

Enhancing retinal SELFF-OCT image quality: A deep learning-based pipeline

Marc S. Seibel^{a*}, Marc Rowedder^{a*}, Julia Andresen^a, Tobias Neumann^c, Richard Neffin^c, Helge Sudkamp^c, Heinz Handels^{a, b}, and Timo Kepp^b

^aInstitute of Medical Informatics, University of Lübeck, Germany

^bGerman Research Center for Artificial Intelligence (DFKI), Lübeck, Germany

^cVisotec GmbH, Lübeck, Germany

ABSTRACT

Advances in the development of optical coherence tomographs will make it possible to monitor the progression of eye diseases at home. To this end, Self-Examination Low-Cost Full-Field Optical Coherence Tomography (SELFF-OCT) was recently developed. SELFF-OCT devices are easy to operate and allow patients to take images of the retina themselves without having a doctor present. However, images produced by these devices are of lower quality compared to traditional OCT devices. In this work, we propose a deep-learning assisted pipeline which enhances the quality of SELFF-OCT images. The pipeline consists of four steps: 1. Quality assessment, 2. Image denoising, 3. Registration and fusion of multiple OCT-scans, 4. Averaging of multiple neighboring B-scans. Our preprocessing pipeline enhances the image quality in terms of signal-to-noise ratio (SNR), artifacts and distortions, measured with the blind/referenceless image spatial quality (BRISQUE) evaluator, and detectability of retinal layers, measured with Fisher's linear discriminant score. Starting from the recorded images, our method increases the SNR from 2.8 to 4.8, lowers the BRISQUE from 67 to 18 which indicates a reduced number of artifacts, and increases Fisher's discriminant from 2.1 to 5.7 which indicates a better detectability of retinal layers. These results indicate that the proposed pipeline will be useful for improving the detection of biomarkers in future studies utilizing SELFF-OCT.

Keywords: Optical coherence tomography, Retina, Image registration, Denoising, Deep learning

1. INTRODUCTION

Age-related macular disease and diabetic retinopathy are leading causes of blindness. Monitoring eye diseases at home provides the opportunity to continuously gather information about the state of these eye diseases and thereby opens the avenue for individualized treatment at the right time. With the advent of Self-Examination Low-Cost Full-Field Optical Coherence Tomography (SELFF-OCT)¹ a technology exists which allows patients to take images of their retina themselves. While the handling of the SELFF-OCT devices does not require a technical service person, the image quality lags behind clinical OCT imaging. To remove speckle noise, artifacts, and missing signal due to movement, we propose a preprocessing pipeline for image enhancement as depicted in Fig. 1. Finally, we verify that our pipeline improves the image quality both from a qualitative and quantitative point of view.

Corresponding author: Marc S. Seibel, marc.seibel@uni-luebeck.de

*Equal contribution

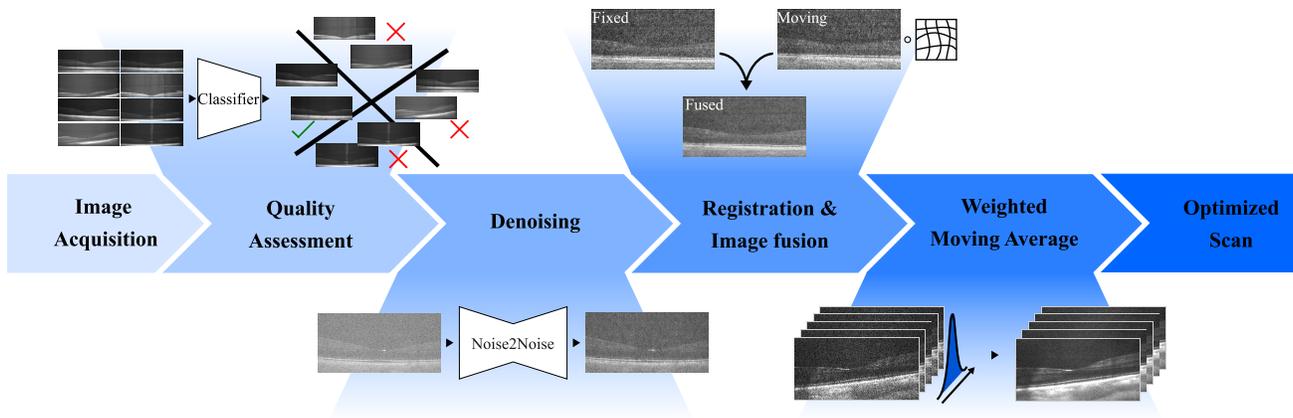


Figure 1. Our pipeline consists of four step. We start by assessing the image quality. Images with appropriate quality are included for further quality enhancement. Subsequently, images are being denoised. Following this, we register two OCT-scans and fuse them together. Finally, we apply a weighted moving average filter to enhance the contrast of retinal layers.

