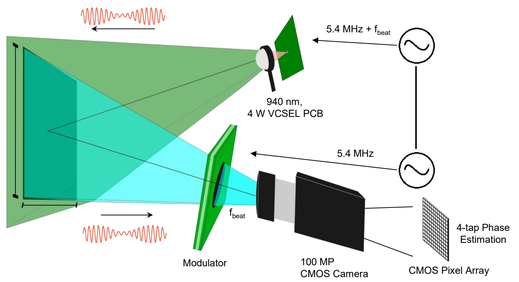
Self-Supervised Denoising of 4-Tap iToF Measurements with Phasor Consistency

Saner Halil Baskaya
EE367, Stanford University

Shot Noise Limits Precision in Indirect Time-of-Flight Sensing

- iToF estimates depth from the phase of modulated light.
- Measurements are corrupted by photon **shot noise**.
- Classical depth denoising approaches often require **clean or near-ground-truth supervision** [1], [2].
- Noise2Noise (N2N) shows that neural networks can learn denoising from **noisy-noisy training pairs** [3], [4].
- **Key Question:** Can self-supervised denoising of raw iToF taps improve phase and depth precision without clean targets?



Related Work

Class	Insight	Limitation
Temporal Averaging	Reduce noise via temporal integration	Reduces temporal resolution
Supervised Denoising	Learn noisy → clean mapping	Clean data often unavailable
Self-Supervised (Noisy-Pair)	Train using noisy observations only	Need independent captures

References

[1] Dong, Zhang & Xiong — Spatial Hierarchy Aware Residual Pyramid Network for ToF Depth Denoising, European Conference on Computer Vision (ECCV), 2020.
[2] Yan, Wang, Kao, Gong, Shen & Fu — *DDRNet: Depth Map Denoising and Refinement for Consumer Depth Cameras*, ECCV, 2018.
[3] Lehtinen, Munkberg, Hasselgren, Laine, Karras, Aittala & Aila — Noise2Noise: Learning Image Restoration without Clean Data, ICML, 2018.
[4] Krull, Buchholz & Jug — Noise2Void: Learning Denoising from Single Noisy Images, CVPR, 2019.

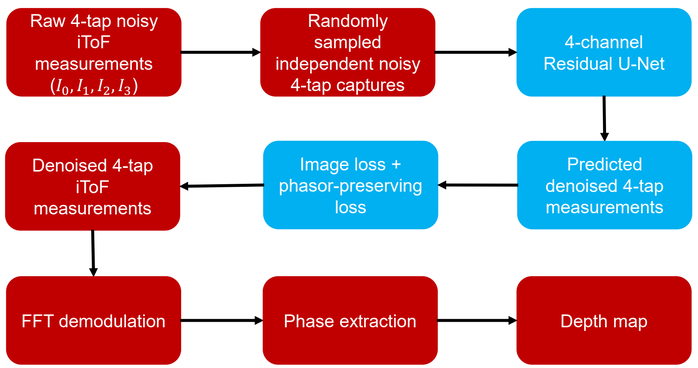
Phase-Preserving Self-Supervised Denoising

- Train a 4-channel residual U-Net to denoise raw 4-tap iToF measurements.
- Training pairs are formed by randomly sampling two independent 4-tap captures of the same static scene.
- Add a phasor-consistency loss on derived I/Q components.

$$Q = I_0 - I_2, \qquad I = I_1 - I_3,$$

- This preserves phase information for downstream depth reconstruction.

Training/Evaluation Pipeline



Training Objective

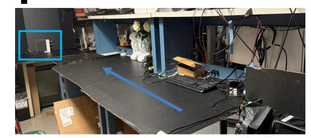
$$\mathcal{L} = \lambda_{img}\, MSE(\tilde{x}, y) + \lambda_{phasor}\left[MSE(\tilde{I}, I_y) + MSE(\tilde{Q}, Q_y)\right]$$

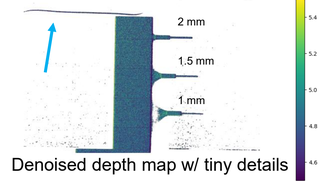
$$Q = I_0 - I_2, \qquad I = I_1 - I_3$$

- The network denoises raw taps, while the phasor term preserves downstream phase consistency.

Denoising Setup

- 100 MP captures from flat targets at 1.0 - 5.0 m.
- Training uses random 96×96 patch pairs from independent noisy observations.
- **Network:** Residual U-Net, 4 input / 4 output channels
- **Reconstruction:** FFT → phase → depth conversion.
- **Evaluation:** amplitude gating + plane-fit RMS in ROI.
- **Loss:** image MSE + phasor-consistency loss

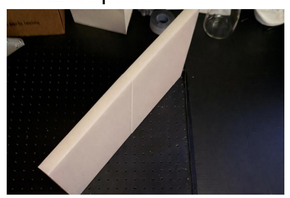

1.5 mm wire & tiny rods @ 5 m

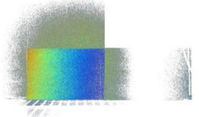


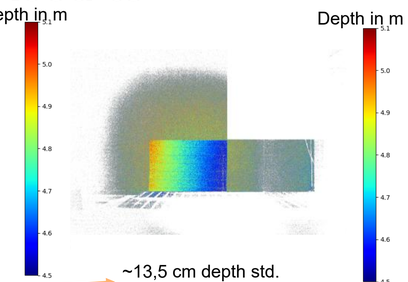
Denoised depth map w/ tiny details

Experimental Results

1 Hz Capture

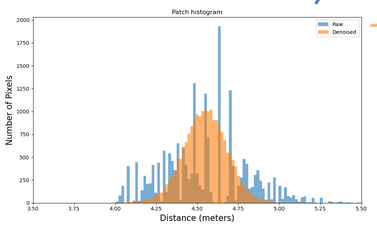

Diagonal plate @ 5 m





~25,6 cm depth std.
On region-of-interest

~13,5 cm depth std.
On region-of-interest



Distance (m)	Raw Phase Std (rad)	Denoised Phase Std (rad)	Phase Improv-ement (%)	Raw RMS (mm)	Denoised RMS (mm)	RMS Improve-ment (%)
1.0	0.0524	0.0305	41.2	207.7	66.9	67.9
1.5	0.0626	0.0286	54.0	244.2	89.5	64.0
2.0	0.0678	0.0278	59.5	254.2	96.2	62.2
2.5	0.0823	0.0360	55.8	266.5	158.9	40.3
3.5	0.1344	0.0419	68.2	325.2	185.2	42.8
5.0	0.0000*	0.0507	23.8*	277.9	209.9	18.1

Flat plates at various distances, ROI at the centers
*At 5 m, very low SNR causes unstable raw amplitude gating, so improvements should be interpreted cautiously

Discussion

- Current method still requires independent noisy captures and is evaluated on controlled flat-target scenes.
- **Key Takeaway:** Self-supervised denoising of raw 4-tap measurements improves phase stability and depth precision without requiring clean ground-truth data.