Energy-Regularized Self-Supervised Deep Network for Speckle Noise Reduction in Optical Coherence Tomography

S. KETHAVATH[1], VASUNDHARA[1,*], M. K. MANEPALLI[2], B. R. CHINTADA[3], AND P. K. YALAVARTHY[2,4]

[1] Department of Electronics and Communication Engineering, National Institute of Technology Warangal, 506004, India
[2] TANUH: AI-Centre of Excellence in Healthcare, Indian Institute of Science, Bengaluru, 560012, India
[3] Department of Bioengineering, Indian Institute of Science, Bengaluru, 560012, India
[4] Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, 560012, India
[*] vasundhara@nitw.ac.in

Compiled July 20, 2026

Optical Coherence Tomography (OCT) speckle degrades the contrast and impairs automated analysis. Classical speckle suppression methods blur boundaries, supervised deep learning requires clinically impractical training pairs, and existing self-supervised approaches, such as S2SNet, tend to over-smooth retinal layers. To address these limitations, this study proposes a novel energy-based regularization term that penalizes overly smooth, low-energy reconstructions, yielding an Energy-regularized Self-Supervised Network (E-S2SNet). Evaluation of the 18 B-scan SD-OCT dataset and the NR206 dataset demonstrated that E-S2SNet achieves a superior balance between speckle suppression and structural preservation, outperforming both classical and deep learning baselines without requiring clean reference (ground truth/training) data. Furthermore, improved downstream retinal layer segmentation indicates the practical utility of the proposed approach, showing promise for future integration into automated OCT analysis.

http://dx.doi.org/XX.XXXX/ol.XX.XXXXXX

Optical coherence tomography (OCT) is an optical cross-sectional imaging technique that generates high-resolution images of biological tissue and is widely used as a clinical diagnostic tool in ophthalmology [1]. OCT images are inherently corrupted by speckle from coherent interference of backscattered light [2]. This reduces contrast, obscures subtle structural changes, and impairs automated pipelines that use local intensity gradients and texture cues to segment retinal layers. Averaging repeated B-scans can suppress speckle, but it requires motion compensation to avoid blurring and prolonged acquisition is challenging in patients with poor fixation [3]. Classical single-B-scan despeckling methods, such as anisotropic diffusion, non-local means (NLM) [4], and block-matching and 3D filtering (BM3D) [5], often blur fine boundaries and lose thin biomarker-relevant layers. Recently, supervised deep learning methods, such as volumetric speckle suppression using cGAN [6], showed strong performance. However, these methods require speckle-suppressed ground truth for training, obtained through lengthy repeated B-scan averaging or computationally expensive algorithmic suppression, which poses practical challenges for routine clinical deployment.

Self-supervised denoising methods learn from noisy inputs alone without any clean reference [7, 8]. Early blind-spot networks, such as Noise2Void, Noise2Self, and Neighbor2Neighbor, pioneered this paradigm by predicting a target pixel's intensity from its receptive field. Subsequent methods, such as Blind2Unblind, improved these mask-based strategies. For optical imaging, S2SNet [8] adapts the Self2Self framework [9] for OCT speckle reduction, training on Bernoulli-sampled dropout instances of a single noisy image and averaging predictions across stochastic forward passes. Although effective, S2SNet lacks explicit anatomical priors and remains susceptible to blurring fine structural boundaries, potentially disrupting retinal layer continuity.

This energy regularization addresses OCT's physical characteristics better than standard constraints. Edge-preserving losses can induce staircasing artifacts and blur fine retinal layers, while pre-trained perceptual losses miss unique backscattering morphologies and require clean reference targets. Because OCT speckle is high-frequency, standard high-frequency penalties struggle to separate noise from structural details. The proposed E-S2SNet differs from standard self-supervised OCT speckle reduction methods by preventing statistical over smoothing. Standard self-supervised models optimize pixel-wise averages, treating thin retinal layer boundaries as noise and producing smooth outputs. The proposed method adds a localized structural energy regularization term that penalizes flat reconstructions, balancing speckle suppression with sharp edge preservation, maintaining boundary contrast without external priors and improving retinal segmentation.

