

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Semantic denoising autoencoders for retinal optical coherence tomography

Laves, Max-Heinrich, Ihler, Sontje, Kahrs, Lüder
Alexander, Ortmaier, Tobias

Max-Heinrich Laves, Sontje Ihler, Lüder Alexander Kahrs, Tobias Ortmaier,
"Semantic denoising autoencoders for retinal optical coherence tomography,"
Proc. SPIE 11078, Optical Coherence Imaging Techniques and Imaging in
Scattering Media III, 1107818 (19 July 2019); doi: 10.1117/12.2526936

SPIE.

Event: European Conferences on Biomedical Optics, 2019, Munich, Germany

Semantic denoising autoencoders for retinal optical coherence tomography

Max-Heinrich Laves, Sontje Ihler, Lüder Alexander Kahrs, and Tobias Ortmaier

Institute of Mechatronic Systems, Leibniz Universität Hannover,
Appelstr. 11A, 30167 Hannover, Germany

ABSTRACT

Noise in speckle-prone optical coherence tomography tends to obfuscate important details necessary for medical diagnosis. In this paper, a denoising approach that preserves disease characteristics on retinal optical coherence tomography images in ophthalmology is presented. We propose *semantic denoising autoencoders*, which combine a convolutional denoising autoencoder with a priorly trained ResNet image classifier as regularizer during training. This promotes the perceptibility of delicate details in the denoised images that are important for diagnosis and filters out only informationless background noise. With our approach, higher peak signal-to-noise ratios with PSNR = 31.0 dB and higher classification performance of $F_1 = 0.92$ can be achieved for denoised images compared to state-of-the-art denoising. It is shown that semantically regularized autoencoders are capable of denoising retinal OCT images without blurring details of diseases.

Keywords: Computer-aided diagnosis, Image enhancement, Image classification, Machine learning

1. PURPOSE

Optical coherence tomography (OCT) is the most common imaging technique for diagnosis in ophthalmology. However, due to image acquisition based on interference of coherent light, OCT suffers from speckle noise. This results in grainy images with low contrast and obscured features where the diagnosis of medical conditions requires trained expert observers. Denoising of OCT has been addressed in the literature already and can be separated into two categories.¹ The first one employs denoising during OCT acquisition by e.g. averaging multiple frames of the same object. This prolongs the acquisition process and is therefore not applicable for dynamic objects. The second category comprises post-processing methods as inverse image problems, which try to reconstruct a clean image $\hat{\mathbf{x}}$ from a noisy observation $\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{c}$. A common assumption of the noise model \mathbf{c} of the observation $\tilde{\mathbf{x}}$ in OCT imaging is additive white Gaussian noise with zero mean and standard deviation σ .^{1,2}

Given a noisy OCT observation $\tilde{\mathbf{x}}$, the denoising can be expressed as optimization problem of the form

$$\hat{\mathbf{x}} = \arg \min \left\{ \mathcal{L}(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) + \lambda \mathcal{R}(\hat{\mathbf{x}}) \right\}, \quad (1)$$

which tries to find a reconstruction $\hat{\mathbf{x}}$ that is close to $\tilde{\mathbf{x}}$ by means of some similarity measure \mathcal{L} , but has considerably less noise. The term \mathcal{R} , weighted by a trade-off factor λ , regularizes the optimization of (1) in order to impose the condition of $\hat{\mathbf{x}}$ having less noise than $\tilde{\mathbf{x}}$. The regularizer \mathcal{R} generally expresses a chosen prior on the denoised images, such as the total variation (TV),³ or first and higher order derivatives of the image. In recent years, autoencoders (AE) have been applied to denoising tasks, in which the regularization prior is learned from corrupted and uncorrupted data samples $\{\mathbf{x}, \tilde{\mathbf{x}}\}$.^{2,4} The performance of AEs for denoising is not only due to their ability to learn priors from data, but also due to the structure of the image generator itself.⁵ Denoising autoencoders (DAE) usually have a data bottleneck between the encoding part and the decoding part, which forces the encoding part to extract a meaningful low-dimensional latent representation from a corrupted input image $\tilde{\mathbf{x}}$. This is then fed into the decoding part and mapped back to a reconstructed image $\hat{\mathbf{x}}$ in input space. Although DAEs provide excellent performance in denoising, they suffer from smoothing out subtle details that are important for medical diagnosis.

