OCT B-SCAN ENHANCEMENT USING SVHT PAIRED BM3D, EDSR AND CLASSIFICATION USING RESNET18-BASED CNN

Nadun Rajapaksha, Sasika Amarasinghe, Devnith Wijesinghe, Ravindu Weerakoon, Nidula Gunawardana, Kavishka Abeywardane, Hasitha Gallella, Chamith Ranathunga Sandushan Ranaweera, Chamira Edussooriya

Department of Electronic and Telecommunications Engineering,
Department of Computer Science and Engineering,
University of Moratuwa

ABSTRACT

This work addresses the computational challenges faced by real-time Optical Coherence Tomography (OCT) image analysis, particularly in the context of noise reduction, resolution enhancement, and automated classification. We propose a novel and efficient methodology that begins with Block-matching 3D Filtering (BM3D) followed by Singular Value Hard Thresholding (SVHT) for effective noise reduction, significantly improving image quality. For resolution enhancement, we employ a pre-trained Enhanced Deep Residual Networks for Single Image Super-Resolution model (EDSR) network, which efficiently upscale images while preserving detail. The final stage integrates a deep learning framework, utilizing a 2D convolutional neural network with a ResNet18 backbone, to automate disease classification. This comprehensive approach aims to overcome the limitations of current methods, providing a robust solution for real-time OCT image analysis.

***Index Terms*—** OCT, OCT denoising, OCT super resolution, OCT classification, medical imaging

1. INTRODUCTION

Retinal imaging plays a vital role in diagnosing ocular and systemic diseases. Optical coherence tomography (OCT) has emerged as a powerful non-invasive tool, leveraging coherence light and interferometry to capture cross-sectional images of the retina. This technology provides valuable insights for identifying pathologies like age-related macular degeneration, diabetic macular edema, diabetic retinopathy, and choroidal neovascularization. While OCT offers high-resolution imaging, to achieve real-time acquisition, lower sampling rates might be employed, leading to a trade-off between speed and resolution. This can manifest as reduced inherent image resolution and the introduction of aliasing artifacts. Additionally, the inherent coherence of the light source in OCT inherently introduces speckle noise to the images.

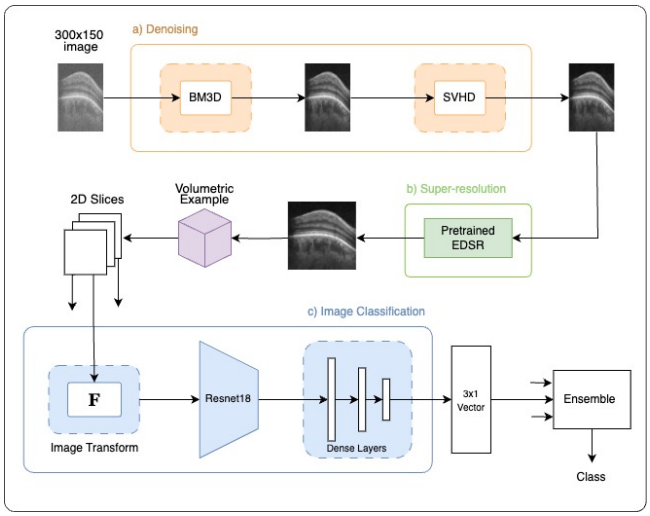

Fig. 1. block diagram of the overall enhancement and classification pipeline

Conventional OCT image denoising and enhancement have been approached in means of statistical modeling and stochastic differential equations to characterize image structure and noise patterns [1, 2, 3]. Other conventional methods for denoising includes Gaussianization transform methods [4]. [5] have explored graph based methods as well.

These methods often struggle with real-time implementation due to their computational demands. Therefore, developing a real-time model capable of accurately capturing noise distributions and enhancing image resolution remains an important challenge for optimal OCT image analysis. In the domain of medical imaging, particularly in OCT, the enhancement of image resolution through super-resolution techniques is crucial for improving diagnostic accuracy and detail from low-resolution scans. This paper explores the evolution of SR methodologies from foundational approaches such as bicubic interpolation, Lanczos resampling to advanced models like EDSR and enhanced GAN models. Emphasizing the

enhancement of perceptual quality and fidelity in OCT image reconstruction, our research addresses significant challenges and evaluates state-of-the-art techniques in medical image processing.

