elsevier.com/locate/ijleo

Original research article

ResCoWNet: A deep convolutional neural network with residual learning based on DT-CWT for despeckling Optical Coherence Tomography images

Arun P.S., Shreyash Ajay Sahare, Varun P. Gopi *

Department of Electronics and Communication Engineering, National Institute of Technology Tiruchirappalli, Tamilnadu 620015, India

ARTICLE INFO

Keywords:
Optical coherence tomography
DT-CWT
Speckle noise
Convolutional neural networks
Residual learning
Batch normalization
Image denoising

ABSTRACT

Various retinal illnesses can be diagnosed and treated using the non-invasive diagnostic technology called OCT (Optical Coherence Tomography). However, the processes used to create OCT images generate speckle noise which considerably reduces the quality of the images, and the subsequent disease diagnosis is negatively impacted by such poor-quality images. The despeckling is effected using conventional methods such as spatial/transform domain filtering, dictionary learning, or a combination of these methods. Deep Convolutional Neural Networks (CNN) have improved the ability to exploit spatial correlations and extract data at various resolutions by adopting a hierarchical network structure, thereby making image-denoising methods more reliable. In deep networks, the Discrete Wavelet Transform (DWT) and its inverse are typically used in place of pooling operations. Nevertheless, the Dual-Tree Complex Wavelet Transform (DT-CWT) can provide a significant performance enhancement compared to deep network models for noise reduction based on DWT pooling. Hence to despeckle OCT images, this paper proposes a novel deep convolutional neural network (ResCoWNet) with residual learning based on DT-CWT. The input noisy image is transformed using DT-CWT into various subbands and is made to pass through several convolutional as well as residual blocks at three different levels in ResCoWNet. The denoised versions of OCT images produced by our model outperform the experimental results both quantitatively and subjectively.

1. Background

Optical Coherence Tomography (OCT) is a physiological imaging method to detect and identify eye diseases. It employs less coherent near-infrared light pulses that move more rapidly than ultrasound to take a cross-sectional image of the target tissue, much like ultrasound, which uses high-frequency sound to scan biological tissue. Generally, the OCT imaging process is plagued by sample-based speckle and detector noise, which deteriorates the total contrast and signal-to-noise ratio of OCT images, leading to reduced discernibility of retinal layer tissues necessary for detecting eye diseases [1]. Unlike other noise counterparts, speckle in OCT is modeled as a multiplicative noise, [2], and speckle reduction in OCT images is, therefore, an important preprocessing step in disease diagnosis. There are various software-based methods for eliminating speckle noise, and they can be classified into [3]:

- Spatial domain filtering approach
- Sparsity-based approach

* Corresponding author.
E-mail addresses: 408321001@nitt.edu (Arun P.S.), 108119111@nitt.edu (S.A. Sahare), varun@nitt.edu (V.P. Gopi).

https://doi.org/10.1016/j.ijleo.2023.170924
Received 7 December 2022; Received in revised form 28 April 2023; Accepted 28 April 2023
Available online 9 May 2023

- Energy-based approach
- Deep learning-based approach

Spatial domain filtering approach:Spatial filters can be classified as local or non-local filters [4]. Local filters like Gaussian filter, Frost filter [5], Lee filter [5], Least Mean Squares Filter (LMSF), bilateral filter [6], anisotropic filtering [6] etc., give a lot of time complexity benefits. However, they fail to eliminate multiplicative noise. Non-local filtering concentrates on related pixels, no matter how far the pixels are apart. Non-local filters provide obvious benefits in high-noise situations [7–9]. However, they have over-smoothing problems. For removing multiplicative speckle noise, several NLM filtering modifications have been proposed. The Probability-based Non-Local Means filtering (PNLM) [9] is one such modification mentioned above, which reduces the weight of those severely corrupted pixels by computing the probability of uncorruption using the ROAD function [10]. This probability then calculates the similarity distances between pixels in non-local patches. The major drawback of spatial-based approaches is over-smoothing and blurring of retinal layers [11]

Sparsity-based approach: Here, a sparse representation dictionary is learned for each of the high-SNR images, and we utilize such dictionaries to denoise the low-SNR images [12]. Sparse signal representation enables transmitting the essential data of the signal using only a few simple components termed ‘‘atoms’’. The most popular dictionary learning algorithm is K-SVD, [13], which is an iterative method that achieves sparse representation by adapting dictionary atoms to fit the data better. Another sparse concept approach is the low-rank representation (LRR) model, which is well exploited in image reconstruction. The traditional LRR model has been modified as a non-local Weighted Gradient Low-Rank Representation (WGLRR) [14] wherein the corrupted probability of each pixel is integrated into the LRR model as a weight to regularize the error term. Most sparse representation techniques preserve edge details of the image; however, they are less efficient in suppressing speckle noise [11]. Although the online dictionary learning (ODL) technique [15] presented in our previous work [16] was effective in retaining retinal layer structure, it only reduced speckle noise to a limited extent.

Energy-based approach: In this approach, the energy or cost function is formulated from different aspects of the image and then minimized using various optimization techniques [17]. The total generalized variation (TGV) decomposition model is one such formulation. It decomposes an OCT picture into a structure image and a texture image, with the former being the denoised image and the latter containing the original image’s speckle noise [18]. In another formulation, the speckle noise in OCT is divided into multiplicative and additive portions, and it is reduced using an Augmented Lagrangian Minimization (ALM) approach combined with a Split-Bregman method (SBM) [19]. Total Variation Denoising (TVD) is a well-known denoising technique in image processing that preserves edge and structural details. To minimize speckle noise in OCT images, TVD has been paired with a median filter, where speckle noise is assumed to be a mix of additive white Gaussian noise and salt and pepper noise [20].

Deep learning-based approach: With the emergence of large image databases and superior computing technologies, deep learning has enabled prospective applications and yielded substantial research achievements in the domain of ophthalmological image processing. By adopting a hierarchical network structure, deep Convolutional Neural Networks (CNN), one type of deep learning architecture, have leveraged the ability to harness spatial correlations and extract image data at multiple resolutions. Residual learning strategies on deep CNN models, [21] combined with a perceptually sensitive loss function, is a recent speckle-noise reduction method [22]. Image restoration methods based on Generative Adversarial Networks (GANs) have attracted much attention recently. GAN is a generative modeling methodology that emerged in contrast to the drawbacks of supervised learning methods, which need a massive amount of labeled data. Speckle noise reduction along with image structure preservation has been achieved using SiameseGAN [23], DN-GAN [24], and by combining CycleGAN and mini cGAN [25]. These networks have a few limitations, like increased testing complexity and loss of structural content in low-pixel intensity areas.

Numerous software-based speckle reduction techniques have the drawback of distorting the structure of the retinal layers, inadequately suppressing the speckle noise, or increasing testing complexity. As evident from our previous work in [16] about the efficacy of DT-CWT in fusing two different algorithms for OCT image denoising, we hereby explore and propose a novel model utilizing DT-CWT in a multi-level deep CNN network together with residual learning strategies for despeckling OCT images that can maintain optimal levels of complexity, OCT image structure preservation, and speckle noise reduction.

2. Related works

In this section, we present a brief overview of the evolution of residual learning in CNN and the relevance of including wavelets/complex wavelet layers in CNN for image denoising.

