2026 IEEE International Conference on Image Processing (ICIP) | 979-8-3315-5151-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICIP61757.2026.11629902

PHOTON-STATISTICS-DRIVEN LEARNING FOR UNDERWATER IMAGING

Zaichang Lu[1,2], *Dongyu Du*[1,2], *Zhiheng Li*[2,1], *and Xin Jin*[1]*

[1]Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
[2]Tsinghua Innovation Center in Zhuhai, Zhuhai, China

ABSTRACT

Transient imaging enables improved underwater imaging by exploiting time-of-flight (ToF) information to separate signal from backscatter. However, complex spatiotemporal photon mixing in dynamic scattering media makes large-scale ToF data collection difficult and realistic simulation inaccurate, hindering data-driven learning. To address these challenges, we develop a photon-statistics-based scattering forward model that explicitly links scene reflectance to scattering measurement. Building on this model, a parameterized ToF scattering simulator is implemented that generates diverse and realistic training data across scattering conditions, effectively bridging the simulation-to-real gap. We further introduce an end-to-end reconstruction network with a likelihood-driven objective that enhances target fidelity while preserving global scene structure. Experiments on real underwater data spanning multiple turbidity levels demonstrate consistent improvements over the baselines, improving PSNR by 42% and SSIM by 74%, and reliable reconstructions even at 10.864 TMFPs (21.728 TMFPs round trip). Together, this work connects physical photon transport with learning-based inference, paving the way for practical underwater transient imaging under severe scattering.

***Index Terms*—** Underwater imaging, Single photon imaging, Scattering imaging, Deep learning

1. INTRODUCTION

Underwater optical imaging is crucial for marine exploration, environmental monitoring, infrastructure inspection, and emergency rescue[1, 2]. However, suspended particles and dissolved substances in turbid water induce strong scattering and absorption, which both attenuate target-return photons and generate dominant backscatter, severely degrading imaging range and quality[3].

Transient imaging has been widely explored to improve visibility in scattering media by leveraging photon timing. In turbid water, ballistic (or weakly scattered) photons typically arrive earlier and yield sharper transients than multiply scattered background, enabling partial time-domain separation. This supports practical strategies such as time-gating[4, 5, 6], cross-correlation with a calibrated instrument response[7, 8], and statistical fitting of photon-count histograms under scattering-aware models[9, 10, 11]. As turbidity increases, temporal broadening and multipath mixing erode these signatures, causing target and backscatter responses to overlap. Physics-based methods[12, 13, 14] address this by modeling photon propagation and solving the resulting inverse problem, but forward models are often simplified (e.g., isotropic scattering or reduced dynamics) and thus struggle to capture complex multipath effects; even more refined formulations[15] can be sensitive to parameter tuning, limiting robustness in highly turbid conditions.

Learning-based methods[16, 17, 18], especially CNN [19]and attention-based [20] architectures with spatiotemporal fusion, perform well in photon-limited imaging for low light, long range, and atmospheric scattering by learning spatiotemporal mappings from photon counts to scene estimates. However, these methods transfer poorly to underwater imaging due to the lack of statistical constraints tailored to underwater photon migration. Moreover, the difficulty of reproducing and accurately modeling underwater scattering limits scalable data collection and realistic simulation, reducing supervision quality and hindering generalization.

To overcome these limitations, we propose a statistical forward model for turbid water that accurately captures the spatiotemporal photon distribution in scattering media. Built upon this model, a dedicated simulation pipeline is developed to generate scalable, high-fidelity synthetic training data, reducing reliance on costly real-world acquisition while preserving strong generalization. Using these measurements, we further design an end-to-end network for target intensity reconstruction as a first step toward the more challenging joint recovery of intensity and depth under severe scattering. Extensive experiments show that our approach significantly outperforms the baselines, improving PSNR by 42% and SSIM by 74%, particularly under extreme scattering, and extends the practical imaging limit to 10.864 TMFPs.

