NOWA: Null-space Optical Watermark for Invisible Capture Fingerprinting and Tamper Localization

Supplementary Material

7. Hardware Prototype and Fabrication

In this section, we provide the specific manufacturing parameters and assembly details for the physical prototype used in our real-world validation.

7.1. Phase Mask Fabrication

The height profile h_ϕ of the phase mask was fabricated via Two-Photon Polymerization 3D lithography on a (10mm $\times$ 10mm) fused-silica substrate. We utilized a Nanoscribe Photonic Professional GT laser writing system operating in Dip-in Liquid Lithography (DiLL) mode. For reliable printing, the height map of the designed phase mask was discretized into 6 layers with steps of 200 nm. We employed the IP-DIP photoresist, which has a refractive index of $n = 1.52$ at 550 nm, closely matching the fused-silica substrate to ensure structural stability. A $63\times$ objective was used to focus the laser into the photoresist.

7.2. Optical Assembly

The optical setup consists of a Canon EOS 5D Mark IV DSLR (Full-frame sensor, $6.5\,\mu$m pixel pitch) and a commercial YONGNUO 50mm f/1.8 lens. To integrate the phase mask into the system, the lens was disassembled to access the physical aperture stop (see Fig. 6 left). Then, the fabricated phase mask was cut to a smaller size to fit the camera's aperture and attached to it using carbon tape. The substrate was cut to a smaller size to fit into the camera's aperture (see Fig. 6 right).

Figure 6. **Optical encoding**. A commercial YONGNUO lens is disassembled to attach the design phase mask on the back side of the lens aperture. The inset shows the quantized designed phase mask.

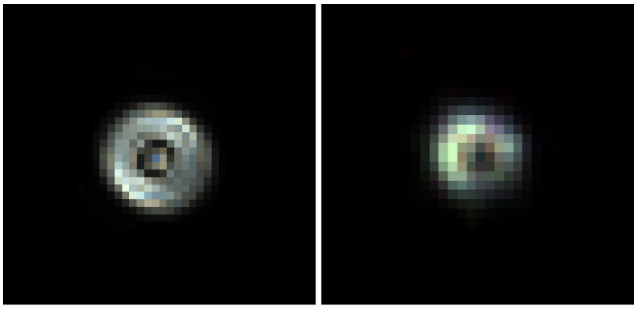

Figure 7. **PSF Calibration.** Comparison between the optimized simulated PSF (left) and the experimentally measured PSF from the physical prototype (right). The similarity in structure confirms the fidelity of the fabrication process.

8. PSF Measurement

A critical step in bridging the simulation-to-reality gap is the accurate characterization of the physical Point Spread Function (PSF) generated by the fabricated mask. We measured the system PSF using a point source setup: 1. A $20\,\mu$m pinhole was back-illuminated by a broadband white LED source. 2. The prototype camera was aligned to the optical axis. 3. The distance between the pinhole and camera is 50 cm. Figure 7 demonstrates the agreement between the simulated PSF and the physically measured PSF. The observed deviations are attributed to fabrication tolerances and misalignment of the mask relative to the optical axis.

9. Fine-Tuning and Implementation Details

While the Null-Space Network f_θ and detector d_ψ were initially trained in a purely simulated environment, real-world optical aberrations and sensor noise introduce a domain shift. To ensure robust protection in physical experiments, we performed a fine-tuning stage. Ideally, one would collect thousands of real-world captures to fine-tune the model; however, acquiring such a large-scale paired dataset imposes a prohibitive workload.

To address this, we adopted a hybrid training strategy. Instead of relying on the analytical wave-optics model, we employ the *measured* PSFs described in Section 8 combined with clean images from the FFHQ dataset to synthetically generate optically encoded measurements $\mathbf{y}$. We further augmented these measurements by adding read noise

and photon shot noise calibrated to the Canon 5D sensor characteristics (ISO 100). With these measurements, we fine-tune the Null-Space Network f_θ and the detector d_ψ for 30 epochs with a learning rate of 1×10^{-5}.