2.1. Residual learning

With the residual learning and batch normalization [26], Zhang et al. [21] have proposed a feed-forward denoising CNN (DnCNN) that removes the latent clean image from the noisy input image in the hidden network layers. The DnCNN model denoised the images contaminated by Gaussian noise at unknown levels. Loss functions are important in deep learning model training because they impact the neural network’s efficacy and accuracy. Qiu et al. [22] proposed a modified DnCNN model that used a perceptually-sensitive loss function to measure the disparity between the denoised and labeled OCT images. In [27], the residual Gaussian denoiser proposed by Zhang et al. [21] is trained with speckle noise added to natural images. The trained network was then utilized to despeckle OCT images. Zhang et al. [28] introduced a dense residual network that completely leveraged the hierarchical characteristics of the original poor-quality image and demonstrated excellent restoration abilities while using fewer network parameters than the residual network and outperforming the dense network.

2.2. Wavelets in CNN

A larger receptive field enhances CNN's capacity to fit data and fosters quality and accuracy by allowing for more spatial content. In general, raising the network depth, expanding the filter size, or performing a pooling operation can increase the receptive field. However, the price paid for the larger network depth or filter size is an increase in computational costs. Pooling can broaden the receptive field and guarantee effectiveness by directly decreasing the spatial resolution of the feature map. But it may lead to data loss. Hence as a solution, integrating discrete wavelet transform (DWT) into CNN architecture instead of pooling operations decreases feature map resolution while boosting the receptive field. In [29], DWT was used in a residual learning network for the Gaussian denoising task. The residue obtained as the output is subtracted from the wavelet sub-bands, and the difference is applied to inverse DWT to get the denoised output image. In [30], multilevel wavelet CNN (MWCNN) was proposed with DWT, and its inverse was used as a replacement for upsampling/downsampling operations in different levels of its U-net architecture. The proposed MWCNN was effective in recovering fine textures and distinct structures from degraded observations because of DWT's invertibility and its frequency and location properties. Peng et al. [31] introduced a multi-level wavelet residual network(MWRN) architecture and progressive training procedure, which improved image denoising ability. As opposed to MWCNN, MWRN has several residual blocks added after each level of DWT and before each level of IDWT. In [32], WaveCNet proposed employing DWT to downsample images to reduce the aliasing effect and improve classification accuracy by integrating it into commonly used CNN like VGG, DenseNet, and ResNet. To eliminate the major part of the signal noises, the high-frequency components after DWT are eliminated in this model. Although the literature mentioned above did a good job of eliminating Gaussian noise, it failed to reduce speckle noise. Qin et al. [33] proposed a speckle suppressing network, Wavelet-SRNet, that used multilevel DWT and a trainable threshold block separately dealing with high pass wavelet decomposed images. The network increased classification accuracy for target detection in noisy synthetic aperture radar (SAR) pictures.

2.3. Complex wavelets in CNN

Duan et al. [34] introduced a novel SAR segmentation method based on convolutional wavelet neural network (CWNN) and Markov random field (MRF). The wavelet pooling incorporated in CWNN had single-level DT-CWT, whose low-frequency subband outputs were only passed to subsequent convolutional layers. After CWNN fused with MRF, the superpixel approach improved the labeling consistency and ensured the preservation of the edges and the details. Farhadiani et al. [35] performed SAR despeckling in the complex wavelet domain. In their methodology, approximate coefficients after DT-CWT were despeckled using a multichannel logarithm with a fast and flexible CNN trained on Gaussian noise, and the detailed coefficients were despeckled by log transforming for averaged maximum a posteriori estimation. The despeckled coefficients were then inverse-transformed to obtain the noise-suppressed output.

Therefore, in this paper, we aim to design an effective despeckling network (ResCoWNet) for speckled OCT images that adopt residual learning on DT-CWT transformed images to suppress the speckle noise. The effectiveness of MWRN [31] in exploiting residual blocks in multilevel wavelet networks to extract the latent clean image from noisy observations served as the inspiration for the proposed network.

3. Methodology

3.1. Proposed method

The residual complex wavelet network (ResCoWNet) architecture is a 3-level CNN model. We examine the single and double-level architecture of ResCoWNet, as shown in Figs. 1 and 2, to validate the requirement for three levels, as shown in Fig. 3. Residual blocks and convolutional blocks form the basic elements of our deep network, the schematic of which is included in Fig. 1. The input of our model is an OCT observation contaminated by speckle noise whose mathematical model is given below in Eq. (1) [8,36],

$$g(x, y) = f(x, y) + f(x, y)^{\gamma} \times n(x, y) \tag{1}$$

where (x, y) are the spatial coordinates of the image, f corresponds to the original uncorrupted image, and n is the zero-mean Gaussian noise. When the noise is multiplicative, the factor γ equals unity and depends on the OCT device and processing method for image formation. Therefore, the term $f(x, y)^{\gamma} \times n(x, y)$ in Eq. (1) represents the image-dependent noise and can be called the image residue. The effort to eliminate this residue term is laborious, although residual learning can help achieve the same.

ResCowNet employs a 3-level DT-CWT layer that decomposes noisy images into high-pass and low-pass sub-bands, which are then concatenated into a frame of double the resolution of the input image. The image residue from the sub-band frames is successfully extracted using our three-level deep learning model, which includes residual and convolutional blocks. This residue is then combined with the corresponding input of each stage, and finally, the image is inversely transformed to produce the despeckled image.