The remainder of this paper is organized as follows. Section 2 presents the proposed forward model and the network architecture. Section 3 reports the hardware prototype, implementation details, reconstruction limits and different objects recovery effects. Section 4 concludes the paper.

*Corresponding author.

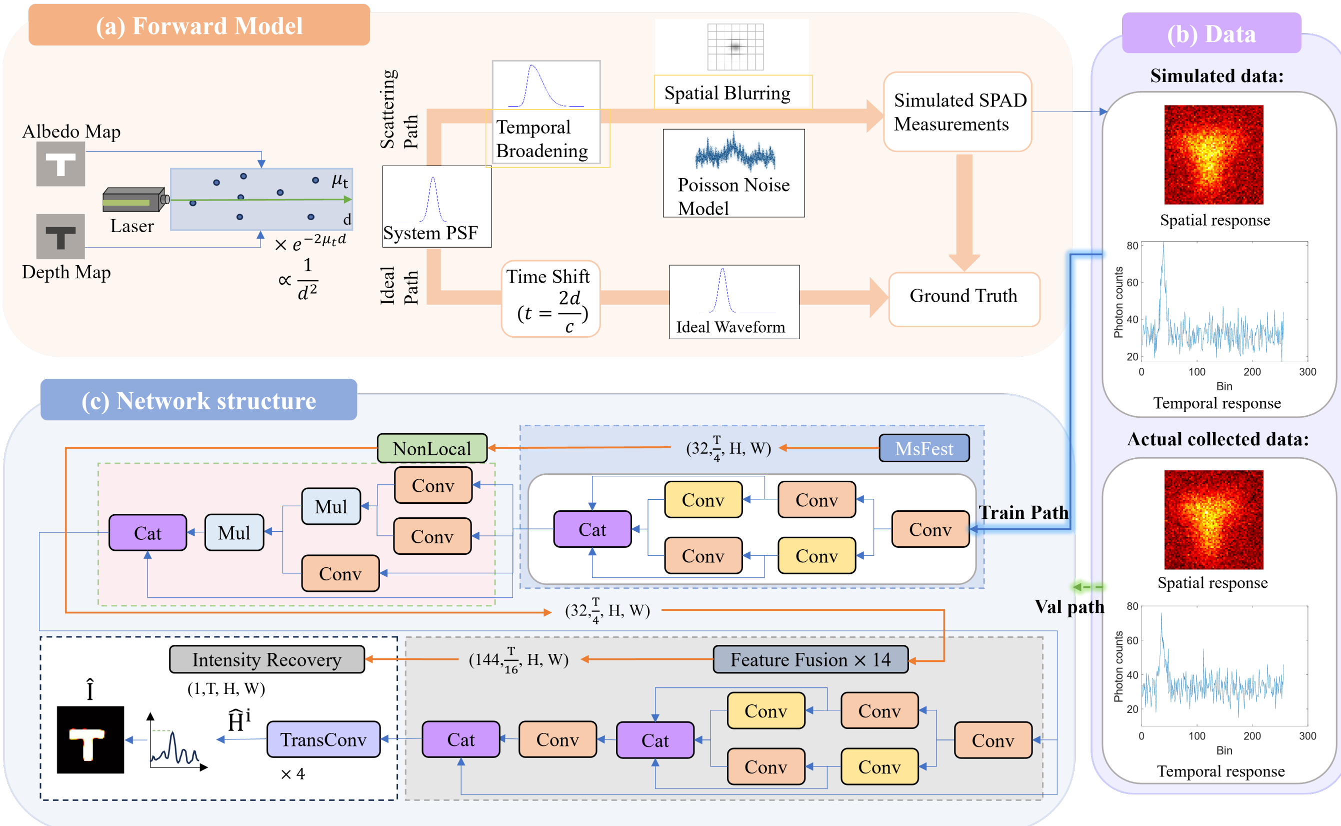

Fig. 1. **Overview of the proposed pipeline.** (a) A statistical forward model that maps depth and albedo maps to SPAD measurements by incorporating system PSF, scattering-induced temporal broadening and spatial blurring, and Poisson photon-counting noise, enabling controllable simulation. (b) Temporal and spatial comparison between simulated and actual collected data. (c) Reconstruction network for underwater intensity recovery from measurements. "Conv" denotes convolution including dilated convolutions, "Cat" feature concatenation, "Mul" matrix multiplication, and "TransConv" 3D transposed convolution. The network outputs a 2D intensity image.