To address these limitations, this study utilized the S2SNet [8] framework and incorporated a novel energy-based regularization term that penalized low-energy, overly smooth reconstructions, thereby deriving the Energy-regularized Self-Supervised Network *(E-S2SNet)*. This approach promoted the preservation of high-frequency anatomical details and sharp structural transitions. The derived E-S2SNet framework was evaluated on two datasets: (i) an 18 B-scan macular OCT benchmark dataset [10], and (ii) the NR206 dataset, comprising 206 healthy retinal B-scans with nine annotated retinal layers [11].

This study adopted S2SNet [8] training and inference strategies, with its self-prediction and background noise attenuation (BNA) loss functions. S2SNet [8] adapts Self2Self for OCT

speckle reduction by training an encoder-decoder network with gated convolutions on Bernoulli-sampled pairs from a noisy image, without ground truth. Given a noisy image $Y \in \mathbb{R}^{H\times W}$, a binary Bernoulli mask $b_m \in \{0,1\}^{H\times W}$ is sampled at each training iteration, where each pixel (i,j) is independently retained $(b_m(i,j)=1)$ with probability p or dropped $(b_m(i,j)=0)$ with a probability of $1-p$. Importantly, this masking probability p remains spatially uniform and fixed throughout training; the mask generation is strictly stochastic and is not dynamically guided by local image features. Both the original noisy image and its Bernoulli-sampled counterpart are fed simultaneously into the network with shared weights, yielding two complementary outputs.

$$\bar{Y}_m = b_m \odot Y, \qquad \widetilde{Y}_m = (1-b_m)\odot Y \tag{1}$$

where $\odot$ denotes element-wise multiplication and M denotes the total number of random mask samples generated per training iteration. Let $\mathcal{F}_\theta$ denote the U-Net-based denoising network with parameters θ. Two predictions were computed: the masked prediction $\hat{X}_m = \mathcal{F}_\theta(\bar{Y}_m)$ and the full prediction $\hat{X} = \mathcal{F}_\theta(Y)$. The self-prediction loss encourages consistency between these two outputs, exploiting the statistical consistency of the Bernoulli-sampled instances of the same underlying signal:

$$\mathcal{L}_{\text{self-prediction}} = \min \sum_{m=1}^{M} \left\| (1-b_m)\odot(\hat{X}_m - \widetilde{Y}_m) \right\|_2^2 \tag{2}$$

The BNA loss additionally suppresses the background noise:

$$\mathcal{L}_{\text{bna}} = \min \frac{1}{M} \sum_{m=1}^{M} \left\| \hat{X} - \hat{X}_m \right\|_1 \tag{3}$$

During inference, dropout is retained, and the final speckle-suppressed result is obtained by averaging the predictions across multiple stochastic forward passes, thereby reducing the prediction variance. While $\mathcal{L}_{\text{self-prediction}} + \mathcal{L}_{\text{bna}}$ suppressed speckle, it often led to over-smoothed images where fine anatomical details were lost.

To prevent this, this study proposes E-S2SNet. Unlike traditional dynamic alternation of the input masking mechanism, E-S2SNet achieves structural preservation by introducing an energy-guided regularization term during the optimization phase. This term acts as a penalty for weak or uncertain feature responses. Let $\Phi(\bar{Y}_m)$ denote the intermediate feature maps extracted from the penultimate layer of the decoder. The proposed framework quantifies the structural confidence at each spatial pixel location (i,j) by aggregating feature activations across all channel dimensions into a scalar energy map:

$$E_{i,j}(\bar{Y}_m) = \log \sum_{c=1}^{C} \exp\left(\Phi_{c,i,j}(\bar{Y}_m)\right) \tag{4}$$

where c is the channel index, C represents the total number of feature channels (set to 96 in the utilized architecture), and $\Phi_{c,i,j}(\bar{Y}_m)$ is the network activation at channel c and spatial coordinates (i,j). To encourage sharp transitions, we enforced high energy magnitudes by minimizing the reciprocal of the absolute energy across all pixels and mask realizations:

$$\mathcal{L}_{\text{energy}} = \min \frac{1}{M} \sum_{m=1}^{M} \frac{1}{H\cdot W} \sum_{i,j} \frac{1}{\left| E_{i,j}(\bar{Y}_m) \right| + \epsilon} \tag{5}$$

where H and W represent the spatial height and width of the feature maps, respectively, and $\epsilon = 10^{-8}$ is a small constant that ensures numerical stability. By minimizing this reciprocal, the loss increases drastically if the feature energy approaches zero, effectively preventing convergence to over-smoothed solutions.