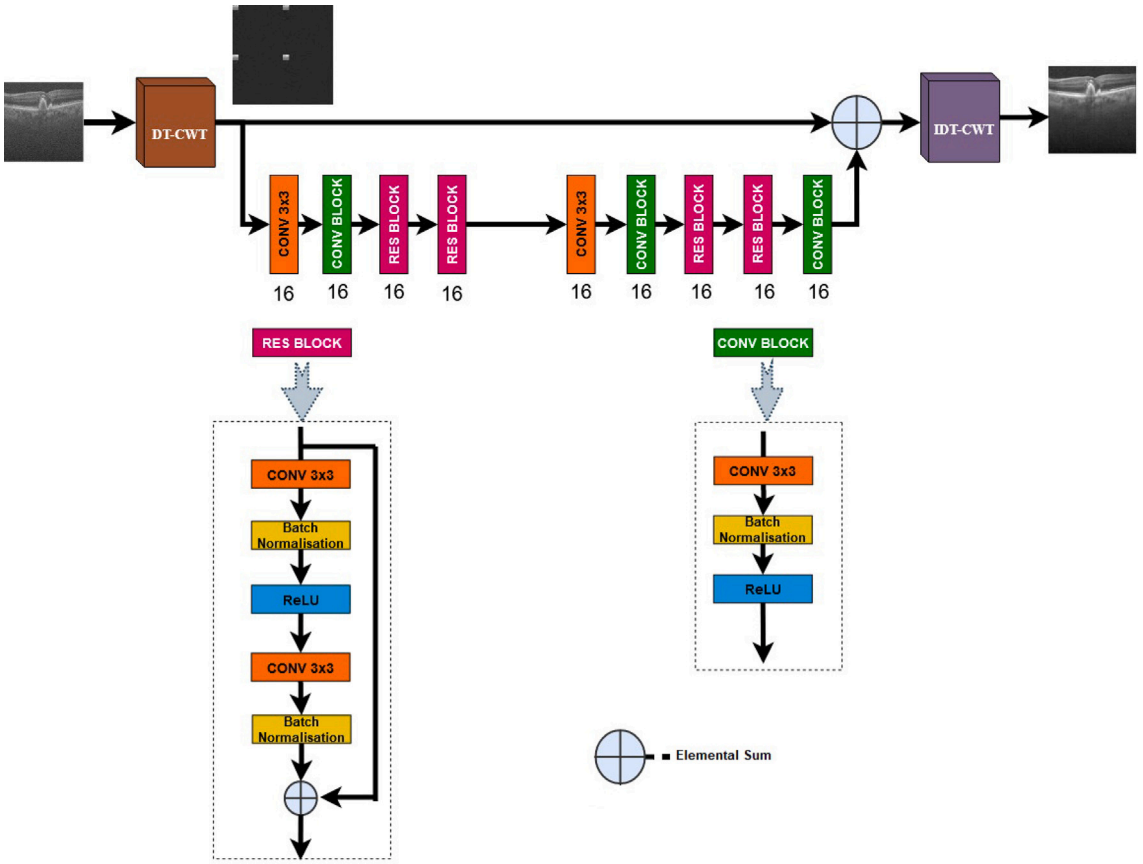

Fig. 1. Single level ResCoWNet architecture with convolutional blocks (shown in green) and residual blocks (shown in ruby pink). Numbers shown under each block denote the number of filters used in the respective convolution layer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

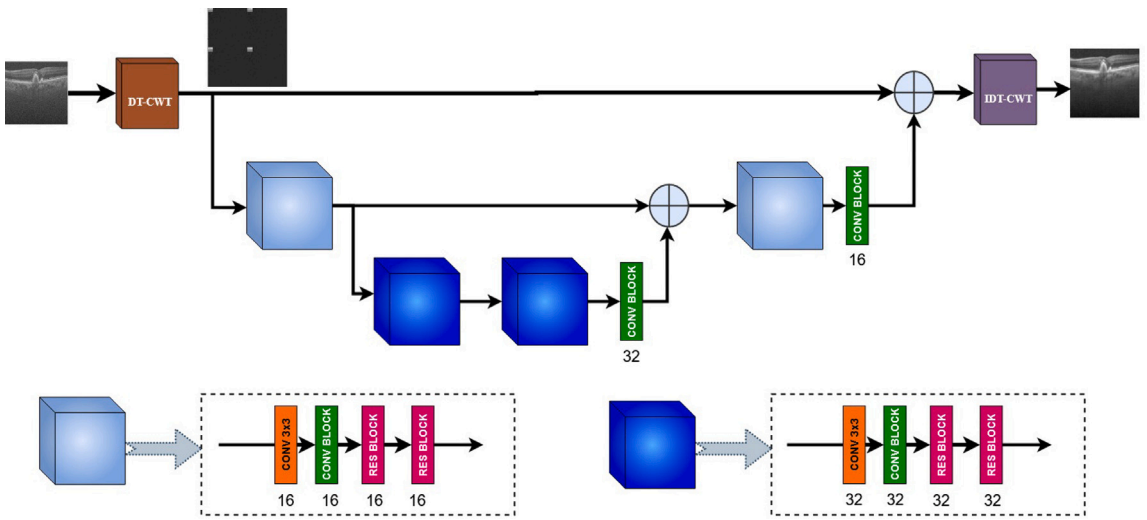

Fig. 2. Double level ResCoWNet architecture. The cubes correspond to a cascade of convolution layers, one convolutional block, and two residual blocks as used in single-level ResCoWNet. The first cube, represented in sky blue, utilizes 16 filters for the convolution layers, while the second cube, shown in lapis blue, implements a similar cascade but with 32 filters in the convolution layers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

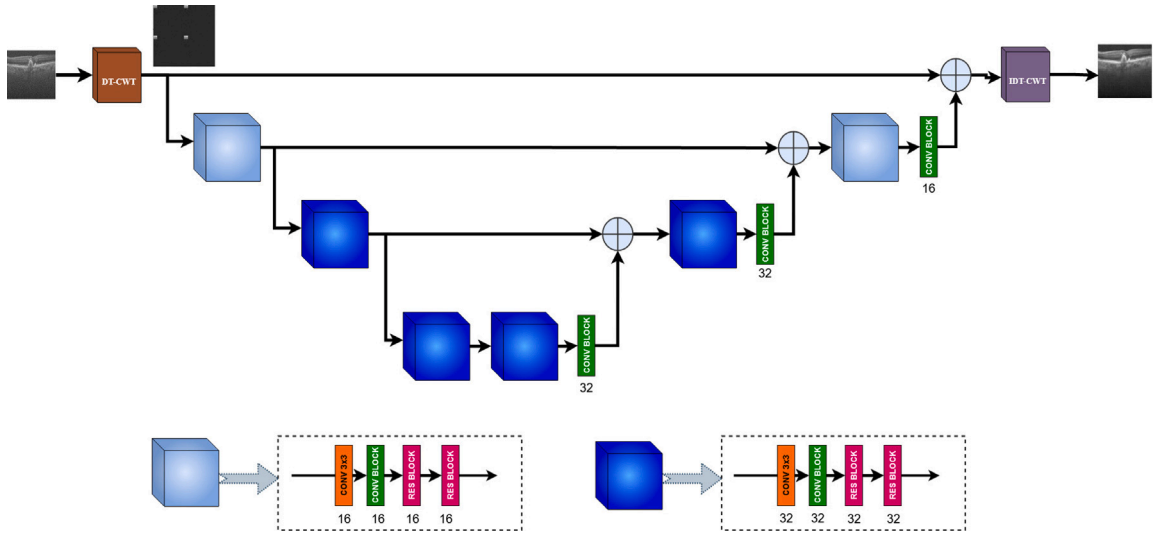

Fig. 3. Three level ResCoWNet architecture. The cubes correspond to a cascade of convolution layer, one convolutional block, and two residual blocks, similar to those used in double-level ResCoWNet.

3.1.1. Dual tree complex wavelet transform (DT-CWT)

Although wavelet bases are in certain ways optimum for a wide class of 1-D signals, the 2-D wavelet transform lacks these optimal qualities for natural images [37]. Moreover, the 2-D wavelet transform fails to segregate orientations resulting in checkerboard artifacts. As a solution, dual-tree complex wavelet transform (DT-CWT) was proposed to produce oriented wavelets. The row–column implementation of 2-D DT-CWT uses six complex wavelets,

$$\xi_1(m,n) = \phi(m) \times \psi(n) \tag{2}$$

$$\xi_2(m,n) = \phi(m) \times \overline{\psi(n)} \tag{3}$$

$$\xi_3(m,n) = \psi(m) \times \phi(n) \tag{4}$$

$$\xi_4(m,n) = \psi(m) \times \overline{\phi(n)} \tag{5}$$

$$\xi_5(m,n) = \psi(m) \times \psi(n) \tag{6}$$

$$\xi_6(m,n) = \psi(m) \times \overline{\psi(n)} \tag{7}$$

where $\phi(m)$ and $\psi(m)$ are complex wavelets denoted as $\phi_h(m)+j\phi_g(m)$ and $\psi_h(m)+j\psi_g(m)$ respectively. $\phi_g(m)$ or $\psi_g(m)$ are the Hilbert transform of $\phi_h(m)$ or $\psi_h(m)$ respectively. $\overline{\psi(m)}$ or $\overline{\phi(m)}$ represent the conjugate of $\psi(m)$ and $\phi(m)$ respectively. The real part of the six complex wavelets ξ_i, $i = 1, 2, \ldots, 6$ results in 6 different orientations viz. 15°, −15°, 45°, −45°, 75° and −75°.

