This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Technical Supplement: Coherent Differentiable Inversion for OCT Digital Phantoms

Ura Modi Mehul S Raval

1 Amplitude-capped coherent decode

Let $R_c = A_c e^{i\phi_c}$ be the complex reflectivity recovered for axial–lateral cell c. The central wavenumber is $k_0 = 2\pi/\lambda$, so the phase is encoded as the sub-pixel depth offset

$$\delta z_c = -\frac{\phi_c}{2k_0}. \tag{1}$$

Cells below 3×10^{-4} of the maximum recovered amplitude are discarded. For each remaining cell, the adaptive multiplicity is

$$K_c = \max\left(1, \left\lceil \frac{A_c}{a_{\max}} \right\rceil\right). \tag{2}$$

We initialise $a_{\max} = \max(\sum_c A_c/B, 10^{-4})$, where $B = 295000$ is the scatterer budget, and multiply $a_{\max}$ by 1.1 until $\sum_c K_c \le B$. Each copy has amplitude $a_{c,k} = A_c/K_c$, energy $E_{c,k} = 100a_{c,k}^2$, and depth $z_{c,k} = z_c + \delta z_c$. Its lateral coordinate is jittered uniformly by half a cell and its out-of-plane coordinate is sampled uniformly over the narrow beam-support interval.

The decode conserves the coherent cell resultant:

$$\sum_{k=1}^{K_c} a_{c,k} e^{i\phi_c} = A_c e^{i\phi_c} = R_c. \tag{3}$$

It intentionally does not conserve the arithmetic sum of scanner input energies. In fact, $\sum_k E_{c,k} = 100A_c^2/K_c$. The scanner adds amplitudes coherently before envelope detection. Therefore, $|\sum_k a_{c,k} e^{i\phi_c}|^2 = A_c^2$ is the relevant resolution-cell quantity. Splitting lowers each scatterer's quadratic shadowing term while retaining the desired coherent output. The subsequent render-in-the-loop stage corrects residual changes caused by lateral beam weights and cumulative attenuation.

2 Metric-aligned loss and map choices

For normalised log-intensity Y, the processor first forms linear intensity $I = 10^{4Y}$. Its three derived maps are

$$\mathrm{OAC}(I)_q = \frac{I_q}{2h\left(\sum_{r\ge q} I_r - I_q/2 + \epsilon\right)}, \tag{4}$$

$$\mathrm{SC}(I) = \frac{\sigma_w(I)}{\mu_w(I) + \epsilon}, \tag{5}$$

$$\mathrm{RSC}(I) = \mathrm{SC}(\mathrm{OAC}(I)), \tag{6}$$

where $h = 6\,\mu$m and w is the processor's local window.

The final exact objective contains pixel $\ell_1 + \ell_2$, SC ℓ_1, OAC ℓ_1, OAC LPIPS, OAC MS-SSIM, Struct MS-SSIM, Struct LPIPS, and both SC and RSC LPIPS terms. Their frozen weights in this order are $(1, 0.1, 1, 0.4, 1, 1.2, 0.01, 0.05)$. SC retains a direct local anchor because it is less processed. RSC remains

explicitly supervised by LPIPS. OAC receives three direct terms because it was a prominent preliminary deficit. The weights came from a small coarse manual sweep on a development subset, not an exhaustive grid search, and were frozen before the final executable was packaged.

3 Optimisation and termination

The complex field is initialised from the de-logged target magnitude with deterministic seed zero. Adam optimises its real and imaginary components for 3000 iterations using an image ℓ_1-plus-MSE term, a high-pass ℓ_1 term with weight 0.5, and a field-energy coefficient of 2.0. The learning rate starts at 0.05 and follows cosine annealing. Cumulative shadowing is disabled during this stage and restored after decoding. Thereafter, sparse refinement runs for 800 legacy-loss iterations and 200 scorer-exact polish iterations. The final executable then performs one 100-iteration restart at learning rate 0.002, followed by input-local 100-iteration restarts at 0.002, 0.001, and 0.0005. The first restart retains its iteration-100 output. At each later restart, the parent, iteration-50, and iteration-100 candidates are scored. Only complete candidates may replace the incumbent. There is no gradient-norm convergence test. A monotonic 450-second internal deadline preserves margin under the 600-second external limit. An atomically written valid output remains available if a later stage cannot complete.

4 Input and output contract

The entry point requires a 256×512 PNG. It converts the image to RGB, scales each channel to $[0, 1]$ in float32, and averages the channels. It performs no cropping, resizing, or denoising. The output is a newline-delimited text file containing x, y, z, E in six-digit scientific notation. The generator uses $x \in [-1536, 1536]$ μm, $y \in [-3, 3]$ μm, $z \in [0, 1536]$ μm, $E \in [0, 100]$, and at most 295000 scatterers. A schema-valid output is written before CUDA initialisation and is atomically replaced only after a complete valid stage.

5 Compute requirements

Three NVIDIA A6000 profiles took 239.96–241.31 seconds per scan and 0.339–0.371 GiB peak allocated memory. An exact Python 3.12.12/PyTorch 2.10 Linux replay on a GTX 1650 took 252.56 seconds. A network-disabled packaged smoke test on that device took 447.37 seconds wall time, 2.01 GB resident memory, and 1.746 GiB allocated memory. A fresh-process audit with `set_per_process_memory_fraction(0.70)` took 445.73 seconds, with 1.733 GiB peak allocated and 1.953 GiB peak reserved memory. A 60-scan rehearsal on mixed RTX 5060/A4000 hardware had zero failures and maximum process time 172.13 seconds. The final code path is CUDA-enabled and emits an emergency valid phantom when CUDA is unavailable. That fallback is not intended for challenge-quality scoring. We therefore regard a GPU as operationally required. No CPU-only end-to-end benchmark or direct Windows/RTX 3060 run has been completed, and we do not extrapolate either result.

6 Identifiability and interpretation

The recovered phantom is one element of a large equivalence class. A single 2-D B-scan does not uniquely determine a 3-D scatterer cloud, especially its out-of-plane coordinates. phase wrapping and coherent interference introduce further ambiguities. The minimum-energy prior, positivity and coordinate constraints, amplitude cap, and dense-cell construction select a low-energy scanner-compliant representative. These priors do not prove that the cloud matches the tissue's true histology. The output should therefore be interpreted as a metric-consistent digital phantom for the specified forward operator. Tissue-faithful recovery would require additional anatomical or statistical priors, multi-view data, or independent microstructural validation.