Send correspondence to:

M.H. Laves: E-mail: laves@imes.uni-hannover.de, Telephone: +49 511 762 19617

Optical Coherence Imaging Techniques and Imaging in Scattering Media III, edited by
Maciej Wojtkowski, Stephen A. Boppart, Wang-Yuhl Oh, Proc. of SPIE-OSA Vol. 11078,
1107818 · © 2019 SPIE-OSA · CCC code: 1605-7422/19/\$21 · doi: 10.1117/12.2526936

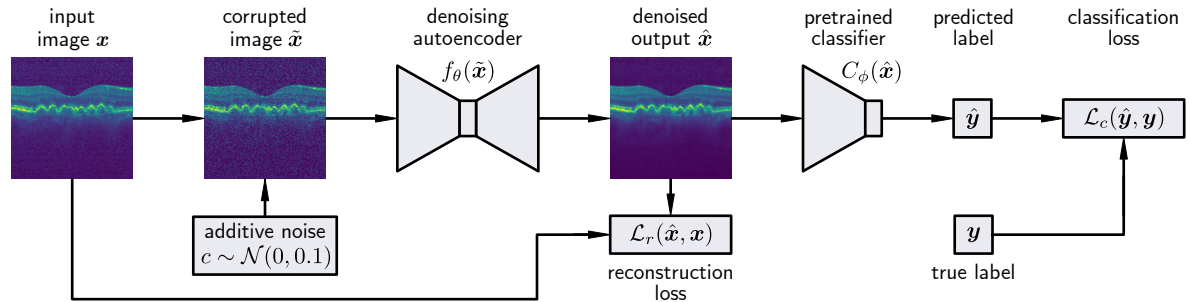


Figure 1. Overview of sDAE training procedure. The autoencoder f_θ tries to reconstruct a clean image \hat{x} from a noisy observation \tilde{x} while being regulated by a pretrained classifier C_ϕ .

AEs have recently been used to regularize the training of diagnostic classifiers in medical imaging.^{6,7} However, the opposing approach where a diagnostic classifier regularizes the process of DAE has not been addressed so far. Therefore, this paper describes a domain-specific post-processing method for denoising medical images with preservation of delicate disease characteristics by proposing *semantic denoising autoencoders* (sDAE).

2. METHODS

In this section, the sDAE approach is presented in detail. First, a ResNet-34 image classifier⁸ C_ϕ pretrained on ImageNet is fine-tuned on a dataset of OCT images described below. This acts as medical expert as it has been shown that the performance of convolutional neural networks (CNNs) in classifying retinal conditions is on par to that of trained ophthalmologists.⁹ Second, the ErfNet CNN autoencoder¹⁰ f_θ is trained to reconstruct input images x corrupted by additive gaussian white noise resulting in $\tilde{x} = x + c$ with $c \sim \mathcal{N}(\mathbf{0}, 0.1 \mathbf{I})$. The parameters θ of the AE are optimized by minimizing the pixel-wise mean squared reconstruction error $\mathcal{L}_r(f_\theta(\tilde{x}), x)$. Essentially, an autoencoder learns a low-dimensional representation similar to principal component analysis (PCA). When training with a large dataset, noise tends to “average out” and the AE reconstructs distinct and relevant (noise-free) image features. In order to promote enhancement of these features, the trained ResNet with fixed weights ϕ is used as additional optimization criterion \mathcal{L}_c . It is applied to the reconstructed, denoised image and tries to predict the retinal disease class (see Fig. 1). This regularizes the AE during training and enhances disease characteristics in denoised images. The proposed approach is therefore optimized using the weighted loss function

$$\arg \min_{\theta} \left\{ \mathcal{L}_r(f_\theta(\tilde{x}), x) + \lambda \mathcal{L}_c(C_\phi(f_\theta(\tilde{x})), y) \right\} \quad (2)$$

with true disease label y of image x and cross entropy loss for \mathcal{L}_c . The denoised reconstruction $\hat{x} = f_\theta(\tilde{x})$ can be obtained after convergence of the training. Regularization factor for \mathcal{L}_c was empirically set to $\lambda = 0.01$.

The dataset used to train the sDAE consists of 84,484 retinal OCT images from 4,657 patients showing the disease states drusen, diabetic macular edema (DME), choroidal neovascularization (CNV) and normal and is publicly available.⁹ We split the dataset using 4,000 OCT scans (1,000 from each class) for validation during training and another 4,000 scans for reporting final results (test set). The images for validation and testing were extracted patient by patient in order to prevent that data from one patient is included in more than one of the partial datasets.

The aforementioned method is implemented with PyTorch 1.0 and trained for 100 epochs using the Adam optimizer with an initial learning rate of $\eta = 10^{-4}$.¹¹ A reduce-on-plateau learning rate scheduling is realized to reduce η with a factor of 10^{-1} when observing saturation of the training loss. The weight configuration with lowest loss value on the validation set is chosen for testing (early stopping).