Motivated by the limitations of current methods in real-time OCT image analysis, our work explores a novel and efficient approach for denoising, resolution enhancement, and automated classification. Our approach BM3D is followed by SVHT for effective noise reduction, leading to improved image quality and interpretability. Subsequently, EDSR [6], a CNN model, is employed for efficient image upscaling and resolution enhancement. Finally, a deep learning model consisting of a ResNet18 backbone followed by a Multi-Layer Perceptron is implemented to achieve automated disease classification.

2. PROPOSED METHODOLOGY

The problem description addresses OCT image enhancement and classification in three approaches. The given OCT dataset, consisting of noisy, low-resolution B-scans, must be processed in order to **(1) reduce noise, (2) increase resolution**, and **(3) classify each subject into three categories**. We propose the pipeline system in Fig. 1 to complete all the tasks mentioned above.

2.1. Denoising with BM3D + SVHT

In our proposed solution, we tackle the denoising problem using a combination of two popular conventional denoising techniques namely; SVHT and BM3D. We found that thresholding singular values of an OCT B-scan improves the Signal-to-Noise Ratio (SNR) of the image and reduces fine noise. However it was not sufficient to enhance the contrast and remove speckle noise. Alternatively, using BM3D independently showed more positive results in terms of improved Contrast-to-Noise Ratio (CNR) and contrast. Since both these methods contribute to remove a specific component of noise, we combined both these methods to create a pipelined system. In this flow, we first pass the image into the BM3D algorithm for initial enhancement. We opted for the BM3D algorithm over state-of-the-art deep-learning methods due to the absence of ground-truth denoised images in the provided dataset. Furthermore, conventional methods such as BM3D offer computational efficiency that surpasses that of deep-learning methods, making them more suitable for our application.

Here, we use an implementation of BM3D with collaborative filtering [7]. More specifically, we selected Hard Thresholding and Wiener Filtering techniques included in this algorithm. The standard deviation of Noise Power Spectral Density (NPSD) is estimated to be $\sigma = 29$. The filtered image is then SV decomposed and reconstructed by hard thresholding the singular values.

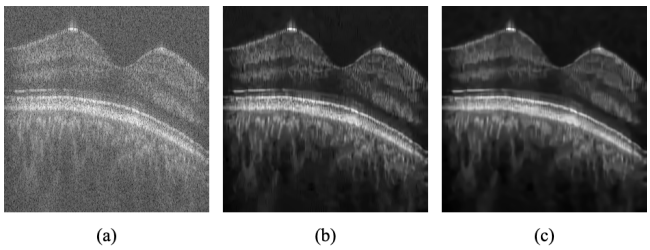

Fig. 2. Results of denoising pipeline. (a) raw image, (b) image after BM3D, (c) image after BM3D+SVHT

SVHT is expressed by [8],

$$\hat{X}_\tau = \sum_{i=1}^{m} \eta_H(y_i;\tau)\mathbf{u}_i\mathbf{v}_i^{'}$$

$$\eta_H(y,\tau) = y\mathbf{1}\{y \geq \tau\}$$

where τ is the threshold. We selected $\tau = 400$ as the thresholding value for SVHT.

This system is carried out in a single processing pipeline in order to avoid the noise added by the encoder-decoder when saving at an intermediary stage. The completed images are then saved as .tiff files to preserve complete information.

2.2. Super-resolution with EDSR

In our work, we used EDSR(*Enhanced Deep Residual Networks for Single Image Super-Resolution*)[6] model to enhance the resolution of test images from 150x300, 200x300 pixels to 300x300 pixels.

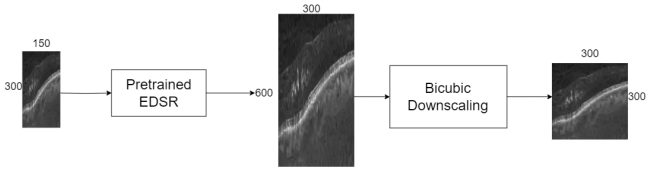

Fig. 3. Super-scaling data flow

The denoised image undergoes processing through a pre-trained EDSR model, which enhances its quality. Subsequently, the image resolution is increased by a factor of 2. Following this enhancement, the image undergoes a downsizing process to achieve a resolution of 300x300 pixels using bicubic interpolation. Please refer to Fig. 3 for a visual representation of these steps. This downsizing via bicubic interpolation filters the image, affecting its sharpness and overall visual fidelity in the downscaled version.