2. METHODS

2.1 Data

The OCT data used in this work were recorded with a Self-Examination Low-Cost Full-Field Optical Coherence Tomography (SELFF-OCT) device of the sixth generation produced by Visotec.¹ Full-field OCT devices record 2-dimensional planar images along the optical axis of the eye. These images are called en face scans. The coordinate vectors, which span en face scans, are called lateral dimensions. The optical axis, also called the axial dimension, is perpendicular to the two lateral dimensions. The en face scans are stacked along the axial dimension and a 3-dimensional volume is obtained. By slicing the image volume along one lateral dimension, we obtain B-scans. These B-scans can be seen in Figures 2 and 3.

We recorded OCT images from 50 subjects with healthy eyes. For each subject, we measured the left and right eye. Three sets of measurements per eye were performed. During one set of measurements, the SELFF-OCT device took a sequence of 15 volumetric scans within less than two minutes. Image variations across these 15 scans were smaller than the image variations in comparison to another set of measurements. For ten subjects, we labeled the retinal layers using the Iowa Reference algorithm.²⁻⁴

2.2 Pipeline

Our pipeline for image quality enhancement is depicted in Fig. 1 and consists of four steps: First, we remove images of bad quality and rank the remaining images; second, we apply a denoising algorithm on the images; third, we take the two best recordings of the same eye, register them to each other, and average them in order to reduce noise. To further reduce noise, we average adjacent B-scans with a one-dimensional Gaussian low-pass filter ($\sigma_x = 35 \mu\text{m}$).

2.2.1 Quality assessment

During the image acquisition, artifacts impact the images. These artifacts vary in severity. Images which contain too many artifacts or do not show the region of interest cannot be repaired through means of registration. We defined four classes of image qualities, which are depicted in Fig. 2. For an image, in order to be considered for further processing, three quality conditions need to be met: First, the retina needs to be visible, second a sufficient amount of the choroid needs to be present, and third only a few artifacts are allowed. If the second condition is met, then the first condition is also met. If an image suffices all three conditions, then we call it *clean*. If an image suffices the first and second but not the third condition, then we call it *corrupted*. If an image meets the first and third but not the second condition, then we categorize it as *missing choroid*. If an image meets only the third condition, then we classify it as *cropped retina*. These conditions are intentionally formulated in a vague way because we have no objective way to practically rate the images. With the intention

to train and evaluate a single layer convolutional neural network, we manually labelled 312 images according to the four classes. The neural network is described in Tab. 1. The algorithm was then used to predict the quality class of the remaining 4368 images.

After identifying all images of high quality, we chose the two best images from a given set of measurements. To this end, we used a handcrafted image quality score developed and tested by Visotec. The score characterizes the images based on three metrics. The first metric describes how many en face scans are empty due to movements of the head relative to the devices. The second metric describes vignetting (reduction of the brightness of the image towards the periphery). The last metric estimates the signal-to-noise ratio.

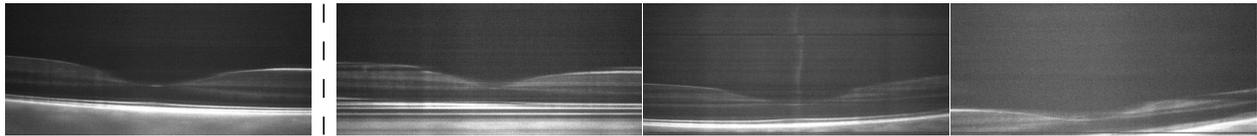


Figure 2. From left to right, images of the four quality classes *clean*, *corrupted*, *missing choroid*, and *cropped retina* are shown. Only images in the first class are considered for further processing.

Table 1. Convolutional neural network for classifying the image quality.

Layer type	Output shape	Comment
0. Input	$1 \times 200 \times 550$	B-scan as input
1. 1D Average pooling (1×50 , 1×25 stride)	$1 \times 200 \times 21$	Average lateral dim.
2. 1D Conv. (8 kernels, 100×1 , 2×1 stride, ReLU activation)	$8 \times 51 \times 21$	Conv. axial dim.
3. Dense (Softmax activation)	4	Quality classification

2.2.2 Denoising

To reduce the speckle noise, we denoised the OCT images using the Noise-to-noise (N2N) approach.^{5,6} The N2N network was trained using paired image data which we obtained from a phantom eye which was fixed in front of a SELFF-OCT device. Thereby, we acquired images where everything but the noise was constant. The full approach is described in.⁶