In this paper, the TensorFlow-wavelets library is used to implement the DT-CWT layer in ResCoWNet. The six subbands produced after DT-CWT is concatenated to form a bigger image twice the size of the input, as shown in Fig. 4. This concatenated transform image is then passed on to different levels of ResCoWNet.

3.1.2. CONV+BN+ReLU

A convolution layer (CONV) is a crucial component of the CNN design that extracts features from the input images. A convolution is a linear operation used for extracting features in which a fixed array of numbers called a kernel is slid over the input, a tensor of numbers [38]. At each point of the tensor, a dot product of each kernel or filter element and the input tensor is computed and added to generate the output tensor, also known as the activation map or feature map. To apply multiple layers, there is zero padding to retain image dimensions and a stride; the distance between two successive filter positions is chosen as unity. Convolution results in the sharing of kernels across all image positions, enabling us to detect local patterns more easily. The feature maps are passed through a nonlinear activation function following a linear operation such as convolution. The rectified linear unit (ReLU), which computes the function $\mathbf{f_j}$ as given in Eq. (8), is the most frequently used nonlinear activation function.

$$a_{j,i} = \mathbf{f_j}(Z_{1,i}, Z_{2,i} \ldots Z_{n,i}) = max(0, Z_{j,i}) = \begin{cases} Z_{j,i}, Z_{j,i} > 0 \\ 0, otherwise \end{cases} \tag{8}$$

where $a_{j,i}$, $j = 1, 2 \ldots n$ and $i = 1, 2 \ldots m$ corresponds to the activation layer output of jth neuron of the current layer fed with ith training sample. m and n are the total numbers of training samples and the number of neurons in the current layer, respectively. $Z_{j,i}$ denotes the weighted sum of the activation outputs of the previous layer shifted by appropriate bias.

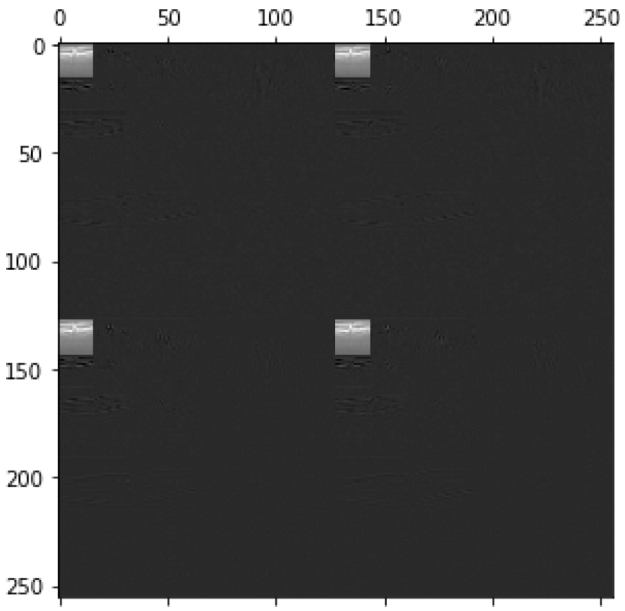

Fig. 4. Subbands of DT-CWT arranged as a single frame.

As the parameters of the preceding layers change during training, the distribution of each layer's inputs changes. This makes it challenging for training models with saturating nonlinearities since they need lower learning rates and careful parameter selection [26]. This problem, called internal covariate shift, is addressed by normalizing the layer inputs. The BN transform is included immediately before the nonlinearity or the ReLU by normalizing $Z_{j,i}$. For different components of the same feature map, normalized in the same way at various locations, it is necessary to normalize all the activations in a mini-batch across all locations. BN transform is implemented by Eq. (3) across the network in mini-batches.

$$\hat{Z}_{j,i}^{(k)} = \frac{Z_{j,i}^{(k)} - \mu_k}{\sqrt{\sigma_k^2 + c}} \tag{9}$$

where $\hat{Z}_{j,i}^{(k)}$ is the normalized activation map of the kth mini-batch, c is the constant for numerical stability, μ_k is the mini-batch mean, and σ_k^2 is the mini-batch variance. Issues such as vanishing gradients can be addressed by batch normalization by preventing small parameter changes from amplifying into bigger and less optimum ones. BN also allows faster convergence while training, aiding a faster algorithm implementation.

3.1.3. Residual learning

ResCoWNet uses residual blocks that adopt residual learning methodology wherein a single residual unit is utilized to sieve out the residue $R(y)$ of the noisy image y. Our residual block consists of CONV+BN+ReLU, CONV+BN layers, and a skip connection from input, as demonstrated in Fig. 1. A total of 4 residual blocks are used in each layer, and residual mapping in our network results in faster and more stable convergence than the original mapping.

3.2. Network training

The ResCoWNet architecture uses the 26 OCT noisy-averaged image pairs provided in [39]. 536 training images, resized to 128×128, are created using 67 various augmentation processes on 8 randomly chosen raw OCT images and their related averaged versions. Four noisy-averaged OCT image pairs are used for validation. We have run the model for 300 epochs for single, double, and triple levels to substantiate the requirement of the three levels in our network. The averaged mean square error between residual images and estimated noise from the noisy images is the loss function, l used to learn the trainable parameters Φ in our network.

$$l(\Phi) = \frac{1}{2N} \sum_{i=1}^{N} \| R(y_i; \Phi) - (y_i - x_i) \|_F^2 \tag{10}$$

where (y_i, x_i) pair indicates N noisy-clean training image pairs and the subscript F denoting Frobenius norm represents L_2 norm.

As it is required to alter the weights of each epoch and minimize the mean square error loss, choosing the right weights and learning rate is an important task. Optimizers alleviate the severity of this task. In our work, we have used an Adaptive Moment Estimation (Adam) optimizer with a learning rate of 10^{-4}. Rather than any optimization strategy, our network's best denoising performance comes from integrating residual learning with batch normalization.

4. Experimental results and discussions

4.1. Evaluation metrics

For the performance evaluation of our DespNet, we use the following objective evaluation metrics:

1. **Peak Signal to Noise Ratio (PSNR)**: It is defined as the decibel ratio of the highest pixel intensity of the input image to the Mean Square Error (MSE). [40,41]. That is:

$$PSNR = 10log_{10}\left(\frac{MAX_I}{MSE}\right) \tag{11}$$

$$where, MSE = \frac{1}{M \times N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}|I(x,y) - I_n(x,y)|^2 \tag{12}$$

where I is the original input image and I_n is its noisy approximation.