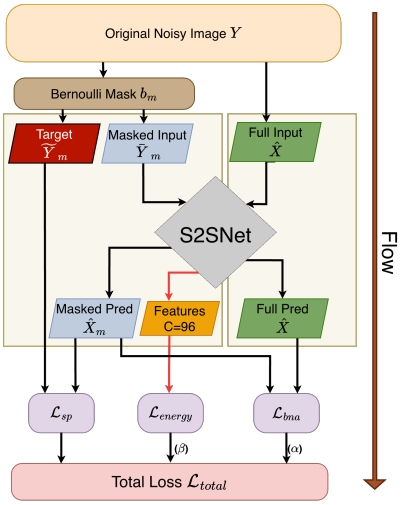

Figure 1. Training strategy flow diagram for the proposed E-S2SNet framework.

The final self-supervised training objective for the E-S2SNet was a weighted combination of these terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{self-prediction}} + \alpha\,\mathcal{L}_{\text{bna}} + \beta\,\mathcal{L}_{\text{energy}} \tag{6}$$

where α and β are the weighting coefficients of the bna loss and energy-based regularization loss, respectively.

The training strategy using $\mathcal{L}_{\text{energy}}$ is shown in Figure 1. The E-S2SNet framework used a U-Net-style encoder-decoder with skip connections, GatedConv2D layers, and LeakyReLU activations. Before unsupervised denoising, input B-scans were resized to 256×256 pixels, cast to floating-point arrays, and z-score normalized. Experiments were run on an Ubuntu 22.04 LTS workstation with dual Intel Xeon Silver 4110 CPUs (@ 2.10GHz), dual NVIDIA TITAN RTX 24GB VRAM GPUs, 256GB RAM, and a 1 TB SSD. The model was trained for 2000 epochs with batch size 1 using Adam ($\beta_1 = 0.9, \beta_2 = 0.999$, weight decay=0.0) at learning rate 10^{-3}. For the main results, the optimal hyperparameter configuration established by the ablation studies was utilized: an internal network dropout rate of 0.3 for model regularization, and applying a 0.5 Bernoulli masking probability to generate the paired inputs for self-supervised training. Additionally, the network employed a background noise attenuation (BNA) weight of $\alpha = 0.3$, an energy loss weight of $\beta = 0.10$, and $N = 400$ stochastic forward passes for inference.

To validate the architectural design and hyperparameter configurations of the proposed E-S2SNet, ablation studies systematically evaluated the impact of the energy loss weighting factor (β), number of stochastic forward passes (N), dropout rate, Bernoulli masking probability, and decoder layer used for regularization injection. The network achieves its peak noise suppression and structural fidelity with $\beta = 0.10$, dropout rate 0.3, masking probability 0.5, while $N = 400$ stochastic passes provides the optimal inference saturation point.

During inference, stochastic ensemble averaging was employed, as in S2SNet [8]. With dropout layers active, N independent forward passes were performed for each input image, producing different predictions from dropout patterns. Each prediction contained dropout-induced stochastic variation and residual speckle, and averaging N passes reduced variance, suppressing fluctuations while preserving deterministic anatomical structural details consistent across passes. We observed that optimal N varied with inherent noise level and structural complexity: $N = 200$ passes achieved convergence for the NR206

Table 1. Quantitative analysis of the SD-OCT Dataset. The TP, EP, MSR, CNR, PSNR, SSIM, and HFEN-L1 were computed for the selected patches. Metrics marked with an asterisk (*) do not require a reference image. The best results are in bold.

