Coherent Differentiable Inversion for High-Fidelity Digital Phantom Generation in Optical Coherence Tomography

Ura Modi* and Mehul S Raval

Ahmedabad University, Ahmedabad, India
{ura.m,mehul.raval}@ahduni.edu.in

Abstract. Most synthetic optical coherence tomography (OCT) methods work in image space. SynthOCT 2026 instead asks for a *digital phantom*, represented as a distribution of optical scatterers, that reproduces a target real B-scan after processing by a fixed physics-based Virtual Scanner. We use train-free coherent inversion performed separately for each scan. A differentiable implementation of the published Fourier-domain OCT model fits a complex reflectivity field, decodes it into approximately 2.95×10^5 amplitude-capped sub-scatterers, and refines their depths and energies with scorer-aligned losses. The objective covers the structural image, optical attenuation coefficient (OAC), speckle contrast (SC), refined speckle contrast (RSC), MS-SSIM, and LPIPS. The submitted preliminary set achieved an official global score of 0.9947, median MS-SSIM of 0.9971, and median LPIPS of 0.0054, tying for first on the preliminary leaderboard. Optimisation uses only the differentiable surrogate, while the fixed scanner is reserved for verification and organiser scoring. All recorded Linux/GPU runs met the 600-s limit. A capped GTX 1650 audit took 445.73 s and used 1.733 GiB peak allocated VRAM. Direct Windows/RTX 3060 execution has not been performed.

Keywords: Optical coherence tomography · Digital phantoms · Inverse scattering · Differentiable rendering · Speckle · Synthetic medical imaging.

1 Introduction

Optical coherence tomography (OCT) forms cross-sectional tissue images from coherent near-infrared backscattering. Limited annotated in-vivo data motivate physics-based synthesis for foundation-model training [10]. A useful OCT simulator must reproduce optical attenuation and speckle statistics, not only perceptual appearance.

The SynthOCT 2026 challenge, part of SASHIMI at MICCAI 2026, frames this task as an inverse problem [6]. Participants submit lists of point scatterers with spatial coordinates and backscattering energies rather than synthetic images. A fixed Virtual Scanner produces each B-scan. The resulting Struct, OAC,

* Corresponding author: ura.m@ahduni.edu.in

SC, and RSC maps are compared with the real reference using MS-SSIM and LPIPS.

We optimise a separate scanner-compatible phantom for each target through a differentiable approximation of the fixed scanner. In preliminary experiments, OAC remained a clear residual error after coherent initialisation, which motivated direct optimisation of the evaluated maps. Our contributions are:

1. A differentiable, physics-based forward model of the SynthOCT Virtual Scanner, implemented from the published Fourier-domain OCT discrete-scatterer formulation and validated against the provided scanner (Sec. 3.2, Fig. 1).
2. A coherent inversion procedure that combines a minimum-energy field initialisation, an amplitude-capped coherent decode into many sub-scatterers, and render-in-the-loop refinement of sub-wavelength depths (Sec. 3.4).
3. A scorer-aligned refinement objective in which all four evaluated maps contribute explicitly and both challenge metrics are represented by differentiable terms (Sec. 3.5).
4. An official preliminary evaluation that attains global score 0.9947 and tied rank one, together with ablations of the preliminary optimisation stages (Sec. 4).

2 Related Work

Discrete-scatterer OCT simulation. Physics-based OCT simulators model tissue as point scatterers and form each A-scan from their coherent sum, followed by envelope detection and log compression [15,9]. The SynthOCT scanner follows the Fourier-domain formulation of Matveev et al. [7], with a Gaussian lateral beam and depth-cumulative shadowing. We implement this model in differentiable form.

Physics-based OCT maps. Beyond the raw intensity B-scan (Struct), the challenge derives three quantitative maps. These are the optical attenuation coefficient (OAC), computed with the depth-integral estimator of Vermeer et al. [13], speckle contrast (SC), and refined speckle contrast (RSC) [7,8]. Each map captures a different property of the intensity B-scan. Differentiable approximations of the scorer let us optimise these properties directly.