2.3. Subject classification using transfer-learned ensembled CNN

Following the super-resolution of the denoised images, we use the said images to train a *Convolutional Neural Network*

with a pre-trained ResNet18[9] backbone. The ResNet18 encoder is followed by two dense layers with ReLU activation and two dropout layers. The network outputs a 3-vector at the final softmax layer which is used to feed categorical cross-entropy loss. Refer to Fig. 4 for the model architecture.

Before feeding the network, the enhanced greyscale images were transformed into 224x224 images with 3 channels. Images were also normalized in order to avoid vanishing/exploding gradients. During our training loop, we first trained only the fully connected layers for 20 epochs using Adam as the optimizer. This was in order to adapt the overall model to high level features of OCT images. Consequently, for the next 5 epochs we unfreezed the last block of the ResNet18 encoder and trained it with a lower learning rate. The purpose of fine-tuning the ResNet18 backbone is to familiarize the feature encoder to the characteristic features of OCT images. We propose reducing the learning rate to prevent the training of lower-order weights.

At the model input each volumetric example was sliced into 2D images which are subjected to the above mentioned transform. Each sliced image is then forward propagated through the model, to obtain the ensembled average of all slices. When calculating the ensemble average, we only considered the predictions of slices with probability higher than 0.6 to predict the class of the entire volumetric image.

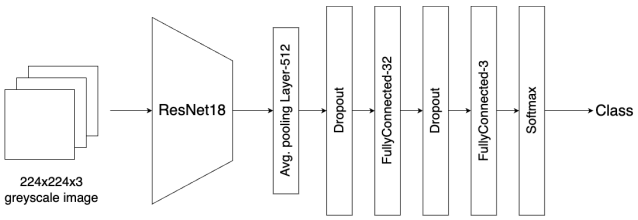

Fig. 4. CNN with ResNet18 backbone model architecture

3. EXPERIMENTS AND RESULTS

The dataset for this task contains 300x300 pixel image volumes of OCT data from 124 subjects. These subjects must be classified into the following groups: Healthy(0), diabetic with DME(1), and non-diabetic patients with other ocular diseases (2).

3.1. Selection of ROIs for evaluation metrics

Region-of-Interest (ROI) was chosen from the denoised image using Otsu's Thresholding method [10] to separate foreground and the background. In Fig. 5, the contour in red separates the foreground and the background image. When calculating Mean-to-Standard Deviation Ratio (MSR), the foreground has been used as the ROI:

$$\text{MSR} = \frac{\mu_f}{\mu_b} \quad .$$

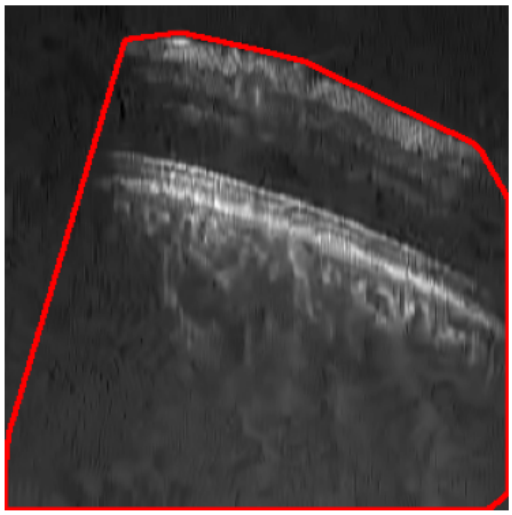

Fig. 5. Contour to separate foreground and background

When calculating CNR [11], the following equation has been used:

$$\text{CNR} = 10 \log_{10} \left(\frac{|\mu_f - \mu_b|}{\sqrt{\frac{\sigma_f^2 + \sigma_b^2}{2}}} \right),$$

Here, μ_f and μ_b represent the mean values, while σ_f^2 and σ_b^2 denote the variances of the foreground and background regions, respectively. These metrics serve to quantitatively assess image quality, reflecting the effectiveness of the denoising process and the clarity achieved in separating foreground details from background noise.

3.2. Denoising OCT images

The task of denoising the image dataset was tackled from different approaches. Conventional denoising techniques such as wavelet transform, anisotropic diffusion, singular value thresholding, total variation, BM3D, and Gaussian transform were experimented on using the train dataset. Out of the above attempts, we received the best results from the BM3D approach. We selected this method as our working technique and fine-tuned our results by approximating the noise level. We used $\sigma = 0.083$ as the standard deviation of NPSD which is a measure of the noise level in the image.