Method	MSR* (↑)	CNR* (↑)	PSNR (dB) (↑)	SSIM (↑)	TP (↑)	EP (↑)	HFEN-L1 (↓)
NLM	20.94	10.86	**27.58**	0.6120	0.75	0.22	0.0724
BM3D	11.11	7.45	22.80	0.4805	0.66	0.25	0.1302
Noise2Void	8.15	1.78	20.93	0.5046	0.35	0.54	0.4825
Noise2Self	8.07	2.40	19.78	0.5778	0.36	0.22	0.1702
Neighbor2Neighbor	8.26	2.71	21.03	0.6154	0.38	0.34	0.1008
Blind2Unblind	8.38	3.17	21.42	0.6559	0.32	0.12	0.2274
Self2Self	19.05	7.96	23.41	0.6501	0.43	0.39	0.1779
S2SNet	20.58	13.62	24.14	0.6211	0.67	0.50	0.0701
E-S2SNet	**23.01**	**15.70**	24.84	**0.6657**	**0.84**	**0.58**	**0.0678**

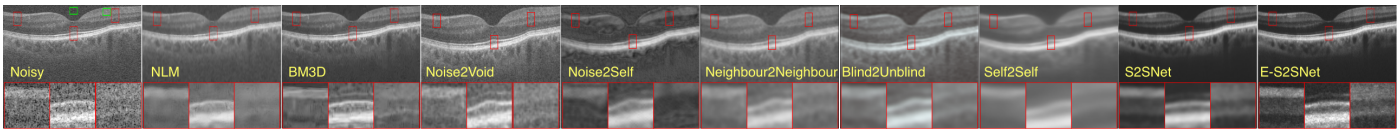

Figure 2. Qualitative comparison on an SD-OCT B-scan. Green boxes indicate the regions of interest used to calculate the CNR, and Red boxes indicate the ROI used to calculate other metrics, with the corresponding Quantitative results detailed in Table 1.

Table 2. Quantitative analysis of the NR206 Dataset. The MSR, CNR, and NIQM were computed for the selected patches. Metrics marked with an asterisk (*) do not require a reference image. The best results are in bold.

Method	MSR* (↑)	CNR* (↑)	NIQM* (↑)
NLM	15.732 ± 2.041	15.843 ± 2.221	0.2983
BM3D	4.264 ± 0.349	5.048 ± 0.342	0.2696
S2SNet	15.734 ± 3.540	15.329 ± 3.655	0.5853
E-S2SNet	$\mathbf{17.850 \pm 1.545}$	$\mathbf{17.105 \pm 1.837}$	**0.7707**

dataset, whereas N =400-500 passes were required for the SD-OCT dataset.

To quantitatively evaluate speckle suppression performance, we computed image metrics used in OCT speckle suppression literature. The Mean Signal-to-Noise Ratio (MSR) assessed smoothness as the mean-to-standard-deviation ratio within foreground ROIs. The Contrast-to-Noise Ratio (CNR) measured distinguishability between foreground and background ROIs from their means and standard deviations. The Peak Signal-to-Noise Ratio (PSNR, in dB) measured reconstruction fidelity relative to the reference image, and the Structural Similarity Index Measure (SSIM) assessed similarity in luminance, contrast, and structure. The Texture Preservation (TP) index quantified intensity variation preservation by comparing variance statistics of speckle-suppressed and reference ROIs, penalizing flattening. The Edge Preservation (EP) index measured longitudinal gradient ratios. The High-Frequency Error Norm (HFEN-L1) quantified the L_1-norm difference between Laplacian of Gaussian (LoG)-filtered speckle-suppressed and reference images. Furthermore, the no-reference image quality metric for image denoising introduced by Kong *et al.* [12] (NIQM score) was used to evaluate noise reduction and structure preservation. It computes the linear correlation coefficient between two structural similarity maps: one comparing the noisy input to the estimated method noise to assess noise reduction in homogeneous regions, and another comparing the noisy input to the denoised output to assess structure preservation. PSNR, SSIM, TP, EP, and HFEN-L1 require a speckle-suppressed reference image; since averaged B-scans are available as ground truth for the SD-OCT dataset, all metrics were computed for it. For the NR206 dataset, only the no-reference metrics MSR, CNR, and NIQM were computed. Table-2 reports figures of merit, and E-S2SNet achieved highest scores across metrics, outperforming classical and self-supervised approaches.