Speckle statistics. Fully developed Rayleigh speckle arises when a resolution cell contains many independent scatterers with random phase [4]. The required density depends on the medium and imaging configuration [1]. Our decoder instead uses phase-aligned sub-scatterers to control the coherent cell resultant. Spatial jitter distributes them within the cell, and scorer-aligned refinement encourages agreement in SC and RSC.

Phase retrieval and differentiable coherent rendering. Recovering a complex field from magnitude measurements is a phase-retrieval problem with non-convex ambiguities. Classical alternating-projection methods and modern gradient methods characterise this landscape and motivate our complex-field initialisation [3,2]. Related coherent-imaging work optimises through a calibrated differentiable propagation model [11]. Our stage is more accurately termed *render-in-the-loop*: the surrogate supplies every optimisation gradient, while the fixed Virtual Scanner is reserved for verification and scoring.

3 Method

3.1 Problem Formulation

A digital phantom is a set of scatterers $\Phi = \{(x_j, y_j, z_j, E_j)\}_{j=1}^{N}$ with coordinates in micrometres and backscattering energy $E_j \in [0, 100]$ in percent. The scanner converts energy to reflection amplitude as $a_j = \sqrt{E_j/100}$. The fixed Virtual Scanner $\mathcal{S}$ maps a phantom to an 8-bit B-scan $\mathcal{S}(\Phi) \in \{0, \ldots, 255\}^{H\times W}$. We target a budget of 295,000, below the competition's 3×10^5 cap. Given a real reference B-scan Y, we seek

$$\Phi^\star = \arg\min_{\Phi} \mathcal{L}\big(\mathcal{S}(\Phi),\, Y\big), \tag{1}$$

subject to $|x_j| \le 1536\,\mu\text{m}$, $z_j \in [0, 1536\,\mu\text{m}]$, and $N \le 3 \times 10^5$. Because $\mathcal{S}$ is a fixed black box at submission time, we optimise Eq. (1) through a differentiable surrogate $\widetilde{\mathcal{S}} \approx \mathcal{S}$ and use $\mathcal{S}$ only to verify.

3.2 Differentiable Forward Model

Following the published discrete-scatterer Fourier-domain formulation [7,9], the complex signal of the A-scan at lateral position x_i is

$$A(z_q, x_i) = \mathrm{IDFT}_{k\to z}\left[S(k) \sum_j a_j\, B(x_i - x_j, y_j)\, L_j\, e^{-2\mathrm{i}kz_j} \right], \tag{2}$$

where $S(k)$ is the source spectrum over wavenumber k, $B(\Delta x, y) = \exp\big(-(\Delta x^2 + y^2)/w_0^2\big)$ is the Gaussian lateral beam with radius $w_0 = 10\,\mu$m, and L_j is a depth-cumulative shadowing factor $L_j = \prod_{m:\, z_m < z_j} \big(1 - \min((B\, a_m)^2, 1)\big)$. The image is the log-compressed envelope $I = \mathrm{clip}\big(20 \log_{10} |A|,\, 51\,\mathrm{dB}\big)$ mapped to 8 bits. The axial carrier $e^{-2\mathrm{i}kz}$ has phase $\varphi(z) = 4\pi n z/\lambda$ with $\lambda = 1.3\,\mu$m. It advances by many cycles over a $6\,\mu$m pixel. As a result, sub-wavelength changes in z_j move the speckle, and the scatterer depths must be resolved to a fraction of a wavelength.

We implement Eq. (2) in PyTorch with fixed physical parameters. Calibration and hard candidate selection use double precision, while gradient-based optimisation uses float32/complex64 to control memory. In a double-precision consistency test, the vectorised and reference implementations agree to a maximum absolute difference of 5×10^{-15}. We check scanner agreement by rendering the same known phantom with both implementations (Fig. 1).

