Supplemental document accompanying submission to *Optics Letters*

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

Authors: Sivalal Kethavath, Vasundhara Vasundhara, MOHAN KUMAR MANEPALLI, Bhaskara Rao Chintada, Phaneendra Yalavarthy

Submitted: 08/04/2026 06:47:06

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

S. Kethavath, Vasundhara, M. K. Manepalli, B. R. Chintada and P. K. Yalavarthy

S1. Ablation Studies

S1.1 Ablation Study 1 – Effect of Energy Loss Weight (β)

Table S1: Quantitative results evaluating the effect of the energy loss weight (β) on model performance.

β	PSNR (↑)	SSIM(↑)	MSR(↑)	CNR (↑)	TP (↑)	EP (↑)	HFEN_L1(↓)
β = 0.00	24.06	0.6284	22.25	14.73	0.8300	0.5273	0.1249
β = 0.01	24.37	0.6557	21.97	15.08	0.7930	0.5434	0.1695
β = 0.02	24.51	0.6337	22.71	14.83	0.7966	0.5590	0.0517
β = 0.03	24.25	0.6386	23.00	14.83	0.8300	0.5698	0.1396
β = 0.04	24.63	0.6557	22.48	15.69	0.7962	0.5749	0.0928
β = 0.05	24.33	0.6508	22.64	15.69	0.8279	0.5700	0.0691
β = 0.075	**24.83**	**0.6691**	22.91	15.69	0.8240	0.5488	0.0375
β = 0.10 (Optimal)	24.81	0.6670	**23.06**	**15.72**	**0.8347**	**0.5774**	**0.0674**
β = 0.15	24.32	0.6571	22.90	15.33	0.7988	0.5465	0.0706
β = 0.20	24.03	0.6403	21.98	15.24	0.8226	0.5700	0.1095
β = 0.25	24.59	0.6521	22.06	14.46	0.7697	0.5147	0.1088
β = 0.30	23.73	0.6240	20.56	13.85	0.7063	0.4947	0.2668
β = 0.35	23.15	0.5886	20.42	13.25	0.6524	0.4725	0.3765
β = 0.40	22.83	0.5462	19.82	12.59	0.6626	0.4443	0.3198
β = 0.45	22.17	0.5417	19.29	12.23	0.5808	0.4247	0.4066
β = 0.50	22.19	0.5121	18.34	11.91	0.5285	0.3641	0.6310

As shown in Table S1, the energy loss component is critical for optimal model performance. Disabling this loss (β = 0.00) results in sub-optimal denoising and structural preservation. Gradually increasing β improves performance up to a distinct peak at **β = 0.10**, which provides the optimal balance of noise suppression and feature preservation. Conversely, setting the weight too high (β > 0.15) leads to a steady decline in performance (e.g., PSNR drops to 22.19 at β = 0.50), indicating that an overly dominant energy loss over-regularizes the network and degrades image fidelity. Therefore, β = 0.10 is established as the optimal default configuration

S1.2 Ablation Study 2 – Effect of Stochastic Passes (N)

Table S2: Qualitative results evaluating the effect of number of stochastic prediction passes(N).

N	Inference Time(sec)	PSNR (↑)	SSIM (↑)	MSR (↑)	CNR (↑)	TP (↑)	EP (↑)	HFEN_L1 (↓)
N = 50	0.59	22.18	0.4958	18.04	11.75	0.5082	0.3894	0.5054
N = 100	1.12	22.71	0.5417	18.48	13.10	0.6035	0.4291	0.4192
N = 200	2.33	23.44	0.5812	20.30	14.17	0.6969	0.5159	0.2478
N = 300	3.51	24.81	0.6504	21.95	15.44	0.7517	0.5508	0.0885
N = 400 (Optimal)	**5.39**	**24.84**	**0.6642**	**23.01**	**15.70**	**0.8379**	**0.5863**	**0.0672**

N = 500	6.04	24.58	0.6557	22.48	15.42	0.8495	0.5700	0.0800
N = 600	7.26	24.42	0.6449	23.00	15.69	0.7979	0.5741	0.1115
N = 700	8.39	24.41	0.6451	22.62	15.69	0.8300	0.5642	0.0696
N = 800	9.34	24.41	0.6554	22.74	14.86	0.8286	0.5571	0.1387
N = 900	10.40	24.17	0.6244	22.05	14.48	0.7944	0.5654	0.1204
N = 1000	12.14	24.18	0.6275	22.04	14.23	0.7238	0.5511	0.2150