3. RESULTS

To assess denoising performance, the proposed method is compared to total variation (TV) minimization,³ BayesShrink wavelet denoising,¹² anisotropic diffusion (AD) denoising¹³ and an unregularized DAE regarding

	uncorrupted	corrupted	TV	wavelet	AD	DAE	sDAE (ours)
PSNR	inf	20.9 ± 0.24	29.3 ± 1.2	28.0 ± 1.0	28.2 ± 1.3	31.4 ± 1.78	31.1 ± 1.65
SSIM	1.0	0.44 ± 0.03	0.85 ± 0.05	0.81 ± 0.05	0.83 ± 0.05	0.89 ± 0.04	0.89 ± 0.04
F_1	0.94	0.89	0.79	0.83	0.55	0.86	0.92

Table 1. Mean results of denoising reported for the test set with mean peak signal-to-noise ratio (PSNR) in dB, structural similarity index (SSIM) and mean classification F_1 scores. Values for uncorrupted \mathbf{x} and corrupted images $\tilde{\mathbf{x}}$ are given for comparison. Bold values denote best results.

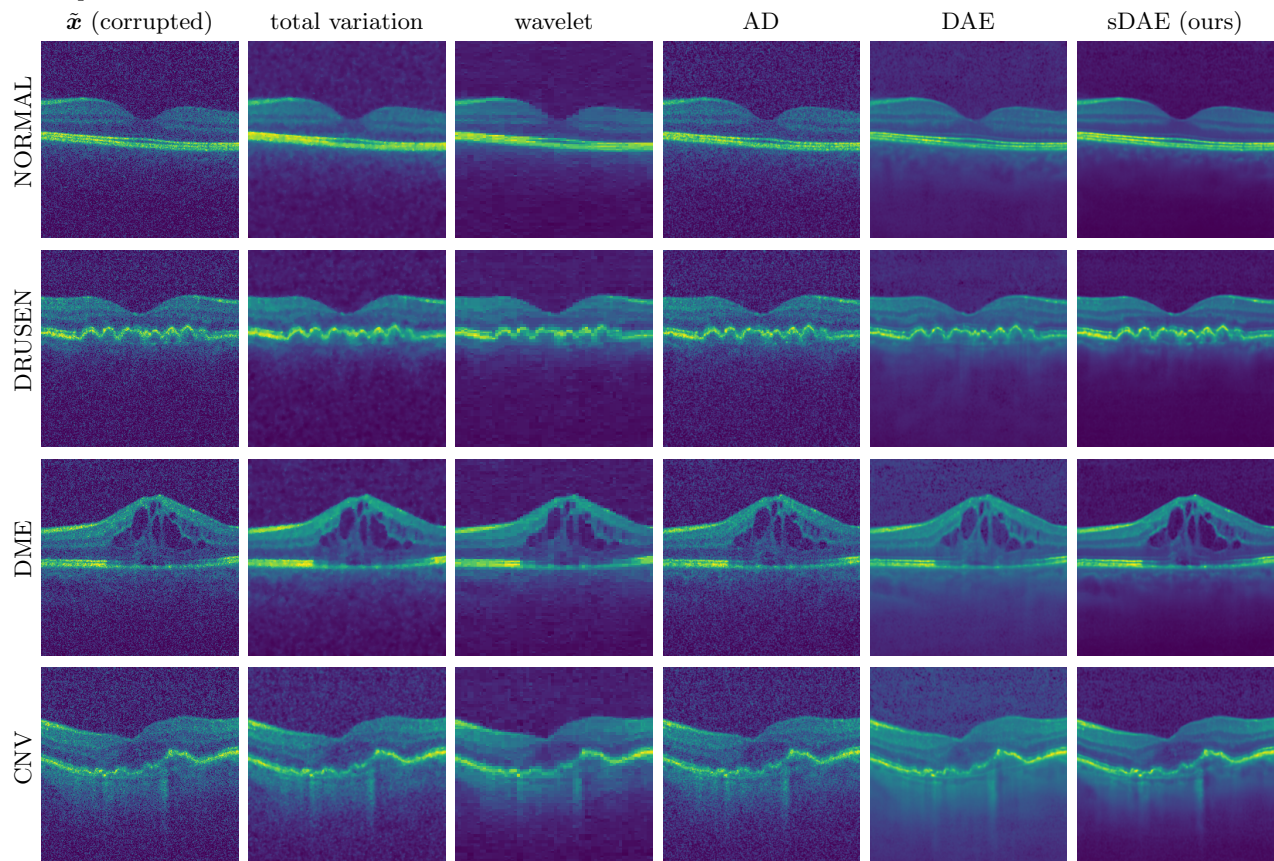


Figure 2. Results of our approach compared to state-of-the-art denoising for retinal OCT disease conditions from the test set. Digital zoom is recommended for optimal comparison.