2. METHOD

Our methodology combines statistical modeling and deep learning to tackle severe scattering in turbid water (Fig. 1). We first formulate a statistical forward model that captures the spatiotemporal photon distribution in scattering media, and leverage it to build a physics-consistent simulator for generating controllable synthetic training data. Based on these data, we train an end-to-end reconstruction network to recover target intensity from distorted measurements. The trained model is then validated on real underwater captures, demonstrating effective sim-to-real transfer and reliable reconstructions under high turbidity.

2.1. Forward model

We consider a SPAD-based underwater ToF imaging system, as shown in Fig. 2. A pulsed laser illuminates the scene, and photons propagate through turbid water where they undergo absorption and multiple scattering. The SPAD array detects both backscatter and target returns. The arrival time are recorded by a time-correlated single photon counting module(TCSPC) to form per-pixel temporal histograms, which serve as the measurements for subsequent reconstruction.

To accurately model underwater photon propagation and generate high-fidelity training dataset, we develop a spatiotemporal degradation model. Unlike atmospheric models, it explicitly captures the temporal broadening and spatial blurring characteristic of underwater scattering media, enabling synthesis of realistic ToF measurements. In photon-starved regimes, the observed photon flux $\psi(t;x,y)$ at pixel location (x,y) and time delay t impulse response as:

$$\psi(t;x,y) = \iint_{x',y'\in\mathrm{FoV}} g_{xy}(\Delta x,\Delta y;d')\cdot\left[r(x',y')\cdot\beta(d')\right]\cdot \left[s\left(t-\frac{2d'}{c}+\delta_\tau\right)*h_\tau(t;d')\right]\mathrm{d}x'\,\mathrm{d}y' + b_\gamma \tag{1}$$

where (x',y') indexes scene points within the receiver field

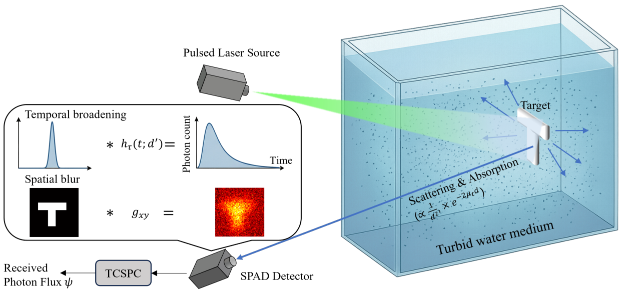

Fig. 2. **Illustration of the forward model for single photon imaging in turbid water.** The system initiates with a pulsed laser emitting a Gaussian pulse $s(t)$ toward a target, where synergistic forward and backward scattering drive signal degradation following the attenuation model $\propto \frac{1}{d^2} e^{-2\mu_t d}$. This physical process induces characteristic temporal tailing in the received flux ψ_t, simulated using a Gamma function to model multi-path delay, alongside scattering-dependent Gaussian blurring g_{xy}. The spatiotemporal statistics captured via a SPAD based TCSPC module.

of view (FoV), $d' = d(x', y')$ denotes scene depth, and $\Delta x = x - x'$, $\Delta y = y - y'$. c is the speed of light. $r(x', y')$ is the intrinsic surface reflectance. $s(t)$ denotes the transmitted laser pulse profile, $*$ indicates convolution over time, and b_γ accounts for ambient background photons. The remaining terms $\beta(d')$, $h_\tau(t; d')$, $g_{xy}(\Delta x, \Delta y; d')$, and δ_τ respectively model attenuation, temporal broadening, spatial blur, and timing jitter induced by underwater scattering, as detailed below.