10. Real-World Capture Procedure and Additional Analysis

We conducted the capture experiments in a controlled, realistic indoor lighting scenario. We displayed all test scenes on a high-resolution LCD monitor rather than using printed targets. This approach ensured high spatial fidelity, stable illumination, and rapid switching between scenes. The monitor was positioned perpendicular to the optical axis at a fixed distance of 50 cm from the camera aperture. All captures were made using the Canon EOS-5D Mark IV with the 50 mm f/1.8 YONGNUO lens and the fabricated phase mask mounted at the aperture stop. All images were captured in RAW format to avoid in-camera processing. Captured images are cropped and resized to 512×512, matching the resolution used in our simulations.

The first column of Fig. 8 shows the raw measurements $\mathbf{y}$ captured with the prototype. Passing these measurements through the fine-tuned NSN f_θ yields the protected reconstructions $\mathbf{x}_p$, shown in the second column. We can recover the scene content with high visual fidelity. To simulate realistic manipulations, we manually edited the protected images using Photoshop (lasso selection + Generative Fill). The resulting tampered images and their corresponding detection masks produced by our NOWA framework are presented in the third and fourth columns, respectively. These results confirm that the robustness observed in simulation carries over to real-world captures, demonstrating the physical-layer security provided by the NOWA. The final column visualizes the discrepancy between the NOWA of the authentic and edited images. High-frequency deviations emerge within the manipulated regions, while mild low-frequency differences appear near the edges. Still, our proposed detector reliably isolates the tampered areas.

Discussion and Handling of Depth Variations. Although our real-capture experiments use a high-resolution LCD monitor positioned at a fixed distance, this setup faithfully represents the optical behavior expected in real scenes for two reasons. First, the phase-mask prototype operates with a small effective aperture, producing a naturally large depth of field. In this regime, objects lying within a broad depth range produce sensor measurements whose point spread function (PSF) and corresponding null-space structure vary only minimally with respect to depth. Second, bsuooecause the NOWA embedding arises from the null space of the imaging operator $\mathbf{A}_\phi$, and because this operator is largely invariant within the depth-of-field limits, the NOWA behaves consistently whether the content is displayed on a monitor or originates from a real 3D environment.

For scenes with large depth variations or macro-scale imaging, the PSF becomes depth-variant. In such cases, one can extend the framework by calibrating a small set of depth-dependent PSFs or by incorporating a depth-conditioned NSN stage. Both are compatible with our current formulation and require the proper augmentations to the reconstruction pipeline. We leave the development of depth-aware calibration strategies as an interesting future direction, enabling the NOWA framework to handle scenes beyond the current depth-of-field regime.

Robustness against real-world camera imperfections The phase mask induces a deterministic, structured PSF, whereas contaminants (e.g., dust) and minor aberrations introduce stochastic optical distortions. The null-space projection acts as a matched filter, isolating the watermark's structured residuals from uncorrelated noise. However, in our real-world experiments, imperfect calibration or slight misalignment degrade null-space estimation that raises the energy baseline of the null space. Even after precise PSF calibration, the residual null-space energy $\|\Pi_{\mathcal{N}}(\hat{\mathbf{x}}_r)\|$ increases from a baseline of $\sim 10^{-5}$ (not 0 due to numerical precision limits) in simulation to $\sim 10^{-4}$ in real-world captures. While this indicates minor spectral leakage of unwanted signals, this "real-world noise floor" remains significantly lower than the residuals induced by tampering. Consequently, the detector successfully differentiates between the stochastic artifacts of physical imperfections and the specific null-space violations of a forgery.