3.3 Phantom Parameterisation

A single scatterer per resolution cell can control local phase, but it produces an unrealistically sparse cloud. An unconstrained random cloud has more plausible ensemble statistics, but it cannot be directed towards one measured realisation. For retained cell magnitude m_c, we choose an amplitude cap $a_{\max}$ and set $K_c = \lceil m_c/a_{\max} \rceil$. The cap is increased by 10% until $\sum_c K_c \leq 295000$. Each child receives amplitude m_c/K_c and scanner energy $E_{cj} = 100(m_c/K_c)^2$. Its axial coordinate encodes the parent phase. Lateral coordinates are jittered within the cell, and out-of-plane positions are sampled within the accepted interval. This construction preserves the coherent cell resultant while reducing the quadratic shadowing contribution of each child. The later refinement encourages agreement in SC and RSC. The supplement describes the complete decoding procedure.

3.4 Coherent Inversion

We solve Eq. (1) for each B-scan in three stages. **(1) Complex-field initialisation.** A complex reflectivity field is initialised from the de-logged target magnitude and optimised for 3000 Adam iterations. We use an ℓ_1-plus-MSE image loss, a high-pass ℓ_1 term with weight 0.5, and an ℓ_2 field-energy coefficient of 2.0. The learning rate starts at 0.05 and follows cosine annealing. Cumulative shadowing is disabled during this stage. **(2) Amplitude-capped coherent decode.** The field is converted into the discrete cloud described above, and the full shadowing model is restored. **(3) Render-in-the-loop refinement.** Sparse depth and log-energy variables are refined for 800 legacy-loss iterations followed by 200 scorer-exact iterations. Four 100-iteration exact-loss restarts use learning rates of 0.002, 0.002, 0.001, and 0.0005. The first restart retains iteration 100. Each later restart selects the best result among its parent, iteration-50, and iteration-100 candidates using the input-local mean of the eight challenge components. Only complete candidates can replace the incumbent. The fixed Virtual Scanner is not queried during optimisation.

3.5 Metric-Aligned Refinement

The challenge evaluates four maps with two metrics (Sec. 4.2), so a B-scan pixel loss is only a proxy for the scored objective. OAC remained a prominent residual after coherent initialisation. We therefore optimise differentiable approximations of the scored functionals. The refinement loss is

$$
\begin{aligned}
\mathcal{L} = {} & \lambda_{\text{pix}}\big(\|P - P^{\text{ref}}\|_1 + \|P - P^{\text{ref}}\|_2^2\big) + \lambda_{\text{sc}} \left\|\mathrm{SC}(P) - \mathrm{SC}(P^{\text{ref}})\right\|_1 \\
& + \lambda_{\text{oac}}\left\|\widehat{\mathrm{OAC}}(P) - \widehat{\mathrm{OAC}}(P^{\text{ref}})\right\|_1 + \lambda_{\text{lo}}\,\mathrm{LPIPS}\big(\widehat{\mathrm{OAC}}(P), \widehat{\mathrm{OAC}}(P^{\text{ref}})\big) \\
& + \lambda_{\text{mo}}\big(1 - \mathrm{MSSSIM}(\widehat{\mathrm{OAC}}(P), \widehat{\mathrm{OAC}}(P^{\text{ref}}))\big) \\
& + \lambda_{\text{ms}}\big(1 - \mathrm{MSSSIM}(P, P^{\text{ref}})\big) \\
& + \lambda_{\text{lp}}\,\mathrm{LPIPS}(P, P^{\text{ref}}) \\
& + \lambda_{\text{ssc}}\,\mathrm{LPIPS}(\widehat{\mathrm{SC}}_P, \widehat{\mathrm{SC}}_{\text{ref}}) \\
& + \lambda_{\text{ssc}}\,\mathrm{LPIPS}(\widehat{\mathrm{RSC}}_P, \widehat{\mathrm{RSC}}_{\text{ref}}),
\end{aligned} \tag{3}
$$