Distance attenuation. We capture distance-dependent signal loss by $\beta(d') = \frac{1}{d'^2} \exp(-2\kappa d')$, which combines geometric spreading with Beer–Lambert attenuation, where κ is the medium-dependent attenuation coefficient.

Temporal broadening and fluctuations. We model multipath-induced temporal spreading using a Gamma kernel

$$h_\tau(t; d') = \frac{t^{\alpha-1} e^{-t/\theta(d')}}{\Gamma(\alpha)\, \theta^\alpha(d')} \tag{2}$$

where α controls the effective scattering order, $\theta(d')$ is a distance-dependent scale parameter, and $\Gamma(\cdot)$ denotes the Gamma function. As d' increases, $\theta(d')$ grows to capture pulse broadening and long-tailed transients.

To account for temporal fluctuations caused by particle Brownian motion and refractive-index inhomogeneities, we introduce a random time offset $\delta_\tau \sim \mathcal{N}(0, \sigma_j^2(d'))$. The depth-dependent variance $\sigma_j(d')$ increases with distance to emulate the heightened range uncertainty inherent in media.

Spatial blur. We model scattering-induced spatial glow with a depth-dependent Gaussian kernel

$$g_{xy}(\Delta x, \Delta y; d') = \frac{1}{2\pi\sigma^2(d')} \exp\left(-\frac{\Delta x^2 + \Delta y^2}{2\sigma^2(d')}\right) \tag{3}$$

with $\sigma(d') = \gamma d' + \sigma_0$, where γ characterizes scattering strength and σ_0 accounts for the system's intrinsic blur.

Following conventional SPAD detection models[19, 18], we characterize the photon counting process as an inhomogeneous Poisson process. The instantaneous detection rate $\Psi(t; x, y) = \eta\psi(t; x, y) + b_d$ accounts for the quantum efficiency η and dark count rate b_d. By integrating Ψ over time bins Δt and accumulating over N pulse periods, the final observed histogram $H_{x,y}[n]$ is modeled as:

$$H_{x,y}[n] \sim \text{Poisson}\left(N \int_{n\Delta t}^{(n+1)\Delta t} \Psi(\tau; x, y)\, d\tau\right) \tag{4}$$

As shown in Fig. 1(b), histograms synthesized by our forward model closely match real captures in both transient shape and noise characteristics, with a JS divergence of 0.0107 and an MSE of 0.0145, demonstrating high temporal and spatial consistency.

2.2. Reconstruction neural network

Given SPAD photon-count histograms, we use an end-to-end neural network to reconstruct a 2D intensity image under strong backscatter, low SNR, and severe spatiotemporal mixing.

As illustrated in Fig. 1(c), we adopt an encoder–decoder reconstruction network[18] that maps the spatiotemporal photon-count distribution to a 2D target-intensity estimate. The encoder extracts compact spatiotemporal features with stacked convolutional blocks and augments them with a non-local module to capture long-range dependencies. The resulting features are then refined by a multi-scale feature extraction stage (MsFest) to enhance robustness under severe scattering. Finally, the decoder upsamples and fuses features via transposed convolutions and feature-fusion blocks, and yields the intensity estimate with an intensity head.

In turbid water, depth cues are often obscured and backscatter is highly non-stationary, making depth-centric or peak-based reconstruction unreliable. We therefore tailor DeepBoosting [18] into an intensity-centric reconstruction network. We aggregate the network 3D output along the temporal dimension and construct a joint loss that combines pixel-wise reconstruction and structural similarity. For each pixel (i, j), the network predicts a temporal response vector $\mathbf{V}_{i,j} \in R^L$. Specifically, we apply a softmax normalization along the temporal dimension and define the reconstructed intensity as the maximum response:

$$\hat{I}(i, j) = \max_{\ell}\Big(\text{softmax}(\mathbf{V}i, j)\ell\Big) \tag{5}$$

where $\hat{I} \in \mathbb{R}^{H \times W}$ denotes the predicted intensity image and $I \in \mathbb{R}^{H \times W}$ is the ground-truth intensity.