2.2.3 Registration and image fusion

In order to increase the image quality, we want to fuse multiple OCT recordings from the same set of measurements. While recording multiple OCT volumes, the position of the eye relative to the scanner changes such that we need to register these images to each other. We performed a two-stage registration scheme for coarse-to-fine alignment.^{7,8} In the first stage, we estimated the axial translation between the retina in two volumes using the deep learning framework for unsupervised affine image registration (AIRNet) of de Vos et al.⁷ For this task, we first averaged OCT-volumes along one lateral dimension. The obtained B-scan exhibits a strong contrast and Bruch’s membrane is clearly visible. The AIRNet learns to predict the correct axial translation by minimizing the mean squared error. For training, fixed and moving images were freely chosen from the whole dataset, i.e., the network learned to predict the axial alignment even for images which come from different eyes and subjects. After training, we computed the vertical alignment between the two images with the highest initial quality within one set of measurements.

In the second stage, we took the pre-aligned images and accurately matched them using the VoxelMorph approach. Given two 3D-images each of size $R^{H \times W \times D}$, VoxelMorph predicts a deformation matrix $\phi \in R^{H \times W \times D \times 3}$ which describes how far and in what direction each voxel in the moving image is shifted. For training the VoxelMorph model and for inferring the deformation matrices, we used images which were first denoised, vertically shifted, smoothed along one lateral dimension with a moving average with a window of size $40 \mu\text{m}$ and then resized by a factor of 0.32 along both lateral dimensions. This changes the lateral resolution from $2.5 \mu\text{m} \times 3.4 \mu\text{m}$ to $7.8 \mu\text{m} \times 10.6 \mu\text{m}$. The axial resolution of $5.6 \mu\text{m}$ was retained unchanged. Resizing was necessary due

to a memory constraint of 40 GB VRAM. We trained VoxelMorph in an unsupervised fashion by minimizing the mean squared error between a fixed image f and a moved image $m \circ \phi$. A smooth displacement field ϕ was encouraged by using a diffusion regularizer on the spatial gradients as described in.⁸ The fixed and moving images are the first and second-best image from each set of measurements.

3. RESULTS

The aim of the preprocessing pipeline is the improvement of image quality such that humans and computer programs can better extract semantic information from the images. To assess whether this aim has been achieved through our pipeline, we chose three metrics which are the signal-to-noise ratio (SNR), the blind/referenceless image spatial quality evaluator score (BRISQUE),⁹ and Fisher’s linear discriminant. Here, we define the SNR as the mean intensity of the voxels within the retina divided by the standard deviation of the voxels in the vitreous humor (upper part of the image). The BRISQUE score is based on handcrafted features, which characterize image distortions and artifacts. These features are passed to a Support Vector Regressor. We rely on the BRISQUE implementation described in.¹⁰ Using the segmentation masks, we calculate Fisher’s discriminant between seven adjacent retinal layers as $(\mu_i^2 - \mu_j^2)/(\sigma_i^2 + \sigma_j^2)$. The intuition behind Fisher’s discriminant is that distinguishable retinal layers differ in their average intensity (numerator) and exhibit a small amount of noise within each layer. The Fisher discriminant thus gives an indication of how well a clinician or an AI program could segment the images. For computing the metrics, all images were first normalized to a minimum value of 0 and a maximum value of 1.

In Fig. 3, we show the output of the pipeline at different stages of the pipeline. For the baseline image, individual retinal layers cannot be distinguished. After applying a logarithmic transform and the N2N network to the image, retinal layers slowly become visible, but at the same time, the image is getting blurred. Registration and fusion of two volumes further highlights individual retinal layers. Finally, by applying the weighted moving average on adjacent B-scans, we can increase the visibility of the retinal layers further. We observe that the quantitative results in Tab. 2 align with the qualitative impression of the images.

Table 2. Average SNR, BRISQUE and Fisher’s discriminant of the images. The best results are printed in bold. The last row of the table shows a comparison with Spectralis OCTs, which can be regarded as the clinical standard.

B-Scans	Preprocessing method	SNR \uparrow	BRISQUE \downarrow	Fisher’s discriminant \uparrow
Single	Baseline	2.8	67	2.1
	Image fusion	4.1	65	3.0
	N2N	7.4	58	2.8
	N2N & Image fusion	6.3	50	4.1
Average	Baseline	4.3	52	4.7
	Image fusion	5.4	40	5.2
	N2N	4.6	41	5.4
	N2N & Image fusion	4.8	18	5.7
-	Spectralis-OCT (clin. standard)	10.4	28.5	30.5