where $P = \widetilde{\mathcal{S}}(\Phi)$ is the modelled log image. Subscripts P and ref denote maps computed from the modelled and reference images. The operators reproduce the scorer's serialised forward values and use straight-through gradients for quantisation. Each OAC map is normalised using its own minimum and 99th percentile. SC and RSC use the fixed range 0.5–5. The frozen weights for pix/sc/oac/lo are 1/0.1/1/0.4, while those for mo/ms/lp/ssc are 1/1.2/0.01/0.05. We selected them with a small coarse manual sweep on development scans and then kept them fixed. SC retains a direct ℓ_1 anchor, while RSC contributes through LPIPS. OAC receives ℓ_1, LPIPS, and MS-SSIM terms because it was a prominent preliminary deficit. All four scored maps therefore contribute to the objective, although they do not use every map–metric pairing with equal weight.

3.6 Implementation and Compliance

Each input is a 256×512 PNG. The program converts it to RGB, scales each channel to $[0, 1]$ in float32, and averages the channels to obtain grayscale without cropping, resizing, or denoising. It targets Python 3.12.12 and PyTorch 2.10, uses deterministic seed zero, and writes newline-delimited $x\ y\ z\ E$ rows with six-digit scientific-notation precision. It enforces $N \le 295000$, $x \in [-1536, 1536]$ μm, $z \in [0, 1536]$ μm, and $E \in [0, 100]$, with $y \in [-3, 3]$ μm. A 450-s internal deadline bounds the fixed iteration schedule, and only complete checkpoints replace the last valid output. A6000 profiles took 239.96–241.31 s per scan. Linux tests took 252.56 s on a GTX 1650 and 447.37 s in a network-disabled packaged smoke test. A fresh GTX 1650 process using Python 3.12.12, PyTorch 2.10.0, and `set_per_process_memory_fraction(0.70)` completed in 445.73 s with 1.733 GiB peak allocated and 1.953 GiB peak reserved VRAM. A mixed RTX 5060/A4000 rehearsal over 60 scans had no failures and a maximum process time of 172.13 s. All tested Linux/GPU configurations met the 600-s limit. Direct Windows/RTX 3060 and CPU-only execution were not measured.

4 Experiments

4.1 Dataset and Development Protocol

We use the official in-vivo human-skin OCT dataset [5,12,6]. The challenge provides 120 B-scans, of which 60 were used for preliminary phantom evaluation. Our method optimises each scan independently without shared learned parameters. We selected global loss weights and iteration counts through coarse manual inspection of development scans, then fixed them for all reported scans. We did not retain the exact development-scan manifest, which limits retrospective assessment of selection bias. No hidden hold-out data or final organiser evaluation was available.

4.2 Metrics and Protocol

Each phantom is rendered by the Virtual Scanner and compared with its matching real reference on the four maps (Struct/OAC/SC/RSC) using multi-scale SSIM (MS-SSIM [14], higher is better) and LPIPS [16] (lower is better). For each map m, we take the median over the N scans for per-scan MS-SSIM and per-scan $(1 - \text{LPIPS})$. The challenge *global score* is the mean of these eight map-level medians,

$$\text{Global} = \tfrac{1}{8} \sum_{m \in \{\text{Struct,OAC,SC,RSC}\}} \big(\widetilde{\text{MS}}_m + (1 - \widetilde{\text{LP}}_m)\big), \tag{4}$$

where $\widetilde{\cdot}$ denotes the median over scans. We report metrics computed by rendering our phantoms through the reference Virtual Scanner and scoring with the challenge harness, and we cite the official server score for the submitted set.

4.3 Forward-Model Fidelity

Figure 1 compares the differentiable renderer and reference Virtual Scanner for the same known phantom. Their outputs attain MS-SSIM 0.999 92 for this example. This supports close agreement for the displayed calibration case but does not establish equivalence outside the calibrated scanner configuration.