In underwater scenes, target regions typically occupy only a small fraction of the image, while background backscatter

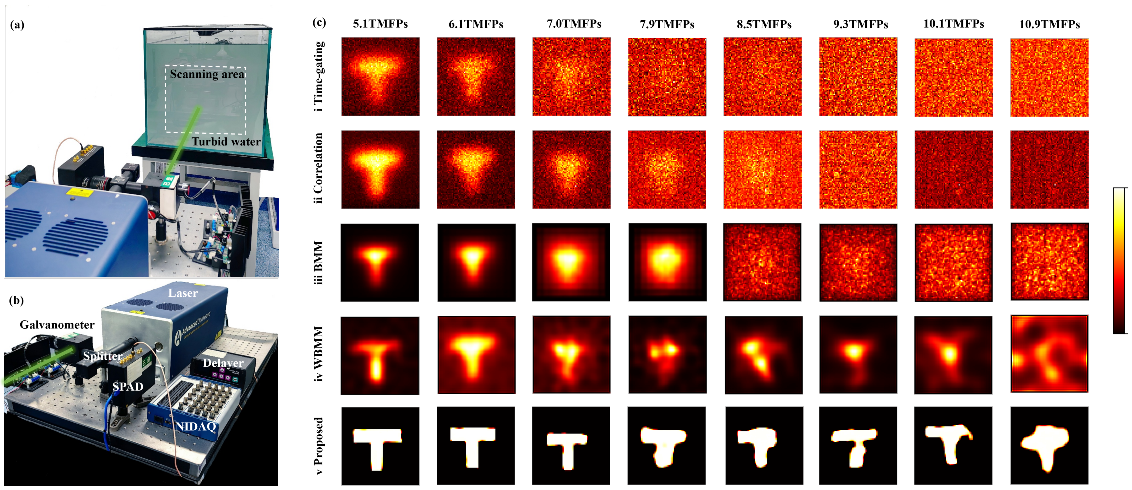

Fig. 3. **Experimental setup and imaging limit evaluation in turbid water.** (a)Experimental environment. (b)Confocal imaging system. (c)Performance comparison of *Time-gating*, *Correlation*, *BMM*, *WBMM*, and *Proposed* under increasing scattering (TMFPs): *Proposed* delivers the most faithful reconstructions and maintains reliable recovery at the strongest scattering regime, indicating the best imaging limit among all methods.

dominates most pixels. A uniform pixel-wise loss is therefore suboptimal, as it biases learning toward background regions. To address this imbalance, we introduce a binary mask $M \in \{0,1\}^{H\times W}$, where $M(i,j) = 1$ indicates target pixels, and define a weighted ℓ_1 loss:

$$\mathcal{L}_{\ell_1}(I,\hat{I}) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\big(1+\alpha\, M(i,j)\big)\,\big|I(i,j)-\hat{I}(i,j)\big| \tag{6}$$

where $\alpha > 0$ emphasizes the target region. To preserve global contrast and structural consistency, we further add an SSIM term computed over the full image:

$$\mathcal{L}_{\mathrm{SSIM}}(I,\hat{I}) = 1 - \mathrm{SSIM}(I,\hat{I}) \tag{7}$$

The total loss for training the intensity branch is given by the combination of the above terms.