Table S2 illustrates a clear convergence trajectory. At lower values (N = 50 to 100), the network exhibits high prediction variance, leading to inadequate speckle suppression and lower contrast (CNR: 11.75 at N = 50). Performance scales logarithmically, reaching an optimal saturation point at N = 400, where we observe the highest PSNR (24.84) and optimal edge preservation (0.5863). Interestingly, extending the passes beyond N = 500 yields diminishing returns and even a slight degradation in structural metrics (TP drops to 0.7238 at N = 1000). This suggests that excessive averaging begins to aggressively smooth out the deterministic anatomical details alongside the stochastic speckle variance. Thus, N = 400 represents the optimal trade-off between computational overhead and image quality.

Please note that the inference times reported in Table S2 (e.g., 5.39 seconds per image for N=400) reflect initial experiments conducted on an NVIDIA Titan RTX GPU. The improved inference time of 1.52 seconds per image reported in the main text was measured during subsequent evaluations using a newer NVIDIA RTX 6000 (Blackwell architecture) GPU.

S1.3 Ablation Study 3 – Effect of Dropout Rate

Table S3: Effect of Dropout Rate

Dropout	**PSNR(↑)**	**SSIM (↑)**	**MSR (↑)**	**CNR (↑)**	**TP (↑)**	**EP (↑)**	**HFEN_L1(↓)**
Dropout = 0.1	21.99	0.5121	17.80	12.02	0.5259	0.3952	0.5197
Dropout = 0.2	23.49	0.6341	20.59	14.18	0.7340	0.5291	0.2933
Dropout =0.3 (Optimal)	**24.84**	**0.6652**	**23.06**	**15.66**	**0.8362**	**0.5851**	**0.0671**
Dropout = 0.4	24.83	0.6353	22.50	14.96	0.8262	0.5700	0.1281
Dropout = 0.5	24.39	0.6263	21.96	15.02	0.7596	0.5529	0.1639
Dropout = 0.6	24.14	0.6025	21.92	13.90	0.6907	0.5189	0.2095
Dropout = 0.7	19.85	0.3966	15.47	8.64	0.3764	0.2722	0.9379
Dropout = 0.8	17.92	0.3018	11.53	6.51	0.1402	0.0843	1.2633
Dropout = 0.9	15.48	0.1682	6.52	2.87	-0.1783	-0.0874	1.7492

As shown in Table S3, a conservative dropout rate (0.1) fails to introduce enough variance across the forward passes, resulting in sub-optimal noise suppression (MSR: 17.80) because the ensemble instances are too similar. Conversely, aggressive dropout rates (> 0.5) introduce severe information bottlenecking; the network loses too much spatial context to accurately reconstruct the retinal boundaries, leading to a catastrophic collapse in SSIM (0.1682 at Dropout = 0.9) and a spike in high-frequency errors. The empirical sweet spot

is clearly located at a dropout rate of 0.3, which perfectly balances the need for prediction variance (for speckle averaging) with feature retention (for anatomical reconstruction).

S1.4 Ablation Study 4 – Effect of Masking Probability

Table S4: Effect of Masking Probability.

Mask Prob	**PSNR (↑)**	**SSIM (↑)**	**MSR (↑)**	**CNR (↑)**	**TP (↑)**	**EP (↑)**	**HFEN_L1 (↓)**
Mask =0.1	21.76	0.5123	18.17	11.70	0.5098	0.3528	0.5641
Mask =0.2	23.13	0.5875	20.27	12.87	0.6781	0.4481	0.4252
Mask =0.3	23.41	0.6101	21.31	14.02	0.7065	0.5357	0.3094
Mask =0.4	24.83	0.6427	22.82	14.73	0.8088	0.5707	0.1399
Mask=0.5 (Optimal)	**24.83**	**0.6657**	**22.97**	**15.67**	**0.8359**	**0.5726**	**0.0694**
Mask =0.6	24.08	0.6557	21.65	15.39	0.8067	0.5275	0.0641
Mask =0.7	23.99	0.6297	21.15	14.04	0.7057	0.4861	0.2189
Mask =0.8	23.18	0.5553	19.05	12.94	0.6139	0.4336	0.4343
Mask =0.9	21.99	0.5128	18.10	11.57	0.5835	0.3798	0.6364

Table S4 indicates that if the masking probability is too low (e.g., 0.1), the network faces a trivial identity-mapping task, leading to poor generalization and weak denoising (PSNR: 21.76). Conversely, if the masking probability is excessively high (e.g., 0.8 or 0.9), too much of the input B-scan is obscured, depriving the network of the necessary local receptive field context to rebuild continuous retinal layers. A mask probability of 0.5 forces the network to rely on complex, non-local spatial correlations to fill in the blind spots, yielding the most robust feature representations and resulting in peak performance across all clinical metrics (PSNR: 24.83, CNR: 15.67).