4.4 Main Results

The submitted preliminary set achieved an official global score of 0.9947 and was tied for first on the preliminary leaderboard. The official summary reports median MS-SSIM of 0.9971 and median LPIPS of 0.0054. Table 1 gives the locally reconstructed map-level medians for the same submitted set. The organiser's hidden hold-out evaluation had not yet been performed, so these results are not the final challenge ranking.

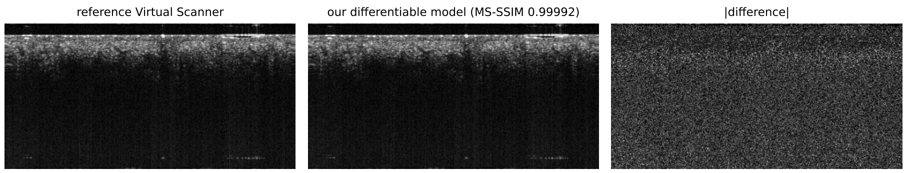

Fig. 1. Forward-model fidelity on the same known phantom. Left: reference Virtual Scanner. Middle: differentiable renderer (MS-SSIM 0.999 92). Right: absolute pixelwise difference over the displayed range. This panel is qualitative because the source artwork does not include a calibrated colour bar.

Table 1. Submitted preliminary set: locally reconstructed per-map medians over 60 scans. MS-SSIM and 1−LPIPS are both higher-is-better. Their eight-value mean is 0.994 675, consistent with the official global score 0.9947 after rounding.

Metric	Struct	OAC	SC	RSC	Global
MS-SSIM	0.9907	0.9967	0.9974	0.9977	**0.9947**
1−LPIPS	0.9905	0.9952	0.9946	0.9946	

4.5 Ablation

Figure 3 compares three archived preliminary configurations over the 60 scans. Coherent inversion with a pixel-plus-speckle-contrast objective reaches 0.9823. Replacing that objective with the scorer-aligned loss raises the score to 0.9935. Longer preliminary refinement with fixed reweighting reaches the submitted score of 0.9947. These configurations test the main stages of the preliminary optimisation pipeline rather than every restart in the final executable.

4.6 Forward-Model Characterisation

Figure 4 summarises three calibration probes: the logarithmic energy-to-intensity law and 51 dB display range, measured axial/lateral point-spread profiles, and the periodic two-scatterer interference response to sub-wavelength separation.

5 Discussion and Limitations

Direct inversion works well in this challenge setting, but we have not compared it with every learned alternative. The inverse problem is non-unique because different phases, out-of-plane positions, and scatterer configurations can produce similar envelopes and maps. We therefore describe the output as a *scanner-equivalent phantom*. It follows the forward physics and imposed constraints but is not guaranteed to recover the biological microstructure. Histology or independent multimodal data would be needed for biological interpretation of individual

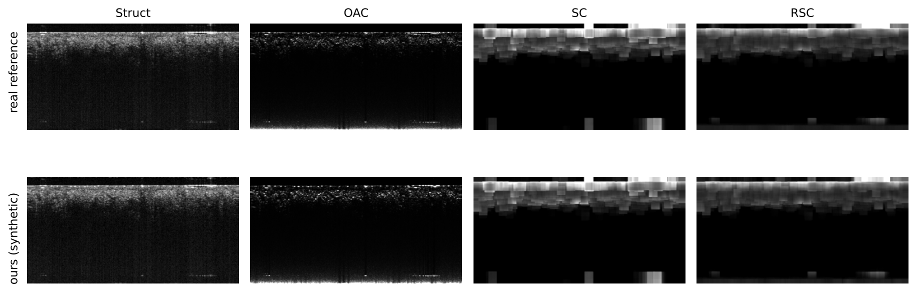

Fig. 2. Qualitative preliminary result. Top: real reference maps (Struct, OAC, SC, RSC). Bottom: the submitted phantom rendered by the fixed Virtual Scanner and processed identically. Processing is matched within each map. Absolute intensities should not be compared across maps because the panels show different physical quantities.