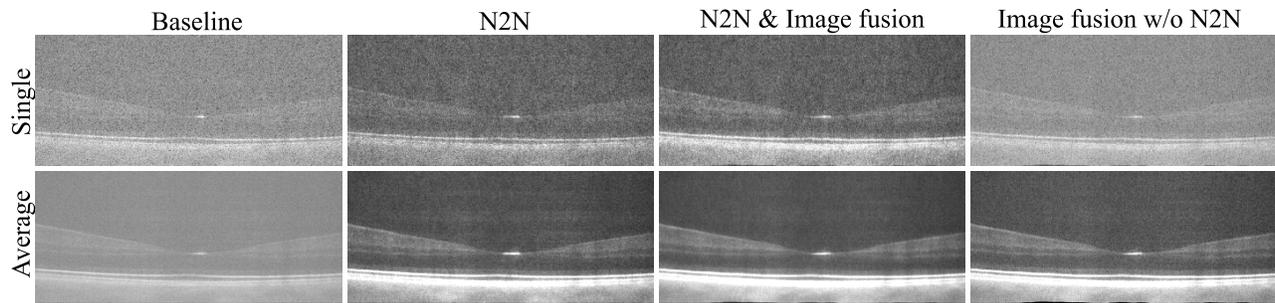


Figure 3. The first row shows a single B-scan at different stages of the preprocessing pipeline. In the second row, we averaged multiple adjacent B-scans from the same volume.

4. DISCUSSION

Our presented pipeline addresses the problem of low image quality in SELFF-OCT images. By removing noise using the N2N approach, we increased the image quality in terms of SNR, BRISQUE, and Fisher’s discriminant. Furthermore, by registering and fusing two OCT scans from the same set of measurements, we were able to improve BRISQUE and Fisher’s discriminant even further. Our best result in terms of BRISQUE and Fisher’s discriminant was achieved by averaging adjacent B-scans. This, however, comes at the cost of reducing high-frequency information in one lateral dimension. Future work will involve the fusion of more than two OCT scans. Regarding the validity of the evaluation of our pipeline, we need to note that all 50 subjects were used for training the VoxelMorph model. To demonstrate that the registration and fusion step generalizes to unseen images, further tests with a held-out data set are required.

5. CONCLUSION

Denoising and combining images from multiple SELFF-OCT recordings improves their quality on a qualitative and quantitative level. The developed pipeline brings us closer to the applicability of SELFF-OCT for monitoring patients with eye diseases at their home. The improved contrast of different retinal layers suggests that the presented pipeline will also improve the detection of biomarkers and retinal fluids in future works.

Acknowledgments This work was supported through funding by the federal state of Schleswig-Holstein.

REFERENCES

- [1] Sudkamp, H., Koch, P., Spahr, H., Hillmann, D., Franke, G., Müntz, M., Reinholz, F., Birngruber, R., and Hüttmann, G., “In-vivo retinal imaging with off-axis full-field time-domain optical coherence tomography,” *Optics Letters* **41**, 4987 (Nov. 2016).
- [2] Kang Li, Xiaodong Wu, Chen, D., and Sonka, M., “Optimal Surface Segmentation in Volumetric Images-A Graph-Theoretic Approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 119–134 (Jan. 2006).
- [3] Garvin, M., Abramoff, M., Xiaodong Wu, Russell, S., Burns, T., and Sonka, M., “Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images,” *IEEE Transactions on Medical Imaging* **28**, 1436–1447 (Sept. 2009).
- [4] Abramoff, M. D., Garvin, M. K., and Sonka, M., “Retinal Imaging and Image Analysis,” *IEEE Reviews in Biomedical Engineering* **3**, 169–208 (2010).
- [5] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T., “Noise2Noise: Learning Image Restoration without Clean Data,” in [*Proceedings of the 35th International Conference on Machine Learning*], 2965–2974, PMLR (July 2018). ISSN: 2640-3498.
- [6] Rowedder, M., Kepp, T., Neumann, T., Sudkamp, H., Hüttmann, G., and Handels, H., “Denoising of home OCT images using Noise2Noise trained on artificial eye data,” in [*Medical Imaging 2024: Image Processing*], **12926**, 583–589, SPIE (Apr. 2024).

- [7] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Isgum, I., “A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration,” *Medical Image Analysis* **52**, 128–143 (Feb. 2019). arXiv:1809.06130 [cs].
- [8] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V., “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Transactions on Medical Imaging* **38**, 1788–1800 (Aug. 2019). arXiv:1809.05231 [cs].
- [9] Mittal, A., Moorthy, A. K., and Bovik, A. C., “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Transactions on Image Processing* **21**, 4695–4708 (Dec. 2012). Conference Name: IEEE Transactions on Image Processing.
- [10] Kastyulin, S., Zakirov, J., Prokopenko, D., and Dylov, D. V., “PyTorch Image Quality: Metrics for Image Quality Assessment,” (Aug. 2022). arXiv:2208.14818 [cs, eess].