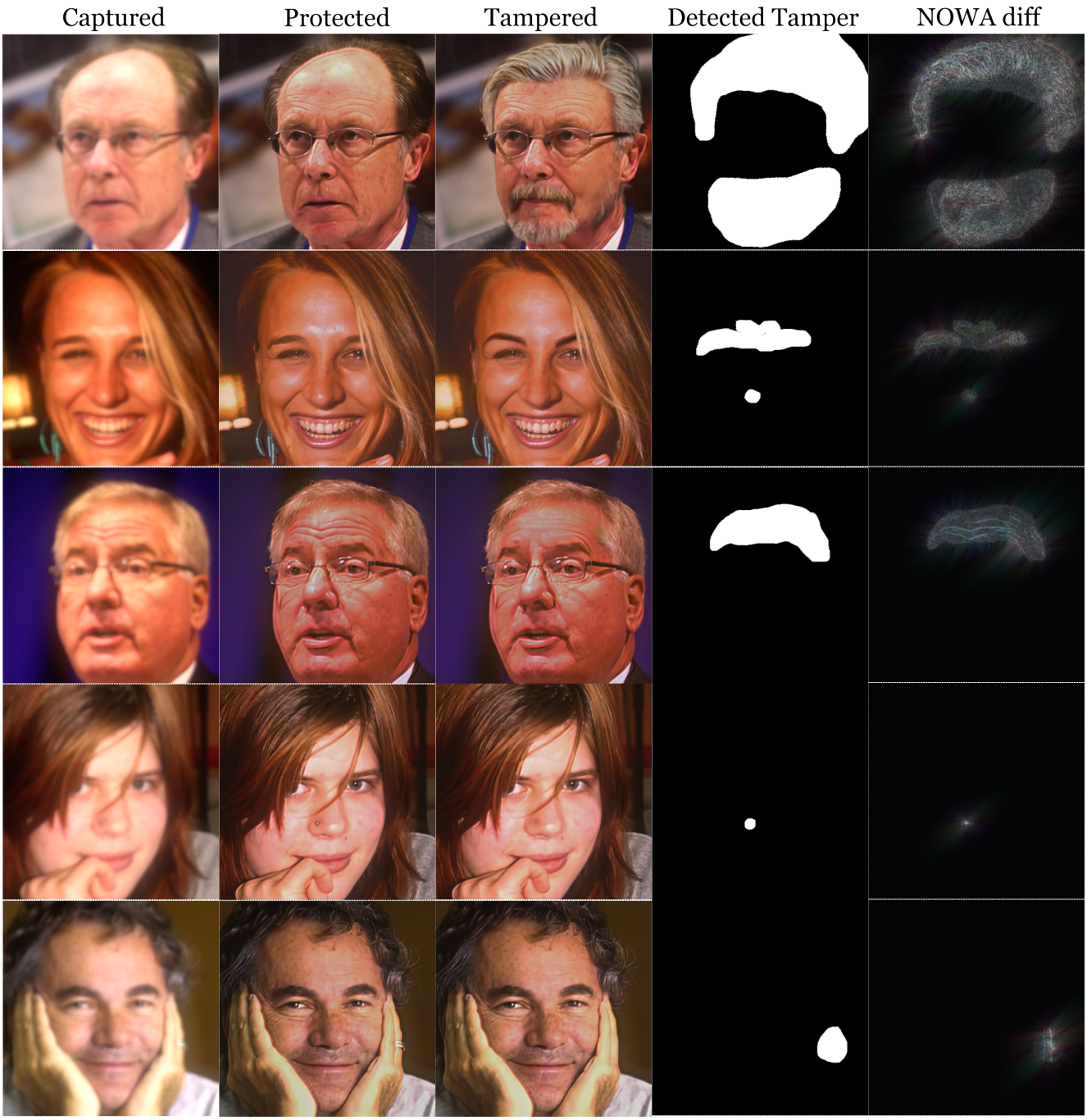

Figure 8. **Real-world prototype results.** From left to right: raw capture $\mathbf{y}$, protected reconstruction $\mathbf{x}_p$, tampered image (Photoshop Generative Fill), detected tamper mask, and NOWA discrepancy between authentic and edited captures. Manipulations introduce measurement-inconsistent NOWA residuals that are reliably localized by our detector.