S1.5 Ablation Study 5 – Effect of Layer Injection Depth

Table S5: Effect of Masking Probability.

Block	**Layer**	**PSNR(↑)**	**SSIM(↑)**	**MSR(↑)**	**CNR(↑)**	**TP (↑)**	**EP (↑)**	**HFEN_ L1(↓)**
Decoder (H/4 ×W /4)	dec_ conv3b	24.34	0.6557	**23.00**	15.02	0.8190	0.5700	0.0891
Decoder (H/2 × W/2) Optimal	dec_ conv2b	24.80	**0.6658**	22.94	**15.67**	**0.8470**	**0.5737**	**0.0691**
Pre-Conv (H × W)	dec_ conv1b	**24.83**	0.6557	22.46	14.97	0.8300	0.5700	0.1209

Table S5 compares the effect of applying the regularizer at different spatial resolutions within the decoder. Applying the regularizer at an early high-resolution stage (dec_conv1b) may cause residual speckle variations to be interpreted as structural information. In contrast, applying it deeper in the bottleneck (dec_conv3b, H/4 × W/4) operates on overly compressed features, reducing sensitivity to fine retinal layer boundaries. The dec_conv2b layer (H/2 × W/2) provides a better abstraction level, where features are sufficiently refined to suppress speckle fluctuations while retaining adequate spatial information to represent retinal anatomical interfaces.

S1.6 Ablation Study 6 [Table S6-S8] : Statistical and no-reference evaluations of tissue structure preservation.

Table S6: Dice Score Significance.

Comparison	**t-statistic**	**t-test p-test**	**Statistical Significance**	**Winner**
S2SNet vs E-S2SNet	-5.5957	0.0072	Yes	E-S2SNet
S2SNet vs Noisy	6.4512	0.0046	Yes	S2SNet
E-S2SNet vs Noisy	9.7960	0.0037	Yes	E-S2SNet

Table S6: Statistical Significance of Downstream Segmentation Accuracy (Dice Score). Two-tailed t-tests confirm that the proposed E-S2SNet achieves a statistically significant improvement ($p < 0.05$) in segmentation overlap compared to both the noisy baseline and the base S2SNet architecture, providing indirect quantitative evidence of enhanced tissue structure preservation.

Table S7: Hausdorff Distance Significance (pixels)

Comparison	**t-statistic**	**t-test p-test**	**Statistical Significance**	**Winner**
S2SNet vs ES2SNet	4.1623	0.0087	Yes	E-S2SNet
S2SNet vs Noisy	7.1858	0.0690	No	S2SNet
E-S2SNet vs Noisy	-8.6653	0.0290	Yes	E-S2SNet

Table S7 illustrates the Statistical Significance of Boundary Preservation (Hausdorff Distance). To further validate the preservation of layer boundaries without relying on direct image ground truth, topological accuracy was assessed via Hausdorff Distance (in pixels). E-S2SNet yields a statistically significant reduction in boundary error against both S2SNet and the noisy baseline.

Table S8: Mean NIQM Score for Various Denoising Methods on the NR206 Dataset.

Denoising Method	**Mean NIQM Score (↑)**
Raw Noisy	N/A
Noise2Void	0.4501
NLM	0.2983
BM3D	0.2696
Blind2Unblind	0.1828
S2SNet	0.5853
E-S2SNet(Ours)	**0.7707**

Because the NR206 dataset lacks speckle-suppressed ground truth, quantitative evaluation was expanded beyond MSR and CNR to include the completely blind NIQM image quality metric. Higher scores (↑) indicate superior structural fidelity, edge preservation, and natural texture statistics. E-S2SNet significantly outperforms both traditional (NLM, BM3D) and deep learning baselines.

S1.7 Ablation Study 7

Table S9: Performance of SegResNet50 across various Denoising configurations.