To systematically evaluate the downstream segmentation utility of our speckle suppression framework, three Fully Convolutional Networks (FCNs) [13] with ResNet-50 encoder backbones were independently trained on original noisy inputs, S2SNet-speckle-suppressed images, and E-S2SNet-speckle-suppressed images to segment retinal layers. The FCN was chosen for its foundational role in dense prediction and compatibility with pre-trained classification backbones; using ResNet-50 [13] lets the model leverage ImageNet-pretrained representations, suiting limited medical imaging data. Since this experiment measures denoising's relative effect on segmentation performance rather than absolute state-of-the-art results, using the architecture across all three training conditions ensures a controlled, fair comparison. The NR206 dataset, which consists of a single, fovea-centered B-scan from each subject's volumetric scan, was split at the patient level into 60%, 20%, and 20% for training, validation and testing, respectively. The networks were trained using conventional Dice loss and the Adam optimizer with a learning rate of 10^{-4} for up to 300 epochs and a batch size of 16, with early stopping to ensure optimal convergence. The performance was assessed using a holdout test dataset with the Dice score and Hausdorff distance. The Dice score measures the overlap between the ground-truth mask and network output mask, whereas the Hausdorff distance measures the maximum dissimilarity between the ground-truth segmentation mask and network output mask. Table 1 summarizes the quantitative denoising performance on the SD-OCT dataset, comparing the proposed E-S2SNet against a comprehensive suite of baseline methods. These baselines encompass classical algorithmic filters (NLM and BM3D) as well as state-of-the-art self-supervised deep learning frameworks (Noise2Void, Noise2Self, Neighbor2Neighbor, Blind2Unblind, Self2Self, and S2SNet). The quantitative results indicate that while classical approaches such as NLM achieve a mathematically high PSNR (27.58 dB), this often comes at the cost of severe over-smoothing and loss of subtle clinical features, as reflected by its poor edge preservation score (EP = 0.22). Conversely, early self-supervised spatial blind-spot models, including Noise2Void, Noise2Self, and Neighbor2Neighbor, struggle to effectively reconstruct the complex, signal-dependent speckle statistics inherent to OCT imaging, resulting in lower structural similarity (SSIM) and texture preservation (TP) scores. Blind2Unblind offers a marginal improvement in structural fidelity over early blind-spot methods but still falls short of optimal visual quality.

Table 3. Segmentation performance (Dice scores and Hausdorff distances) on the NR206 dataset. The best results are highlighted in bold. All three FCN models were trained using identical initialization, data augmentation, optimization, and early stopping strategies to ensure a fair comparison. NFL: Nerve Fiber Layer; GCL–IPL: Ganglion Cell Layer–Inner Plexiform Layer; INL: Inner Nuclear Layer; OPL: Outer Plexiform Layer; ONL: Outer Nuclear Layer; IS: Inner Segment; OS: Outer Segment; RPE: Retinal Pigment Epithelium.

Method	Background	NFL	GCL–IPL	INL	OPL	ONL	IS	OS	RPE
Dice Score (%) ↑									
Noisy Input	98.77 ± 0.35	**75.73 ± 3.82**	89.98 ± 2.45	80.99 ± 3.12	46.03 ± 7.94	89.66 ± 2.85	81.47 ± 3.92	31.55 ± 9.15	91.15 ± 2.18
S2SNet	99.36 ± 0.21	54.57 ± 5.12	90.01 ± 1.95	82.19 ± 2.45	58.57 ± 5.84	93.68 ± 1.62	88.35 ± 2.15	79.72 ± 3.84	94.89 ± 1.05
E-S2SNet	**99.57 ± 0.12**	59.76 ± 4.55	**91.12 ± 1.34**	**84.49 ± 1.65**	**75.65 ± 3.25**	**94.66 ± 1.13**	**90.99 ± 1.28**	**85.48 ± 2.45**	**95.95 ± 0.76**
Hausdorff Distance ↓									
Noisy Input	5.45 ± 2.15	**2.46 ± 1.24**	21.22 ± 7.84	14.28 ± 5.45	17.31 ± 6.12	7.18 ± 3.45	6.42 ± 2.84	43.25 ± 11.20	35.94 ± 9.45
S2SNet	2.52 ± 0.95	6.83 ± 2.54	12.09 ± 4.23	8.69 ± 3.12	5.26 ± 2.14	3.26 ± 1.15	2.68 ± 0.94	8.90 ± 3.15	7.42 ± 2.65
E-S2SNet	**2.08 ± 0.65**	2.64 ± 1.18	**8.17 ± 2.85**	**6.32 ± 1.95**	**3.20 ± 0.98**	**2.70 ± 0.82**	**1.98 ± 0.54**	**5.49 ± 1.45**	**4.23 ± 1.42**