3. RESULTS

3.1. Hardware prototype

We acquire underwater measurements using a confocal volumetric scattering imaging system, as shown in Fig. 3(a,b). A 532-nm pulsed laser provides active illumination, and a coaxial transmit–receive path (via a polarizing beam splitter) aligns illumination and detection. The scene is raster-scanned by a galvanometer synchronized with an NI-DAQ; at each scan position, the SPAD array records a TCSPC temporal histogram, and stacking all histograms forms a $64 \times 64 \times 256$ spatiotemporal observation cube. Experiments are conducted in a $0.5 \times 0.5 \times 1$ m water tank with Maalox[8] to control turbidity under fully submerged targets. The laser (INNO AMT-532-1W1M; $\sim$12 ps pulse width, 20 MHz repetition rate, $\sim$800 mW) is coupled into the coaxial path, and returns are imaged (2.8–12 mm, $f/1.6$) onto a Photon Force PF32 SPAD array with an effective temporal resolution of $\sim$55 ps; a delay module converts the laser sync to a TTL trigger for timing synchronization.

3.2. Implementation details

Dataset. We train the model on a self-designed synthetic dataset with 40 scenes covering diverse object arrangements (single/multiple objects) and depth configurations (single/multi-depth). The scattering strength is uniformly sampled from 4 to 11 TMFPs, yielding 13,200 simulated training pairs, which provide large-scale physics-consistent supervision for generalization. For real-world adaptation, we collect 215 Maalox-based turbid water tank measurement pairs. We fine-tune the network on 4/5 of the real pairs, and reserve the remaining 1/5 for validation.

Baselines and evaluation metrics. We compare our method with four representative baselines: time gating[4], cross-correlation[7], BMM[13], and UBMM[15]. All methods are applied to the same measured SPAD transient data for reconstruction. For quantitative evaluation, we compute PSNR and SSIM between the reconstructed intensity images and the reference images. The reference images are captured using the same imaging system without the scattering medium and then binarized to match the target representation.

Training Details. The input is a $64 \times 64 \times 256$ tensor with batch size 4. We use Adam with an initial learning rate of 1×10^{-4} and train for 50 epochs on an NVIDIA A40 GPU.

Table 1. Quantitative Comparison Under Different Scattering Levels

Case	Time-gating[4]		Correlation[7]		BMM[13]		WBMM[15]		Proposed	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
5.1TMFPs	11.4172	0.5529	11.2736	0.4160	11.8532	0.5640	12.2386	0.6388	**23.5451**	**0.9832**
6.1TMFPs	10.9557	0.4448	11.0023	0.3480	12.8316	0.6812	11.7001	0.5781	**16.6008**	**0.9101**
7.0TMFPs	8.3145	0.1577	9.9207	0.1293	12.0166	0.5840	12.4245	0.5742	**21.0356**	**0.9602**
7.9TMFPs	7.0697	0.0940	8.0474	0.1211	11.5393	0.5772	11.1610	0.4519	**14.1976**	**0.8329**
8.5TMFPs	7.6938	0.0890	7.9103	0.1347	9.4337	0.2026	12.4490	0.5306	**14.6645**	**0.8524**
9.3TMFPs	7.6280	0.0703	8.2842	0.1232	9.7242	0.2302	11.9683	0.5222	**15.1526**	**0.8598**
10.1TMFPs	6.7287	0.0013	6.8937	0.1002	8.1530	0.1216	11.2984	0.4945	**13.9669**	**0.8231**
10.9TMFPs	6.2947	0.0011	6.1493	0.0528	7.9856	0.1298	8.9217	0.1898	**11.3626**	**0.7012**
Avg.	8.2628	0.1764	8.6852	0.1782	10.4421	0.3863	11.5202	0.4975	**16.3157**	**0.8654**

3.3. Imaging limit evaluation

Compared with a wide range of baselines (Fig. 3), our approach achieves a clear improvement in imaging limit. As scattering increases, most baselines degrade rapidly: backscatter dominates, details vanish, and artifacts overwhelm the target, eventually yielding unrecognizable reconstructions under severe turbidity. In contrast, our method remains effective at higher scattering levels and enables reliable intensity recovery beyond the operating range of competing approaches, remaining effective up to 10.864 TMFPs. It also delivers consistently higher PSNR and SSIM than all baselines at the same scattering strength (Table 1). These gains reflect lower reconstruction error and better structural fidelity, confirming the robustness of our framework under extreme scattering and low-SNR conditions.