Dataset Configuration	Model	Mean Dice Score (↑)	Hausdorff Distance (↓)
Raw Noisy	SegResNet50	0.8497	9.2813
Denoised(NLM)	SegResNet50	0.8759	5.2117
Denoised(BM3D)	SegResNet50	0.8702	5.9907
Denoised(Noise2Void)	SegResNet50	0.8692	3.6806
Denoised(Blind2Unblind)	SegResNet50	0.8546	8.3619
Denoised(Standard Self2Self)	SegResNet50	0.8804	3.1459
Denoised(S2SNet)	SegResNet50	0.8878	3.9756
Denoised(E-S2SNet) (Ours)	SegResNet50	**0.8942**	**2.4552**

Table S9 shows the impact of various denoising methods on downstream SegResNet50 segmentation. The proposed E-S2SNet significantly outperforms all baselines, achieving the highest Mean Dice Score (0.8942) and the lowest Hausdorff Distance (2.4552). Notably, the drastic reduction in boundary error from the base S2SNet to E-S2SNet demonstrates that our energy-regularized approach effectively preserves critical layer boundaries and structural edges rather than blurring them.

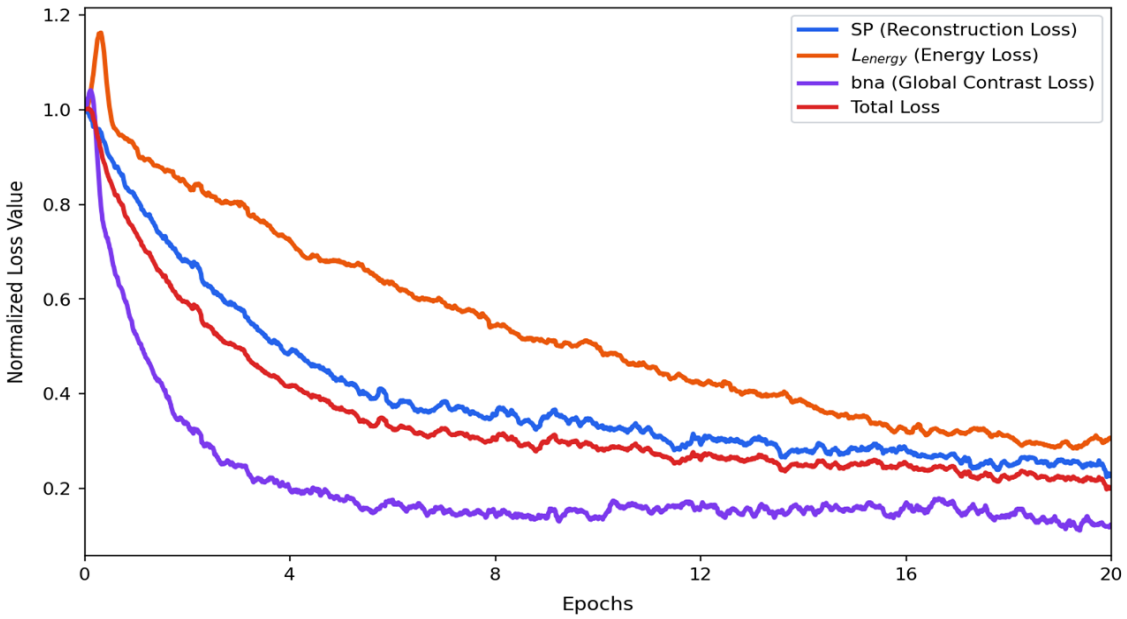

Figure S1: Normalized training loss curves over 2000 epochs. The plot illustrates the steady convergence of the Total Loss (red) alongside its constituent components: Reconstruction Loss (blue), Energy Loss (orange), and Global Contrast Loss (purple).

As shown in the plots, the proposed energy loss becomes effective after initial few iterations (acting as a regularizer). This downward monotonic trend is desirable for not causing vanishing gradients as well as confirming better learning for the task at hand. The convergence of the energy loss along with other task

specific losses validates the proposed loss term being numerically stable and not causing instability into the optimization process. The total loss makes the network learn what feature is relevant and what to reduce. Overall, as shown in the figure, the total loss reaches a plateau (saturation of learning) demonstrating the convergence.

S1.8 Implementation Details for Self-Supervised Baselines

All self-supervised baselines utilized the exact same protocols as our proposed method to ensure a fair comparison. Specifically, all models applied identical cropping and resizing as pre-processing steps and were trained and evaluated on individual images, maintaining the same data configuration as S2SNet. Furthermore, the ROI patch selection was consistent across all evaluations. The specific patch used to calculate Edge Preservation (EP) was selected to span multiple layers, including the Retinal Pigment Epithelium (RPE). The remaining two patches were utilized for Texture Preservation (TP) and Contrast-to-Noise Ratio (CNR) calculations.