After obtaining the BM3D-filtered outputs for the images, we managed to remove the remaining slight grain structures by truncating the singular value decomposition. We used SVHT to selectively truncate all singular values below a threshold of $\tau = 400$. Although this additional step of noise removal did not improve the CNR of the image, it gave better results in terms of visual quality. The functional pipeline was tested for performance under both the train dataset and test

dataset. The results of the metrics MSR and CNR for each dataset is shown in Table 1.

Table 1. Results of denoising for the train and test dataset. All values are in dB.

	Train		**Test**	
Method	MSR	CNR	MSR	CNR
Raw image	1.8525	1.8525	2.0011	2.5633
SVHT	1.8587	1.8529	2.0011	2.5633
BM3D	1.9214	2.0036	2.1358	2.7499
BM3D+SVHT	**1.9215**	**2.0036**	**2.1358**	**2.7499**

3.3. Super Scaling OCT images

Table 2. Results of different super-resolution methods. All values are in dB.

Method	MSR	CNR
Bicubic Interpolation (baseline)	1.6269	2.8461
Lanczos Resampling	1.6259	2.8438
ESR-GAN	1.5524	2.6573
EDSR	**1.6295**	**2.8502**

In Table 2, bicubic interpolation serves as the baseline method for comparison against state-of-the-art super-resolution models employing generative approaches such as ESR-GAN (*Enhanced Super Resolution Generative Adversarial Network*) [12] However, in the context of super-scaling OCT denoised images, ESR-GAN demonstrated suboptimal performance. Consequently, we employed a pretrained EDSR model [6], which integrates a modified ResNet backbone, achieving superior CNR and MSR values for upscaling images.

In Fig. 6, artifacts introduced by ESR-GAN are discernible, despite the apparent sharpness of edges. However, the resulting CNR and MSR metrics show a decline. On the other hand, EDSR yields a visually comparable outcome to conventional methods like bicubic interpolation and Lanczos Resampling. Nevertheless, the CNR and MSR metrics are notably superior in the output of the EDSR model.

3.4. Classification of OCT dataset

To address this problem, we employed two advanced machine learning models, including the fine-tuned ResNet18 CNN classification model using an ensemble method (our proposed method), where we chose the highest-voted disease from classifying each B-scan of the patient. Additionally, we experimented on a hybrid Transformer-based model for volumetric image classification that operates on CNN feature maps.

In the hybrid transformer, we reduced the B-scan size to 128x128 to be compatible with the DenseNet121[13] model. Next, we focused on the Transformer[14] model. The attention layers in a Transformer are inherently order-agnostic, meaning they do not consider the order of input data. Since our images are not ordered for each patient we haven't cared about the positional encodings of each image. Then we extracted the features of each B-Scan using DenseNet121 and then fed these features into the Transformer to obtain the classification results.

Both methods involved splitting the data into a 80:20 ratio for training and testing purposes. This ensured a robust evaluation of our models' performance. Each approach was rigorously tested to determine which yielded the highest evaluation metrics. Precision, recall, specificity, sensitivity, and the F1 score were the key metrics we used to gauge the effectiveness of our models. By doing so, we ensured that our evaluation was comprehensive, covering both the models' accuracy and ability to handle imbalanced datasets.

The results of classification task is shown in Table 3.

Table 3. Results metrics of OCT volumetric image classification. P-*precision*, R-*recall*, SP-*specifity*, Acc-*accuracy*

Method	P	R	SP	F1	Acc
Hybrid transformer	0.85	0.77	0.34	0.74	0.80
Resnet Ensemble	**0.95**	**0.94**	**0.31**	**0.95**	**0.95**

We observe that the Resnet Ensemble model outperforms the Hybrid transformer obtaining an accuracy of 95%. Thus, the model that achieved the highest evaluation metrics was ultimately selected as our best-performing model.

This approach improves accuracy by employing several strategies: utilizing a pre-trained ResNet18 backbone allows the model to start with low-level features that are already useful, rather than learning from scratch. Initially training only the fully connected layers ensures that the high-level features of OCT images are adequately captured before fine-tuning the more complex layers, improving overall model robustness. Finally, aggregating predictions from multiple 2D slices into a final decision reduces the impact of any single inaccurate prediction, enhancing the reliability and accuracy of the classification. To increase theconfidence of each prediction filtering the results using a prediction threshold level. Taking into account, only the predictions with high confidence 2D slices (B-scans), while neglecting the predictions with ambiguity, mitigates subject classification errors.