3.4. Different objects recovery effects

We further evaluate our method on a diverse set of objects to assess its generalization capability. As shown in Fig. 4, the proposed approach consistently produces high-quality reconstructions across different object shapes, including both simple geometric targets and objects with more complex surface textures. Notably, it remains effective in challenging multi-depth scenarios, where returns from different depths are strongly mixed due to scattering and backscatter. Despite these additional ambiguities, our method preserves the overall object integrity and recovers clear structural details with reduced artifacts, demonstrating robust performance across varying object configurations and scene complexity.

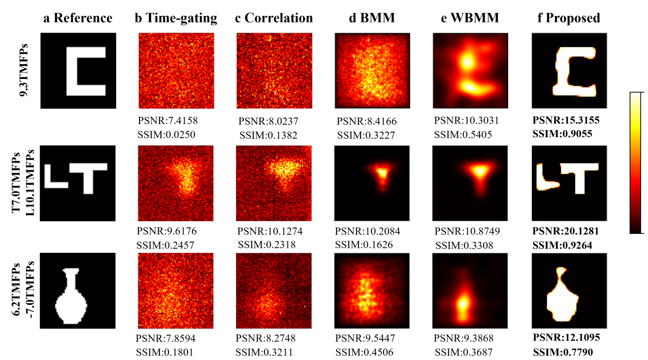

Fig. 4. **Baseline comparison on real measurements across objects and scattering strengths.** PSNR/SSIM are reported for *Time-gating*, *Correlation,*, *BMM*, *WBMM*, and *Proposed*; the best result in each setting is highlighted in **bold**.

4. CONCLUSION

We propose a unified statistical-to-learning framework for time-resolved single-photon imaging in turbid water. A photon-statistics-based forward model and ToF simulator bridge the sim to real gap, while a likelihood-driven end-to-end network reconstructs target intensity from scattering measurements. With simulation pretraining and limited real-data fine-tuning, our method improves PSNR by 42% and SSIM by 74%, and achieves reliable reconstruction up to 10.864 TMFPs one-way (21.728 TMFPs round-trip). Future work will extend the approach to underwater 3D reconstruction and focus on speeding up inference to support real-time use, which could broaden its impact in autonomous navigation, infrastructure inspection, and scientific exploration under severe turbidity.

5. ACKNOWLEDGEMENT

This work was supported in part by NSF of Guangdong Province under Grant 2023A1515012716 and Shenzhen Science and Technology Program under Grant KCXFZ20240903 094301003.

6. REFERENCES

[1] Zao Xu and Dick KP Yue, "Analytical solution of beam spread function for ocean light radiative transfer," *Optics express*, vol. 23, no. 14, pp. 17966–17978, 2015.

[2] Yanyun Pu, Chengyuan Zhu, Gongxin Yao, Chao Li, Yu Pan, Kaixiang Yang, and Qinmin Yang, "Computational imaging based on single-photon detection: a survey," *Artificial Intelligence Review*, vol. 58, no. 8, pp. 251, 2025.

[3] Ying Shen, Chuanjiang Zhao, Yu Liu, Shu Wang, and Feng Huang, "Underwater optical imaging: Key technologies and applications review," *IEEE Access*, vol. 9, pp. 85500–85514, 2021.

[4] Martin Laurenzis, Frank Christnacher, and David Monnin, "Long-range three-dimensional active imaging with superresolution depth mapping," *Optics letters*, vol. 32, no. 21, pp. 3146–3148, 2007.