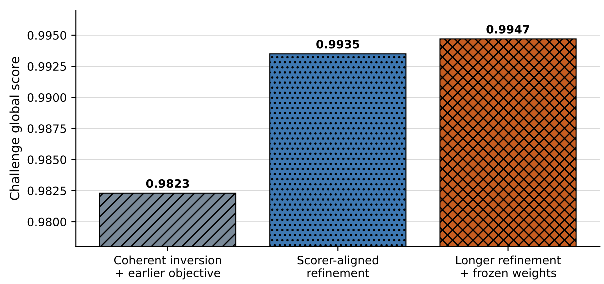

Fig. 3. Archived preliminary configurations evaluated over 60 scans: coherent inversion with the earlier objective (0.9823), scorer-aligned refinement (0.9935), and longer refinement with frozen reweighting (official submitted score 0.9947).

scatterers. Other limitations include calibration to one scanner, diminishing improvements near uint8 quantisation, CUDA dependence, and the absence of a direct Windows/RTX 3060 run. A learned initialiser followed by brief refinement may reduce runtime while preserving fidelity.

6 Conclusion

We cast SynthOCT phantom generation as target-conditioned coherent inversion and combine a differentiable forward model, amplitude-capped coherent decoding, and scorer-aligned render-in-the-loop refinement. The preliminary submission attained official global score 0.9947 and was tied for first on the preliminary leaderboard. The recovered output should be understood as a scanner-equivalent digital phantom for the specified forward operator, not a unique reconstruction of

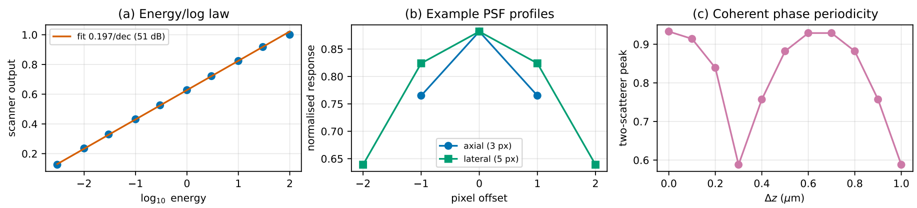

Fig. 4. Forward-model characterisation. (a) Energy-to-output relation and logarithmic fit over the 51 dB display range. (b) Example axial and lateral point-spread profiles. (c) Two-scatterer interference versus sub-wavelength separation Δz, revealing the coherent carrier periodicity.

the biological microstructure. Final performance will be determined by organiser execution on the hidden hold-out dataset.

Acknowledgments. We thank the SynthOCT 2026 organisers for the dataset, the Virtual Scanner, and the evaluation framework.

Funding This work received no external funding.

AI tools disclosure AI-assisted tools, including Anthropic Claude Code, Cursor AI, and OpenAI Codex, were used for code suggestions, debugging support, experiment scripting, and language editing. All scientific decisions, experiments, validation, and interpretation were carried out and checked by the authors. These tools had no access to hidden evaluation data or organiser scoring systems. The authors take full responsibility for the work and its reporting.

Disclosure of Interests. The authors declare no competing interests.