2. **Image Enhancement Factor (IEF)**: The ratio of the mean square error between the original picture and its noisy form, as well as the mean square error between the denoised image and its noisy version, is used to compute IEF [16].

$$IEF = \frac{MSE_{Original,Noisy}}{MSE_{Original,Denoised}} \tag{13}$$

3. **Structural Similarity Index (SSIM)**: The structural similarity index indices the degree of similarity between the images [42]. Better noise-removal methods are achieved with larger SSIM. SSIM is found using Eq. (9), given below.

$$SSIM = \frac{(2\mu_1\ \mu_2 + C_1)(2\sigma_{1,2} + C_2)}{\left((\mu_1)^2 + (\mu_2)^2 + C_1\right)\left((\sigma_1)^2 + (\sigma_2)^2 + C_2\right)} \tag{14}$$

where μ_1, μ_2 are mean, σ_1, σ_2 are the variances of the original noiseless image, and the denoised image respectively. $\sigma_{1,2}$ denotes the covariances between images.

4. **Equivalent Number of Looks (ENL)**: A popular metric for gauging homogeneity in image areas following noise removal [43].

$$ENL = \frac{\mu_{ROI}^2}{\sigma_{ROI}^2} \tag{15}$$

where μ_{ROI} and σ_{ROI}^2 denote the mean and variance of the region of interest, respectively.

5. **Average Gradient (AG)**: It measures the average gradient magnitude of the image [44].

$$AG = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}|G(x,y)|}{M \times N} \tag{16}$$

where, $|G(x,y)|$ is the gradient magnitude. The gradient is computed using the Sobel mask, and for better visibility of edges, the value of AG should be high.

6. **Edge Preservation Index (EPI)**: EPI is determined from the Laplacian (∇^2) of the test image [16].

$$EPI = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(\nabla^2 f(x,y) - \mu_f\right)\left(\nabla^2 \hat{f}(x,y) - \mu_{\hat{f}}\right)}{\sqrt{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(\nabla^2 f(x,y) - \mu_f\right)^2 \times \sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(\nabla^2 \hat{f}(x,y) - \mu_{\hat{f}}\right)^2}} \tag{17}$$

where f and $\hat{f}$ are the image with and without noise respectively.μ_f and $\mu_{\hat{f}}$ denotes the empirical mean of f and $\hat{f}$ respectively. An EPI value lies in the range of 0 and 1, and a value equal to one indicates that the algorithm well preserves the edges [23].

7. **Speckle Suppression and Mean Preservation Index (SMPI)**: SMPI denotes the efficiency of the denoising algorithm in terms of mean preservation and noise reduction [43].

$$SMPI = Q \times \frac{\sigma_{denoised}}{\sigma_{noisy}} \tag{18}$$

where,

$$Q = \frac{max(I_{denoised} - min(I_{noisy}))}{\mu_{noisy}} + |\mu_{denoised} - \mu_{noisy}| \tag{19}$$

where $I_{denoised}$ and I_{noisy} are the noise removed and noisy versions of the image used. Stronger speckle noise reduction is possible with lower SMPI values.

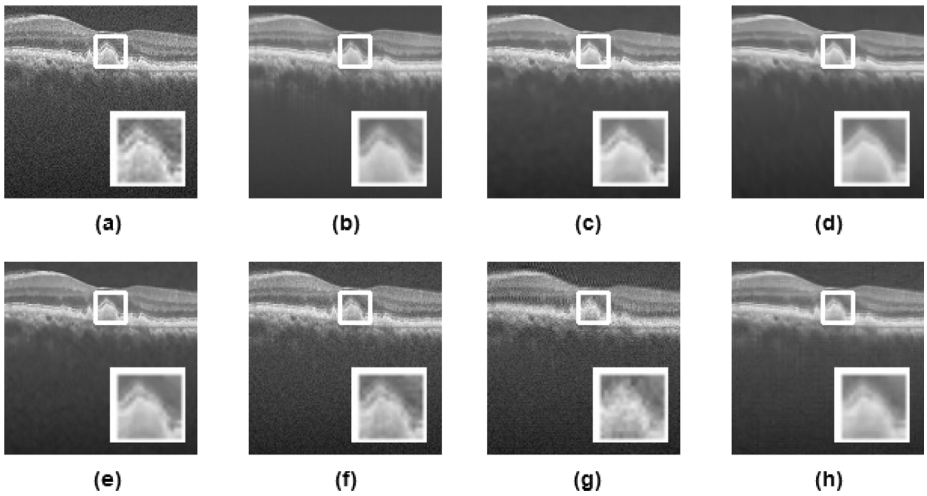

Fig. 5. (a) A noisy OCT image from [39] with speckle noise and a region of interest (ROI) enlarged to see the noise effect,(b) Averaged image from [39] with ROI, which is noise free (c) Despeckled image using PNLM [9] with ROI, which is moderately speckle free (d) Despeckled image using WGLRR [14] with ROI, note that while the image is free of speckles, there is a loss in structural details (e)Despeckled image after DT-CWT fusion algorithm [16] with ROI, which is moderately speckle-free and preserves edge details(f) Despeckled image using MWCNN [30] with ROI which still contains some noise (g) Despeckled image using MWRN [31]with ROI, that fails in preserving edges and also contains speckles (h) Despeckled image using the proposed method with ROI, which is speckle free and preserves edge details.

Table 1
Results of quantitative analysis.

Methodology	PSNR[a]	IEF[a]	ENL[a]	SSIM[a]	AG[a]	EPI[a]	SMPI[a]
PNLM [9]	30.47	2.98	8.85	0.87	0.11	0.69	3.44
WGLRR [14]	31.34	3.99	8.96	0.87	0.11	0.62	3.26
DT-CWT fusion [16]	30.45	3	8.69	0.83	0.13	0.69	3.3
MWCNN [30]	27.83	1.5	8.29	0.56	0.18	0.7	3.29
MWRN [31]	26.86	1.21	7.63	0.54	0.18	0.09	3.42
ResCoWNet	31.05	3.45	9.13	0.8	0.13	0.57	3.22

[a]The figures displayed for each method are the average metric values derived from the collection of 14 test images supplied in [39].

4.2. Performance analysis

After training the network with 536 images of size 128×128 for 300 epochs as described in Section 3.2, we test our ResCoWNet model using the remaining 14 images in the dataset provided in [39] to analyze the despeckling performance of the network. The model is implemented in Jupyter using TensorFlow library on a desktop PC with 64 GB RAM running at 3.4 GHz and dual Nvidia Geforce RTX 2080 Ti 11 GB.

The efficacy of our ResCoWNet model is demonstrated by comparing it with some of the recent OCT denoising algorithms. ResCoWNet results have been juxtaposed with PNLM [9], WGLRR [14], DT-CWT fusion [16], MWWCNN [30] and MWRN [31] in Figs. 5, 6 and 7.

Table 1 summarizes the objective evaluation done with the evaluation metrics discussed in Section 4.1. The graphical plot in Fig. 8 shows how our ResCoWNet and other algorithms respond to the objective assessment metrics (PSNR, IEF, ENL, SSIM, AG EPI and SMPI) when tested with the 14 images in the dataset of [39].

4.3. Ablation study

We have analyzed the architecture of our ResCoWNet level by level. Single and double-level ResCowNet, denoted as ResCoWNet-S and ResCoWNet-D, shown in Figs. 1 and 2, respectively, have been trained for 300 epochs and tested with 14 images. The objective evaluation carried out using the evaluation criteria outlined in Section 4.1 is summarized in Table 2, which leads to a conclusion about the adoption of ResCoWNet's three-level design. The three-level architecture of ResCoWNet, as illustrated in Fig. 3, has improved PSNR, IEF, SSIM, and EPI than its less-level counterparts, such as ResCoWNet-S and ResCoWNet-D.