[5] Petter Risholm, Jostein Thorstensen, Jens T Thielemann, Kristin Kaspersen, Jon Tschudi, Chris Yates, Chris Softley, Igor Abrosimov, Jonathan Alexander, and Karl Henrik Haugholt, "Real-time super-resolved 3d in turbid water using a fast range-gated cmos camera," *Applied optics*, vol. 57, no. 14, pp. 3927–3937, 2018.

[6] Patrizio Mariani, Iñaki Quincoces, Karl H Haugholt, Yves Chardard, Andre W Visser, Chris Yates, Giuliano Piccinno, Giancarlo Reali, Petter Risholm, and Jens T Thielemann, "Range-gated imaging system for underwater monitoring in ocean environment," *Sustainability*, vol. 11, no. 1, pp. 162, 2018.

[7] Aurora Maccarone, Aongus McCarthy, Ximing Ren, Ryan E Warburton, Andy M Wallace, James Moffat, Yvan Petillot, and Gerald S Buller, "Underwater depth imaging using time-correlated single-photon counting," *Optics express*, vol. 23, no. 26, pp. 33911–33926, 2015.

[8] Aurora Maccarone, Francesco Mattioli Della Rocca, Aongus McCarthy, Robert Henderson, and Gerald S Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," *Optics express*, vol. 27, no. 20, pp. 28437–28456, 2019.

[9] Guy Satat, Barmak Heshmat, Dan Raviv, and Ramesh Raskar, "All photons imaging through volumetric scattering," *Scientific reports*, vol. 6, no. 1, pp. 33946, 2016.

[10] Guy Satat, Matthew Tancik, and Ramesh Raskar, "Towards photography through realistic fog," in *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–10.

[11] Stephanie Maruca, Patrick Rehain, Yong Meng Sua, Shenyu Zhu, and Yuping Huang, "Non-invasive single photon imaging through strongly scattering media," *Optics Express*, vol. 29, no. 7, pp. 9981–9990, 2021.

[12] David B Lindell and Gordon Wetzstein, "Three-dimensional imaging through scattering media based on confocal diffuse tomography," *Nature communications*, vol. 11, no. 1, pp. 4517, 2020.

[13] Dongyu Du, Xin Jin, Rujia Deng, Jinshi Kang, Hongkun Cao, Yihui Fan, Zhiheng Li, Haoqian Wang, Xiangyang Ji, and Jingyan Song, "A boundary migration model for imaging within volumetric scattering media," *Nature Communications*, vol. 13, no. 1, pp. 3234, 2022.

[14] Dongyu Du, Xin Jin, and Rujia Deng, "Non-confocal 3d reconstruction in volumetric scattering scenario," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 732–744, 2023.

[15] Xin Jin, Dongyu Du, Jiawei Jin, and Yihui Fan, "Time-of-flight based imaging in strong scattering underwater environments," *Optics Express*, vol. 32, no. 21, pp. 37247–37259, 2024.

[16] Jiayong Peng, Zhiwei Xiong, Xin Huang, Zheng-Ping Li, Dong Liu, and Feihu Xu, "Photon-efficient 3d imaging with a non-local neural network," in *European Conference on Computer Vision*. Springer, 2020, pp. 225–241.

[17] Zhenya Zang, Dong Xiao, and David Day-Uei Li, "Non-fusion time-resolved depth image reconstruction using a highly efficient neural network architecture," *Optics Express*, vol. 29, no. 13, pp. 19278–19291, 2021.

[18] Jiayong Peng, Zhiwei Xiong, Hao Tan, Xin Huang, Zheng-Ping Li, and Feihu Xu, "Boosting photon-efficient image reconstruction with a unified deep neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4180–4197, 2022.

[19] David B Lindell, Matthew O'Toole, and Gordon Wetzstein, "Single-photon 3d imaging with deep sensor fusion.," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 113, 2018.

[20] Dominik Scheuble, Hanno Holzhüter, Steven Peters, Mario Bijelic, and Felix Heide, "Lidar waveforms are worth 40x128x33 words," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 28913–28924.