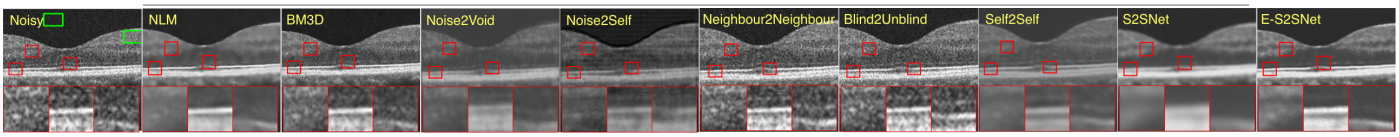

Figure 3. Qualitative comparison of speckle suppression on a representative image from the NR206 dataset. Green boxes indicate the regions of interest used to calculate the CNR, and Red boxes indicate the ROI used to calculate other metrics.

More advanced masking-based ensemble methods, namely Self2Self and its OCT-adapted variant S2SNet, demonstrate a significant leap in performance by leveraging stochastic dropout predictions to better separate deterministic structure from random noise. Although E-S2SNet does not achieve the highest PSNR among all evaluated methods, its performance indicates a more favorable balance between speckle suppression and structural preservation. Specifically, although NLM obtains the highest PSNR and E-S2SNet achieves a marginally higher PSNR than S2SNet, the proposed method consistently delivers superior results on structure-sensitive metrics, including SSIM, TP, EP, and $HFEN\_L1$. This confirms that explicitly penalizing low-energy, overly smooth feature representations successfully mitigates the blurring tendencies of traditional self-supervised networks, ensuring the protection of high-frequency anatomical details.

Figures 2 and 3 illustrate that E-S2SNet produces speckle-suppressed retinal images with well-preserved, sharp layers, contrast, and clear layer boundaries. Zoomed patches retained structural details without over-smoothing or blocking artifacts. The proposed method achieves the trade-off between noise suppression and structural preservation, avoiding over-smoothing present in NLM and S2SNet and blocking artifacts present in BM3D, making retinal layers clinically interpretable.

Table 3 reports segmentation performance of E-S2SNet versus Noisy Input and S2SNet across nine retinal layers on the NR206 dataset. E-S2SNet achieved the highest Dice scores in eight layers and the lowest Hausdorff distances in nearly all, confirming superior accuracy and delineation. Most notably, the OPL layer showed marked improvement with E-S2SNet (75.65%) over Noisy Input (46.03%) and S2SNet (58.57%), highlighting the benefit of speckle suppression for preserving fine detail. Although Noisy Input marginally outperformed E-S2SNet on NFL Dice, E-S2SNet performed best overall for retinal layer delineation in speckle-corrupted OCT images.

Statistical significance analyses assessed robustness using downstream segmentation Dice scores (Table S6 of supplementary material) and Hausdorff Distance (Table S7 supplementary material). Results show statistically significant improvements ($p < 0.05$), supporting E-S2SNet in preserving retinal structures while suppressing speckle noise. The model comprises ≈1.2M parameters, occupying 4.76 MB in single-precision (FP32) format. Evaluated on NVIDIA RTX PRO 6000 GPU ($N = 400$), inference requires 1350.0 MB VRAM, 1.8 TFLOPS, and 1.52 s/image (≈0.66 FPS). Although unsuitable for real-time scanning, footprint enables offline processing on workstations.Near-real-time preview can be accelerated by reducing stochastic forward passes (N), with inference times being presented in Table-S2 of supplementary material.