4.4. Inferences

From Table 1, it can be inferred that WGLRR [14] has the best PSNR, IEF, and SSIM values among all the methods compared. However, the AG and EPI values of this method are lesser than PNLM [9]. Despite having superior edge and structure preservation

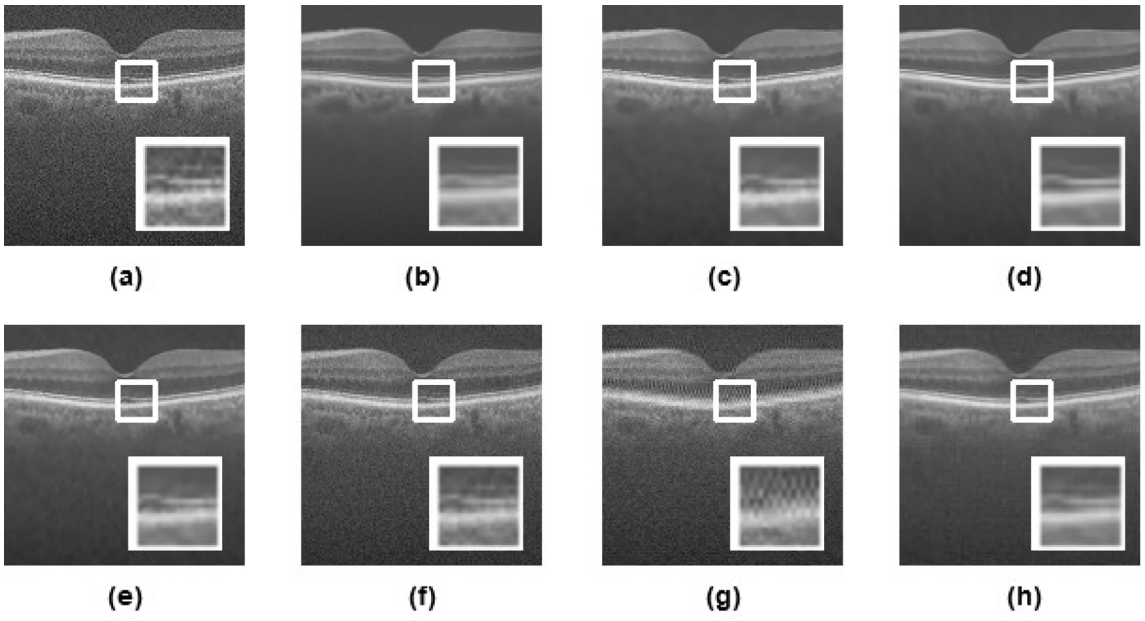

Fig. 6. (a) A noisy OCT image from [39] with speckle noise and a region of interest (ROI) enlarged to see the noise effect,(b) Averaged image from [39] with ROI, which is noise free (c) Despeckled image using PNLM [9] with ROI, which is moderately speckle free (d) Despeckled image using WGLRR [14] with ROI, note that while the image is free of speckles, there is a loss in structural details (e)Despeckled image after DT-CWT fusion algorithm [16] with ROI, which is moderately speckle-free and preserves edge details(f) Despeckled image using MWCNN [30] with ROI which still contains some noise (g) Despeckled image using MWRN [31]with ROI, that fails in preserving edges and also contains speckles (h) Despeckled image using the proposed method with ROI, which is speckle free and preserves edge details.

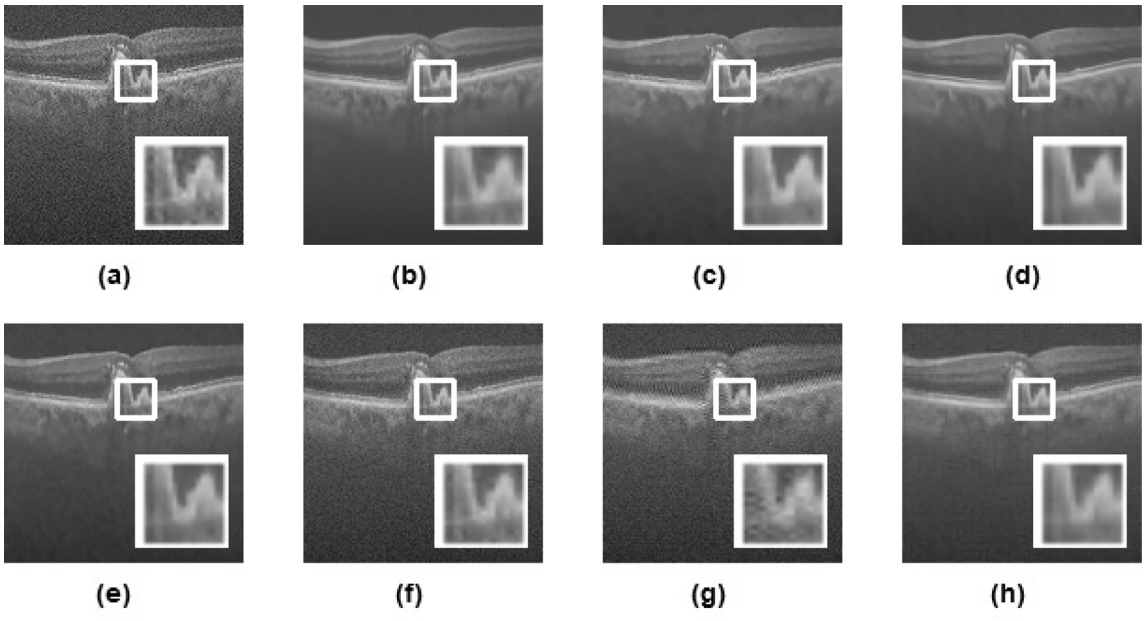

Fig. 7. (a) A noisy OCT image from [39] with speckle noise and a region of interest (ROI) enlarged to see the noise effect,(b) Averaged image from [39] with ROI, which is noise free (c) Despeckled image using PNLM [9] with ROI, which is moderately speckle free (d) Despeckled image using WGLRR [14] with ROI, note that while the image is free of speckles, there is a loss in structural details (e)Despeckled image after DT-CWT fusion algorithm [16] with ROI, which is moderately speckle-free and preserves edge details(f) Despeckled image using MWCNN [30] with ROI which still contains some noise (g) Despeckled image using MWRN [31]with ROI, that fails in preserving edges and also contains speckles (h) Despeckled image using the proposed method with ROI, which is speckle free and preserves edge details.