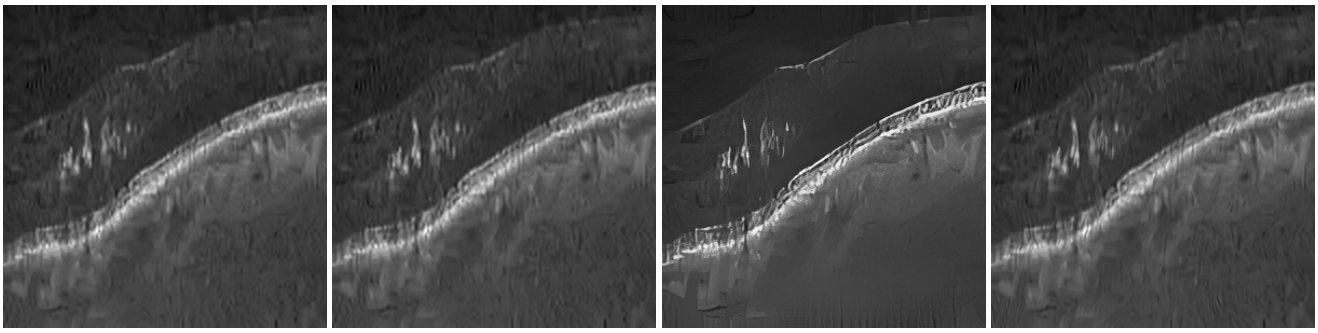

Fig. 6. Super-scaled sample image from left to right in order: bicubic interpolation, Lanczos Resampling for super-scaling, ESR-GAN Super Resolution, EDSR Super Resolution.

4. CONCLUSION

In this study, we presented a comprehensive approach to enhance and classify OCT B-scans using a combination of advanced denoising, super-resolution, and deep learning techniques. The integration of BM3D and SVHT significantly improved the quality of denoised images, effectively reducing speckle noise and enhancing the contrast-to-noise ratio (CNR). Our use of the EDSR model for super-resolution demonstrated superior performance in maintaining image fidelity and detail compared to traditional methods and state-of-the-art GAN-based approaches.

For classification, we implemented a ResNet18-based CNN, which through transfer learning, provided robust performance in distinguishing between healthy individuals, diabetic patients with DME, and non-diabetic patients with other ocular diseases. The ensemble method for classification further improved the accuracy, achieving a high F1 score and precision.

Our results indicate that the proposed pipeline not only enhances OCT image quality but also provides reliable automated classification, making it a valuable tool for real-time OCT image analysis and diagnosis. Future work will focus on further optimizing the deep learning models and exploring additional techniques to improve the robustness and accuracy of the classification system.

5. REFERENCES

[1] M. Tajmirriahi, Z. Amini, A. Hamidi, A. Zam, and H. Rabbani, "Modeling of retinal optical coherence tomography based on stochastic differential equations: Application to denoising," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2129–2141, 2021.

[2] M. Tajmirriahi, R. Rostamian, Z. Amini, A. Hamidi, A. Zam, and H. Rabbani, "Stochastic differential equations for automatic quality control of retinal optical coherence tomography images," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3870–3873, IEEE, 2022.

[3] Z. Amini and H. Rabbani, "Statistical modeling of retinal optical coherence tomography," *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1544–1554, 2016.

[4] Z. Amini and H. Rabbani, "Optical coherence tomography image denoising using gaussianization transform," *Journal of Biomedical Optics*, vol. 22, no. 8, pp. 086011–086011, 2017.

[5] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Curvature correction of retinal octs using graph-based geometry detection," *Physics in Medicine & Biology*, vol. 58, no. 9, p. 2925, 2013.

[6] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

[7] Y. Mäkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Transactions on Image Processing*, vol. 29, pp. 8339–8354, 2020.

[8] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[10] N. Otsu *et al.*, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[11] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "Reduction of speckle noise from optical coherence tomography images using multi-frame weighted nuclear norm minimization method," *Journal of Modern Optics*, vol. 62, no. 21, pp. 1856–1864, 2015.

[12] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.