The proposed approach is less susceptible to dataset shift. However, a limitation of this study is that the current evaluation is restricted to the SD-OCT benchmark and the healthy-subject NR206 dataset. Future work is required to fully validate the clinical generalization of this approach across different OCT devices, multi-center datasets, and a variety of pathological cases.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. D. Huang, E. A. Swanson, C. P. Lin *et al.*, Science **254**, 1178 (1991).
2. B. Karamata, K. Hassler, M. Laubscher, and T. Lasser, J. Opt. Soc. Am. A **22**, 593 (2005).
3. H. Ahmed *et al.*, J. Imaging **10**, 86 (2024).
4. A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, , vol. 2 (2005), pp. 60–65.
5. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, IEEE Trans. Image Process. **16**, 2080 (2007).
6. B. R. Chintada, S. Ruiz-Lopera, R. Restrepo, *et al.*, Biomed. Opt. Express **15**, 4453 (2024).
7. D. Zhang, F. Zhou, Y. Jiang, *et al.*, Comput. Vis. Image Underst. **269**, 104786 (2026).
8. C. Ge, X. Yu, M. Yuan, *et al.*, Biomed. Opt. Express **15**, 1233 (2024).
9. Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2Self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2020), pp. 1890–1898.
10. L. Fang, S. Li, Q. Nie, *et al.*, Biomed. Opt. Express **3**, 927 (2012).
11. X. He, Y. Wang, F. Poiesi, *et al.*, Front. Bioeng. Biotechnol. **11**, 1191803 (2023).
12. X. Kong, K. Li, Q. Yang, *et al.*, "A new image quality metric for image auto-denoising," in *Proceedings of the IEEE International Conference on Computer Vision*, (2013), pp. 2888–2895.
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2016), pp. 770–778.

FULL REFERENCES

1. D. Huang, E. A. Swanson, C. P. Lin *et al.*, "Optical coherence tomography," Science **254**, 1178–1181 (1991).
2. B. Karamata, K. Hassler, M. Laubscher, and T. Lasser, "Speckle statistics in optical coherence tomography," J. Opt. Soc. Am. A **22**, 593–596 (2005).
3. H. Ahmed *et al.*, "Denoising of optical coherence tomography images in clinical practice: A review," J. Imaging **10**, 86 (2024).
4. A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* vol. 2 (2005), pp. 60–65.
5. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Trans. Image Process. **16**, 2080–2095 (2007).
6. B. R. Chintada, S. Ruiz-Lopera, R. Restrepo, *et al.*, "Probabilistic volumetric speckle suppression in oct using deep learning," Biomed. Opt. Express **15**, 4453–4469 (2024).
7. D. Zhang, F. Zhou, Y. Jiang, *et al.*, "Unleashing the power of self-supervised image denoising: A comprehensive review," Comput. Vis. Image Underst. **269**, 104786 (2026).
8. C. Ge, X. Yu, M. Yuan, *et al.*, "Self-supervised self2self denoising strategy for oct speckle reduction with a single noisy image," Biomed. Opt. Express **15**, 1233–1252 (2024).
9. Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2Self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* (2020), pp. 1890–1898.
10. L. Fang, S. Li, Q. Nie, *et al.*, "Sparsity based denoising of spectral domain optical coherence tomography images," Biomed. Opt. Express **3**, 927–942 (2012).
11. X. He, Y. Wang, F. Poiesi, *et al.*, "Exploiting multi-granularity visual features for retinal layer segmentation in human eyes," Front. Bioeng. Biotechnol. **11**, 1191803 (2023).
12. X. Kong, K. Li, Q. Yang, *et al.*, "A new image quality metric for image auto-denoising," in *Proceedings of the IEEE International Conference on Computer Vision,* (2013), pp. 2888–2895.
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* (2016), pp. 770–778.