peak signal-to-noise ratio (PSNR), structural similarity index¹⁴ and classification performance of ResNet using the F_1 score. The DAE can be seen as a special case of our approach, where $\lambda = 0$, such that it is only trained for reconstruction. The results are summarized in Tab. 1. We additionally provide the results of uncorrupted \mathbf{x} and corrupted images $\tilde{\mathbf{x}}$ as baseline. Our approach not only provides the highest disease classification accuracy with $F_1 = 0.92$ after denoising, but also has a peak signal-to-noise ratio with PSNR = 31.1 dB, which is only exceeded by the DAE. However, the SSIM suggests that sDAE and DAE have similar reconstruction performance.

Fig. 2 visualizes qualitative results for example OCT scans from the test set showing different disease conditions. The methods are used to restore the input image \mathbf{x} from the corrupted image $\tilde{\mathbf{x}}$ (first column). In contrast to state-of-the-art denoising, our approach is able to distinctively preserve the retinal layers while removing speckle noise. Pathological alterations of the retina are clearly visible and the explanatory power for diagnosis is not reduced. Mean processing time of sDAE for one image is 13.1 ms on an NVIDIA GeForce GTX 1080 Ti.

4. CONCLUSION

It has been shown that the proposed semantic denoising autoencoder is capable of denoising retinal OCT images without suppressing characteristics of diseases. This was achieved by regularizing the denoising autoencoder during training with another CNN, which was previously trained for disease classification. The denoising performance of sDAE is similar to that of an unregularized autoencoder, but sDAE preserves details important for diagnosis. The trained decoder can also be used to generate new images by sampling the latent space. Future work therefore aims on variational autoencoder and generative adversarial networks for OCT denoising. It should be noted, however, that speckle noise can also contain significant information as it creates a unique fingerprint of tissue. This information is hard to be interpreted by humans, and CNNs can be valuable tools to acquire and utilize this information in the future. The presented approach can also be translated to other inverse image problems such as single image super-resolution or compression artifacts removal or other medical imaging modalities such as computed tomography or magnetic resonance imaging.

Conflict of Interest The authors declare that they have no conflict of interest.

Funding This research has received funding from the European Union as being part of the EFRE OPhonLas project.

Formal Consent The medical images used in this article were made available to the public in a previous study,⁹ therefore formal consent is not required.

REFERENCES

- [1] Salinas, H. M. and Fernandez, D. C., “Comparison of PDE-Based Nonlinear Diffusion Approaches for Image Enhancement and Denoising in Optical Coherence Tomography,” *IEEE Transactions on Medical Imaging* **26**(6), 761–771 (2007).
- [2] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L., “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing* **26**(7), 3142–3155 (2017).
- [3] Chambolle, A., “An Algorithm for Total Variation Minimization and Applications,” *Journal of Mathematical Imaging and Vision* **20**(1–2), 89–97 (2004).
- [4] Bengio, Y., Yao, L., Alain, G., and Vincent, P., “Generalized denoising auto-encoders as generative models,” in [*Advances in Neural Information Processing Systems (NIPS) 26*], 899–907 (2013).
- [5] Lempitsky, V., Vedaldi, A., and Ulyanov, D., “Deep Image Prior,” in [*IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 9446–9454 (June 2018).
- [6] Creswell, A., Pouplin, A., and Bharath, A. A., “Denoising adversarial autoencoders: classifying skin lesions using limited labelled training data,” *IET Computer Vision* **12**(8), 1105–1111 (2018).
- [7] Laves, M.-H., Ihler, S., Kahrs, L. A., and Ortmaier, T., “Retinal OCT disease classification with variational autoencoder regularization,” *arXiv preprint arXiv:1904.00790* (2019).
- [8] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 770–778 (2016).
- [9] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., et al., “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell* **172**(5), 1122–1131 (2018).
- [10] Romera, E., Álvarez, J. M., Bergasa, L. M., and Arroyo, R., “ERFNet: Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation,” *IEEE Trans. Intell. Transp. Syst.* **19**(1), 263–272 (2018).
- [11] Kingma, D. P. and Ba, J., “Adam: A method for stochastic optimization,” *arXiv e-prints* (2014).
- [12] Chang, S. G., Yu, B., and Vetterli, M., “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing* **9**(9), 1532–1546 (2000).
- [13] Perona, P. and Malik, J., “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(7), 629–639 (1990).
- [14] Zhou Wang, Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).