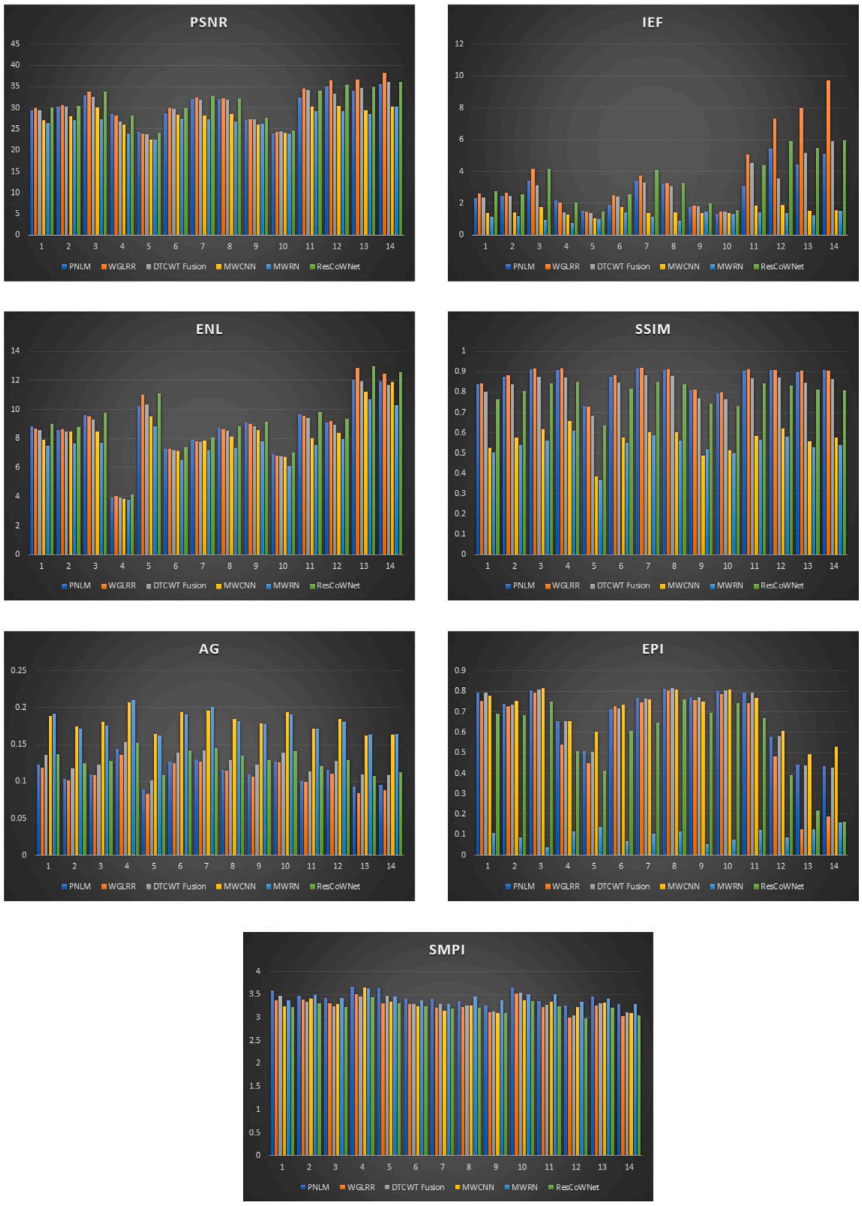

Fig. 8. Graph showing the quantitative assessment measures for each of the 14 test images in the dataset [39].

Table 2
Results of quantitative analysis of level by level ResCoWNet model.

Methodology	PSNR[a]	IEF[a]	ENL[a]	SSIM[a]	AG[a]	EPI[a]	SMPI[a]
ResCoWNet-S	30.84	3.22	9.16	0.76	0.13	0.54	3.20
ResCoWNet-D	31.02	3.4	9.17	0.79	0.13	0.53	3.20
ResCoWNet	31.05	3.45	9.13	0.8	0.13	0.57	3.22

[a]The numbers shown for each approach are the average of metric values obtained from the collection of 14 test images provided in [39].

capabilities than WGLRR, as shown by the EPI and SSIM values in Table 1, PNLM cannot successfully suppress the speckle, as seen from the SMPI value.

Our previous work [16], stays optimum between PNLM and WGLRR, although its PSNR, IEF, and SSIM values are lesser than WGLRR. MWCNN [30], a wavelet-based CNN, has lower PSNR, IEF, SSIM, and ENL values, indicating that the model for OCT denoising needs improvement. The output despeckled OCT pictures of MWRN, which used residual learning in a wavelet-based

CNN, have the lowest assessment results. This approach produced several artifacts that would seriously impair subsequent illness detection.

ResCoWNet, our proposed network, has responded optimally to the objective analysis with a PSNR value close to WGLRR. The approach achieved our main goal of having the lowest SMPI value among other methods while still having adequate IEF and SSIM values. Thus, it can be concluded that our proposed residual complex wavelet CNN or ResCoWNet, has produced a denoised image that has suppressed the speckle noise while maintaining the structural features and overall image quality.

5. Conclusion

We have proposed ResCoWNet—residual learning and complex wavelet-based deep CNN for speckle noise removal from OCT images. The network employs a residual learning strategy to extract clean images from noisy observations by predicting residual images. ResCoWNet is trained only on OCT images that have undergone 3-level DT-CWT transformation and had their subbands processed through three levels of residual-convolutional blocks. The subjective and objective analysis discussed in the paper explains the restoration quality of our model in speckle noise removal. Our network has high PSNR, SSIM, IEF, and a low SMPI value among the compared deep learning approaches, denoting better speckle suppression and preservation of image quality, making it an effective method for preprocessing OCT images that have been heavily tainted by speckle noise.

Declaration of competing interest

The authors declare that they have no conflict of interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to the authors of the datasets in [39], using which we have done a detailed analysis of our algorithm. We also express our gratitude to the authors of [9,14] for making their implementation codes publicly available.

Ethical approval statement

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] S. Adabi, A. Clayton, S. Conforto, A. Hojjat, A.G. Podoleanu, M. Nasiriavanaki, Mitigation of speckle noise in optical coherence tomograms, in: Optics, Photonics and Laser Technology, Springer, 2018, pp. 115–135.
[2] A. Pizurica, L. Jovanov, B. Huysmans, V. Zlokolica, P. De Keyser, F. Dhaenens, W. Philips, Multiresolution denoising for optical coherence tomography: A review and evaluation, Curr. Med. Imaging 4 (4) (2008) 270–284.
[3] M. Samieinasab, Z. Amini, H. Rabbani, Multivariate statistical modeling of retinal optical coherence tomography, IEEE Trans. Med. Imaging 39 (11) (2020) 3475–3487.
[4] J.K. Mandal, S. Banerjee, J. Kacprzyk, Intelligent Computing: Image Processing Based Applications, Springer, 2020.
[5] A. Aksel, A.D. Gilliam, J.A. Hossack, S.T. Acton, Speckle reducing anisotropic diffusion for echocardiography, in: Conference Record - Asilomar Conference on Signals, Systems and Computers, Vol. 11, 2006, pp. 1988–1992, http://dx.doi.org/10.1109/ACSSC.2006.355113.
[6] D. Barash, A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 844–847, http://dx.doi.org/10.1109/TPAMI.2002.1008390.
[7] A. Buades, B. Coll, J.-M. Morel, Non-local means denoising, Image Process. Line 1 (2011) 208–212.
[8] Institute of Electrical and Electronics Engineers, IEEE Engineering in Medicine and Biology Society, IEEE Signal Processing Society, 2008 5th IEEE International Symposium on Biomedical Imaging : from Nano to Macro : Proceedings : May 14-17, 2008, paris Marriott Rive Gauche Hotel & Conference Center, Paris, France, IEEE, 2008.
[9] H. Yu, J. Gao, A. Li, Probability-based non-local means filter for speckle noise suppression in optical coherence tomography images, Opt. Lett. 41 (2016) 994, http://dx.doi.org/10.1364/ol.41.000994.
[10] R. Garnett, T. Huegerich, C. Chui, W. He, A universal noise removal algorithm with an impulse detector, Image (Rochester, N.Y.) 14 (2005) 1747–1754.
[11] W. Fan, H. Yu, T. Chen, S. Ji, OCT image restoration using non-local deep image prior, Electronics (Switzerland) 9 (2020) http://dx.doi.org/10.3390/electronics9050784.
[12] L. Fang, S. Li, Q. Nie, J.A. Izatt, C.A. Toth, S. Farsiu, Sparsity based denoising of spectral domain optical coherence tomography images, Biomed. Optics Express 3 (2012) 927, http://dx.doi.org/10.1364/boe.3.000927.
[13] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (2006) 4311–4322, http://dx.doi.org/10.1109/TSP.2006.881199.
[14] C. Tang, L. Cao, J. Chen, X. Zheng, Speckle noise reduction for optical coherence tomography images via non-local weighted group low-rank representation, Laser Phys. Lett. 14 (2017) http://dx.doi.org/10.1088/1612-202X/aa5690.
[15] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for sparse coding, J. Mach. Learn. Res. 11 (2010) 19–60, http://dx.doi.org/10.1145/1553374.1553463.
[16] A. P.S., V.P. Gopi, P. Palanisamy, Despeckling of OCT images using DT-CWT based fusion technique, Optik (2022) 169332, http://dx.doi.org/10.1016/j.ijleo.2022.169332.