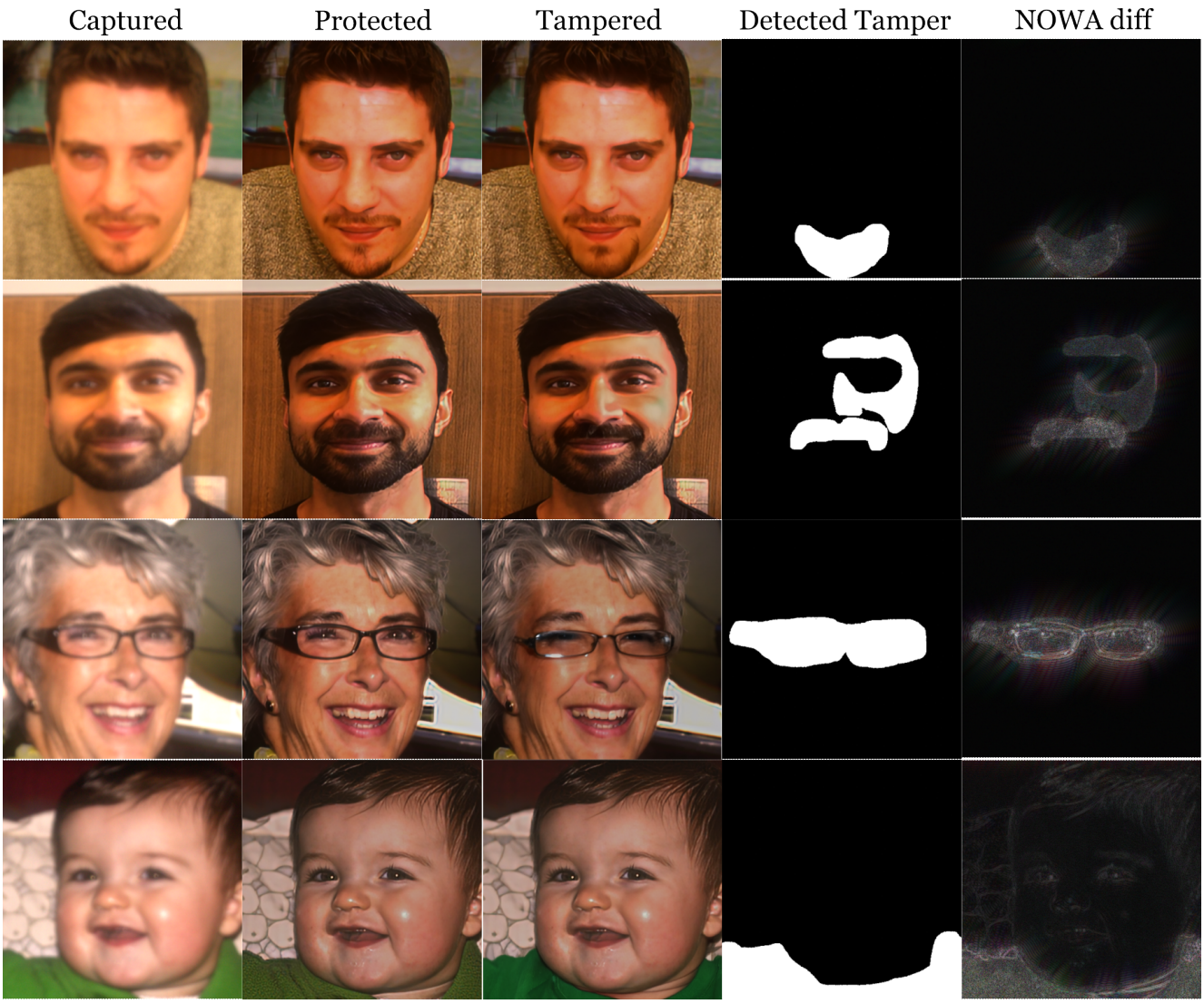

Figure 8. (cont.). **Real-world prototype results.** From left to right: raw capture $\mathbf{y}$, protected reconstruction $\mathbf{x}_p$, tampered image (Photoshop Generative Fill), detected tamper mask, and NOWA discrepancy between authentic and edited captures. Manipulations introduce measurement-inconsistent NOWA residuals that are reliably localized by our detector.