

# Physics-Aware Noise2Noise Denoising for High-Resolution Indirect Time-of-Flight Imaging

## Motivation

Indirect Time-of-Flight (iToF) cameras estimate depth from phase shifts of modulated illumination. While dense imaging architectures enable megapixel-scale sensing, low-SNR measurements suffer from shot noise, multipath effects, and system drift, which degrade phase accuracy. Self-supervised methods such as Noise2Noise [1] offer a promising strategy for improving signal quality without requiring clean ground-truth depth. However, iToF measurements are complex-valued phasors with signal-dependent noise, making direct application of image-domain denoising assumptions challenging. This project investigates physics-aware Noise2Noise training in the complex phasor domain to improve depth stability under realistic low-SNR conditions.

## Related Work

Noise2Noise [1] and related self-supervised approaches [2,8] demonstrate that denoising can be learned from independent noisy observations. Prior work has explored self-supervised depth denoising [3,4] and classical Time-of-Flight noise modeling [5], but most methods operate on depth or intensity images rather than complex correlation signals. Here, we study Noise2Noise directly in the complex iToF domain, accounting for heteroscedastic shot-noise statistics.

## Proposed Approach

We propose to develop a learning pipeline that operates directly on the complex phasor representation obtained from 4-tap iToF measurements. The key idea is to split temporal measurements into independent subsets, forming noisy pairs suitable for Noise2Noise training [1].

The pipeline will include:

- FFT-based demodulation to obtain complex phasors from raw 4-tap frames.
- Construction of independent phasor pairs via cycle splitting.
- Complex-domain denoising using Noise2Noise supervision [1].
- Conversion of denoised phasors into depth maps using phase extraction.

Experimental validation shows that even in low-SNR regimes, temporal splits capture a highly consistent underlying signal while residual–signal correlations reflect expected shot-noise behavior. This motivates physics-aware learning rather than assuming strictly additive noise.

## Goals and Expected Outcomes

Primary goals:

1. Demonstrate SNR improvement and reduced phase variance in dense iToF measurements.
2. Compare classical averaging, Noise2Noise learning [1], and raw data.
3. Evaluate how physics-aware training impacts depth precision.

We expect to produce:

- Improved depth stability in flat-target experiments.
- Quantitative analysis of phase noise and depth variance.
- Visualization of denoised complex phasor fields and reconstructed depth maps.

## Milestones and Timeline

Week 8: Baseline and Noise2Noise Training

- Finalize dataset preprocessing and complex phasor pipeline.
- Construct independent temporal splits for Noise2Noise supervision.
- Train initial complex-domain denoising model.
- Establish quantitative evaluation metrics (phase variance, depth stability, SNR).

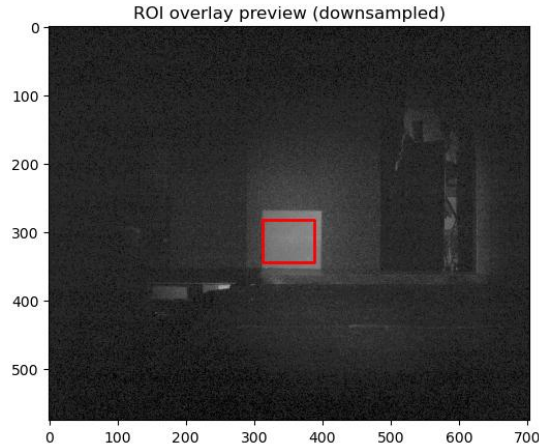
Week 9: Evaluation and Analysis

- Compare classical averaging vs. Noise2Noise denoising on low-SNR regions.
- Perform ablation experiments on ROI selection and SNR regimes.
- Analyze heteroscedastic noise behavior and training stability.
- Generate figures and prepare final report.

## Expected Deliverables

- A working complex-domain Noise2Noise pipeline.
- Quantitative evaluation of phase and depth improvements.
- Final conference-style report including theory, experiments, and analysis.

**Summary:** This project integrates computational imaging, physics-aware learning, and active depth sensing to explore scalable denoising strategies for high-resolution iToF imaging systems.