References

1. Almasian, M., van Leeuwen, T.G., Faber, D.J.: OCT amplitude and speckle statistics of discrete random media. Scientific Reports **7**(1), 14615 (2017). `https://doi.org/10.1038/s41598-017-14115-3`
2. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via Wirtinger flow: theory and algorithms. IEEE Transactions on Information Theory **61**(4), 1985–2007 (2015). `https://doi.org/10.1109/TIT.2015.2399924`
3. Fienup, J.R.: Phase retrieval algorithms: a comparison. Applied Optics **21**(15), 2758–2769 (1982). `https://doi.org/10.1364/AO.21.002758`
4. Goodman, J.W.: Speckle Phenomena in Optics: Theory and Applications. Roberts and Company (2007)
5. Matveev, L.A., Sovetsky, A.A., Petrova, K.S., Ryabkov, M.G., Brueva, M.A., Matveyev, A.L., Zaitsev, V.Y.: In vivo human skin optical coherence tomography (OCT) dataset for classification benchmarking and digital phantom generation. Zenodo (2026). `https://doi.org/10.5281/zenodo.18095266`

6. Matveev, L.A., Remedios, S., Zam, A., Hassan, T.: SynthOCT 2026: Digital phantoms simulation for physics-based scans synthesis in optical coherence tomography. Zenodo (2026). https://doi.org/10.5281/zenodo.19733396
7. Matveev, L.A., Sovetsky, A.A., Matveyev, A.L., Chizhov, P.A., Zabel, W.J., et al.: Online platform for generating realistic digital phantoms of OCT signals and performing multimodal processing towards optical cancer diagnostics. Light: Advanced Manufacturing **7**(2), 270 (2026). https://doi.org/10.37188/lam.2026.006
8. Matveev, L.A., Sovetsky, A.A., Matveyev, A.L., Vitkin, I.A., Assaad, M., Zaitsev, V.Y.: Refined speckle contrast estimation in OCT based on compensation of scattering-related distortions of speckle pattern parameters. Laser Physics Letters **22**(1), 015601 (2025). https://doi.org/10.1088/1612-202X/ad9149
9. Matveyev, A.L., Matveev, L.A., Moiseev, A.A., Sovetsky, A.A., Gelikonov, G.V., Zaitsev, V.Y.: Simulating scan formation in multimodal optical coherence tomography: angular-spectrum formulation based on ballistic scattering of arbitrary-form beams. Biomedical Optics Express **12**(12), 7599–7615 (2021). https://doi.org/10.1364/BOE.440739
10. Nikoshin, D., Mikhailenko, D., Sovetsky, A., Matveyev, A., Zaitsev, V., Matveev, L.: From tissue-mimicking phantoms to physics-based scans: synthetic OCT for few-shot foundation model training. In: Simulation and Synthesis in Medical Imaging (SASHIMI 2025). Lecture Notes in Computer Science, vol. 16085, pp. 44–51 (2026). https://doi.org/10.1007/978-3-032-05573-6_5
11. Peng, Y., Choi, S., Kim, N., Wetzstein, G.: Neural holography with camera-in-the-loop training. In: ACM SIGGRAPH Asia (2020). https://doi.org/10.1145/3414685.3417802
12. Sovetsky, A.A., Petrova, K.S., Brueva, M.A., Ryabkov, M.G., Matveyev, A.L., Matveev, L.A., Zaitsev, V.Y.: Automated segmentation and skin-layer thickness estimation by extracting the optical scattering coefficient and speckle contrast parameter from optical coherence tomography scans. Skin Pharmacology and Physiology (2026). https://doi.org/10.1159/000550613
13. Vermeer, K.A., Mo, J., Weda, J.J.A., Lemij, H.G., de Boer, J.F.: Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. Biomedical Optics Express **5**(1), 322–337 (2014). https://doi.org/10.1364/BOE.5.000322
14. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. Proc. Asilomar Conf. Signals, Systems and Computers **2**, 1398–1402 (2003). https://doi.org/10.1109/ACSSC.2003.1292216
15. Zaitsev, V.Y., Matveev, L.A., Matveyev, A.L., Gelikonov, G.V., Gelikonov, V.M.: A model for simulating speckle-pattern evolution based on close to reality procedures used in spectral-domain OCT. Laser Physics Letters **11**(10), 105601 (2014). https://doi.org/10.1088/1612-2011/11/10/105601
16. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR). pp. 586–595 (2018). https://doi.org/10.1109/CVPR.2018.00068