[17] F. Zhao, X. Xie, Energy minimization in medical image analysis: Methodologies and applications, Int. J. Numer. Methods Biomed. Eng. 32 (2016) http://dx.doi.org/10.1002/cnm.2733.
[18] Denoising optical coherence tomography using second order total generalized variation decomposition, Biomed. Signal Process. Control 24 (2016) 120–127, http://dx.doi.org/10.1016/j.bspc.2015.09.012.
[19] X. Wang, X. Yu, X. Liu, S. Chen, S. Chen, N. Wang, L. Liu, A two-step iteration mechanism for speckle reduction in optical coherence tomography, Biomed. Signal Process. Control 43 (2018) 86–95, http://dx.doi.org/10.1016/j.bspc.2018.02.011.
[20] M. Shamouilian, I. Selesnick, Total variation denoising for optical coherence tomography, in: 2019 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2019 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2019, http://dx.doi.org/10.1109/SPMB47826.2019.9037832.
[21] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising, IEEE Trans. Image Process. 26 (2017) 3142–3155, http://dx.doi.org/10.1109/TIP.2017.2662206.
[22] B. Qiu, Z. Huang, X. Liu, X. Meng, Y. You, G. Liu, K. Yang, A. Maier, Q. Ren, Y. Lu, Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function, Biomed. Optics Exp. 11 (2020) 817, http://dx.doi.org/10.1364/boe.379551.
[23] N.A. Kande, R. Dakhane, A. Dukkipati, P.K. Yalavarthy, SiameseGAN: A generative model for denoising of spectral domain optical coherence tomography images, IEEE Trans. Med. Imaging 40 (2021) 180–192, http://dx.doi.org/10.1109/TMI.2020.3024097.
[24] Z. Chen, Z. Zeng, H. Shen, X. Zheng, P. Dai, P. Ouyang, DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images, Biomed. Signal Process. Control 55 (2020) http://dx.doi.org/10.1016/j.bspc.2019.101632.
[25] Y. Zhou, K. Yu, M. Wang, Y. Ma, Y. Peng, Z. Chen, W. Zhu, F. Shi, X. Chen, Speckle noise reduction for OCT images based on image style transfer and conditional GAN, IEEE J. Biomed. Health Inf. 26 (2022) 139–150, http://dx.doi.org/10.1109/JBHI.2021.3074852.
[26] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML '15, JMLR.org, 2015, pp. 448–456.
[27] N. Gour, P. Khanna, Speckle denoising in optical coherence tomography images using residual deep convolutional neural network, Multimedia Tools Appl. 79 (2020) 15679–15695, http://dx.doi.org/10.1007/s11042-019-07999-y.
[28] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 2480–2495, http://dx.doi.org/10.1109/TPAMI.2020.2968521.
[29] W. Bae, J. Yoo, J. Chul Ye, Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.
[30] P. Liu, H. Zhang, W. Lian, W. Zuo, Multi-level wavelet convolutional neural networks, IEEE Access 7 (2019) 74973–74985, http://dx.doi.org/10.1109/ACCESS.2019.2921451.
[31] Y. Peng, Y. Cao, S. Liu, J. Yang, W. Zuo, Progressive training of multi-level wavelet residual networks for image denoising, 2020, URL http://arxiv.org/abs/2010.12422.
[32] Q. Li, L. Shen, S. Guo, Z. Lai, WaveCNet: Wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification, IEEE Trans. Image Process. 30 (2021) 7074–7089, http://dx.doi.org/10.1109/TIP.2021.3101395.
[33] Multilevel wavelet-SRNet for SAR target recognition, IEEE Geosci. Remote Sens. Lett. 19 (2022) http://dx.doi.org/10.1109/LGRS.2021.3050891.
[34] Y. Duan, F. Liu, L. Jiao, P. Zhao, L. Zhang, SAR image segmentation based on convolutional-wavelet neural network and markov random field, Pattern Recognit. 64 (2017) 255–267, http://dx.doi.org/10.1016/j.patcog.2016.11.015.
[35] SAR despeckling based on CNN and Bayesian estimator in complex wavelet domain, IEEE Geosci. Remote Sens. Lett. 19 (2022) http://dx.doi.org/10.1109/LGRS.2022.3185557.
[36] P. Kokil, S. Sudharson, Despeckling of clinical ultrasound images using deep residual learning, Comput. Methods Programs Biomed. 194 (2020) http://dx.doi.org/10.1016/j.cmpb.2020.105477.
[37] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Process. Mag. 22 (2005) 123–151, http://dx.doi.org/10.1109/MSP.2005.1550194.
[38] R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi, Convolutional neural networks: An overview and application in radiology, Insights Into Imaging 9 (2018) 611–629, http://dx.doi.org/10.1007/s13244-018-0639-9.
[39] H. Chen, Optical Coherence Tomography Image Enhancement Using Residual Encoder-Decoder CycleGAN, IEEE Dataport, 2021, http://dx.doi.org/10.21227/1t28-gb80.
[40] K.J. Sabareesaan, J. Jaya, R. Varahamoorthi, Performance analysis of various filters for noise removal in EDM electrode surface crack images, Int. J. Appl. Sci. Manag. 3 (2017) 317–328.
[41] M. Saeedzarandi, H. Nezamabadi-pour, S. Saryazdi, A. Jamalizadeh, Image denoising in undecimated dual-tree complex wavelet domain using multivariate t-distribution, Multimedia Tools Appl. 79 (2020) 22447–22471.
[42] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.
[43] A. Shamsoddini, J.C. Trinder, A. Shamsoddini, J.C. Trinder, Image texture preservation in speckle noise suppression, 2010.
[44] H. Chen, Fusion denoising algorithm of optical coherence tomography image based on point-estimated and block-estimated, Optik 225 (2021) 165864, http://dx.doi.org/10.1016/j.ijleo.2020.165864.