**Figure 1:** ROI selection for complex-domain Noise2Noise validation, object at 5 m

```
Using 80 frames => 20 complete 4-tap cycles
Mask coverage in ROI: 100.00%

===== RAW (no global drift correction) =====
Complex similarity (coherence-like):
rho      = +0.980320 -0.037607j
|rho|    = 0.981041 (closer to 1 => two views match well)

Phase-difference stats (from complex ratio; wrap-safe):
circular mean(mu) = -3.848675e-02 rad
circular std(sigma) = 1.479085e-01 rad (smaller is better)
resultant length (R) = 0.989121 (closer to 1 => tighter)
linear mean/std   = -3.848707e-02 rad, 1.479574e-01 rad

Depth/phase-domain check (phi_a - phi_b wrapped):
circular mean(mu) = -3.848674e-02 rad
circular std(sigma) = 1.479085e-01 rad
resultant length (R) = 0.989121

Independence / structure checks (want small magnitudes):
Corr(Re(diff), Im(diff)) = +0.000672
Corr(Re(diff), Re(sig))  = +0.701113
Corr(Im(diff), Im(sig))  = +0.336083
Corr(Re(diff), Im(sig))  = +0.042451
Corr(Im(diff), Re(sig))  = -0.000559
...
Corr(Re(diff), Re(sig))  = +0.338305
Corr(Im(diff), Im(sig))  = +0.703977
Corr(Re(diff), Im(sig))  = +0.000941
Corr(Im(diff), Re(sig))  = +0.002705
```

**Figure 2:** Experimental validation of Noise2Noise assumptions for complex iToF phasors.

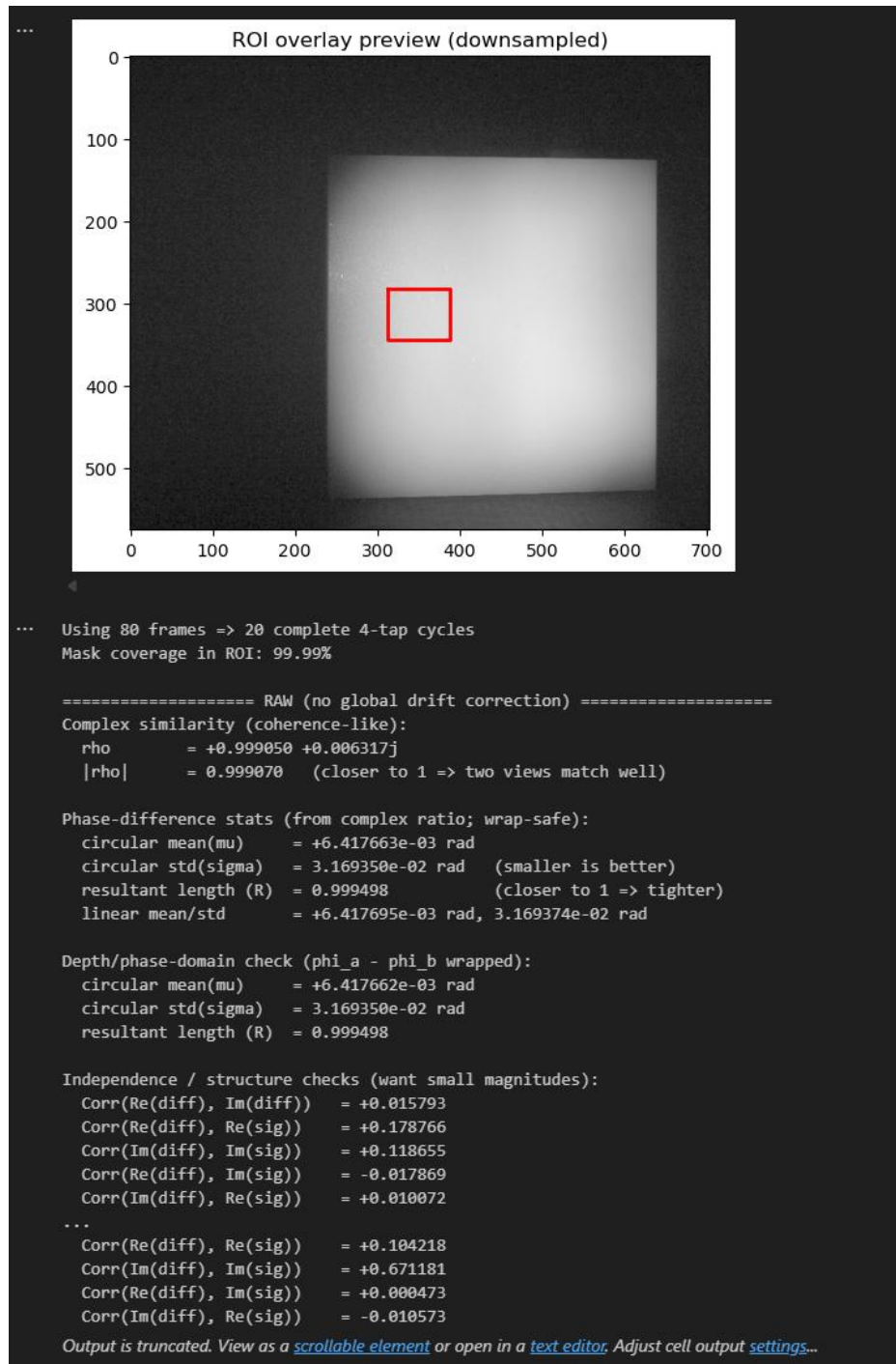
Figure 2 shows the complex-domain similarity metrics between two independent cycle splits. High coherence ( $|\rho| \approx 0.98$ ) and low circular phase variance indicate that the two temporal splits capture a highly consistent underlying signal, while residual–signal correlations reveal the heteroscedastic shot-noise statistics typical of low-SNR iToF measurements. [1].

The figure illustrates the low-SNR regime where signal-dependent noise is prominent. Although residual–signal correlations are non-negligible due to photon shot noise, the zero-mean property required by Noise2Noise remains valid. This motivates learning in the complex domain to improve depth precision under challenging conditions.

## References

- [1] J. Lehtinen, J. Munkberg, J. Hasselgren, et al.,  
“*Noise2Noise: Learning Image Restoration without Clean Data,*”  
Proc. Int. Conf. on Machine Learning (ICML), 2018.
- [2] A. Krull, T.-O. Buchholz, and F. Jug,  
“*Noise2Void – Learning Denoising from Single Noisy Images,*”  
Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.
- [3] A. Sterzentsenko, G. Drettakis, and S. Leutenegger,  
“*Self-Supervised Deep Depth Denoising,*”  
Proc. IEEE Int. Conf. on Computer Vision (ICCV), 2019.
- [4] F. Kienle, D. Cremers, and M. Moeller,  
“*Multi-Object Self-Supervised Depth Denoising,*”  
arXiv preprint arXiv:2305.05778, 2023.
- [5] S. Foix, G. Alenya, and C. Torras,  
“*Lock-In Time-of-Flight (ToF) Cameras: A Survey,*”  
IEEE Sensors Journal, 2011.
- [6] M. B. Lindell, M. O’Toole, and G. Wetzstein,  
“*Wave-Based Non-Line-of-Sight Imaging using Fast  $f$ - $k$  Migration,*”  
ACM SIGGRAPH, 2019.
- [7] S. Gong et al.,  
“*Towards Robust Time-of-Flight Depth Denoising with Diffusion Models,*”  
arXiv preprint arXiv:2503.19448, 2025.
- [8] D. Ulyanov, A. Vedaldi, and V. Lempitsky,  
“*Deep Image Prior,*”  
Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2018.

## Appendices



**Figure 3:** High-SNR capture at 1 m (validation in near-field high-photon regime).

High-SNR measurements exhibit reduced signal-dependent noise, resulting in weaker correlation between residuals and signal amplitude. This suggests that Noise2Noise assumptions are more closely approximated in the high-SNR regime.