SCIENCE CHINA
Information Sciences

• LETTER •

August 2026, Vol. 69, Iss. 8, 189305:1–189305:2
https://doi.org/10.1007/s11432-025-4861-5

Foggy depth estimation via Mamba2-based single-photon Lidar imaging

Yanyun PU[1†], Chengyuan ZHU[1†], Chao LI[1], Kaixiang YANG[2*], Qinmin YANG[1*] & C.L. Philip CHEN[2]

[1]*College of Control Science and Engineering, Zhejiang University, Hangzhou 310000, China*
[2]*College of Computer Science and Engineering, South China University of Technology, Guangzhou 510000, China*

Received 15 July 2025/Revised 18 December 2025/Accepted 11 March 2026/Published online 8 July 2026

Citation Pu Y Y, Zhu C Y, Li C, et al. Foggy depth estimation via Mamba2-based single-photon Lidar imaging. Sci China Inf Sci, 2026, 69(8): 189305, https://doi.org/10.1007/s11432-025-4861-5

Single-photon imaging is a photon-counting computational imaging paradigm that forms images or depth maps from extremely sparse return photons [1]. By modeling Poisson photon arrivals, it enables sensing in photon-starved regimes where conventional intensity imaging or high-flux LiDAR is ineffective. However, fog and haze remain challenging because atmospheric scattering attenuates the target signal while introducing path-dependent backscatter and background photons, distorting the temporal photon histogram and driving the signal-to-background ratio to an extreme level [2]. The resulting physical degradation and sparsity can bias depth estimates, induce structural artifacts, and destabilize standard denoising or model-based inversion pipelines. To address this robustness bottleneck, our Mamba-Unet-Depth work showed that a Mamba-based reconstruction architecture can exploit long-range dependencies in sparse photon measurements and recover accurate images under extremely low SBR [3].

Existing photon-counting reconstruction methods in scattering media typically rely on simplified histogram peak picking, maximum-likelihood fitting with handcrafted priors, or variational formulations with regularizers such as total variation [4]. These pipelines are effective when the return signal is well separated, but degrade in fog and haze, where backscatter and ambient photons reshape the histogram and weaken the target peak, making inversion ill-conditioned and sensitive to calibration, thresholding, and initialization. Recent deep networks improve robustness by learning data-driven priors, yet CNN-based designs remain limited by local receptive fields and struggle to capture long-range temporal dependencies in sparse histograms, while Transformer-style attention is often compute- and memory-intensive for long sequences and high-resolution maps, limiting practical deployment and sometimes producing overly smooth reconstructions under extreme signal-to-background ratios [5].

Motivated by these gaps, we leverage the Mamba architecture for efficient long-context modeling with linear complexity, enabling the network to aggregate weak temporal and spatial evidence without quadratic attention costs. We propose a fog-aware photon-efficient reconstruction approach, Fog-MaNet, which couples a Mamba-driven temporal encoder with a U-Net-style spatial decoder to jointly recover depth and reflectivity from extremely noisy photon histograms. Each per-pixel histogram is treated as a sequence, and multi-scale Mamba blocks are used to capture global temporal structure and suppress backscatter-induced ambiguity. The decoded features are then fused through skip connections to preserve fine spatial edges. To enforce scattering-aware physical consistency, the model is trained with a Poisson likelihood objective augmented by fog-specific constraints that penalize depth bias and structure collapse under heavy backscatter, yielding stable and accurate reconstructions in fog.

FogMaNet architecture. FogMaNet is an end-to-end deep architecture for foggy single-photon LiDAR depth estimation, built in an encoder-decoder paradigm to process per-pixel photon-count histograms as a spatial-temporal feature volume, as shown in Figure 1(a). Starting from the raw photon-count tensor, the network applies an initial temporal aggregation and normalization to stabilize extremely sparse measurements under strong background and back-scatter. The encoder then performs multi-scale feature extraction and long-context spatial-temporal modeling to progressively strengthen semantic representations while preserving geometric detail via skip connections. The decoder upsamples and fuses multi-level features to recover dense depth maps, explicitly targeting the coupled degradations induced by fog, namely signal attenuation, foggy back-scatter, and severe noise sparsity. This design enables robust separation of meaningful return structures from fog-induced noise, leading to stable and accurate depth reconstructions.

Within FogMaNet, DDFS (dense dilated fusion strategy) is placed at the early encoding stage to capture both fine local details and large-scale context without changing the spatial-temporal resolution; it combines standard 3D convolutions and dilated 3D convolutions through dense concatenation, followed by a $1\times1\times1$ compression and residual fusion to suppress isolated noisy counts under strong background, as shown in Figure 1(b). AGSC (adaptive gated spatial convolution) is inserted before state-space sequence modeling to retain explicit local spatial structure that would oth-

* Corresponding author (email: yangkx@scut.edu.cn, qmyang@zju.edu.cn)
† These authors contributed equally to this work.

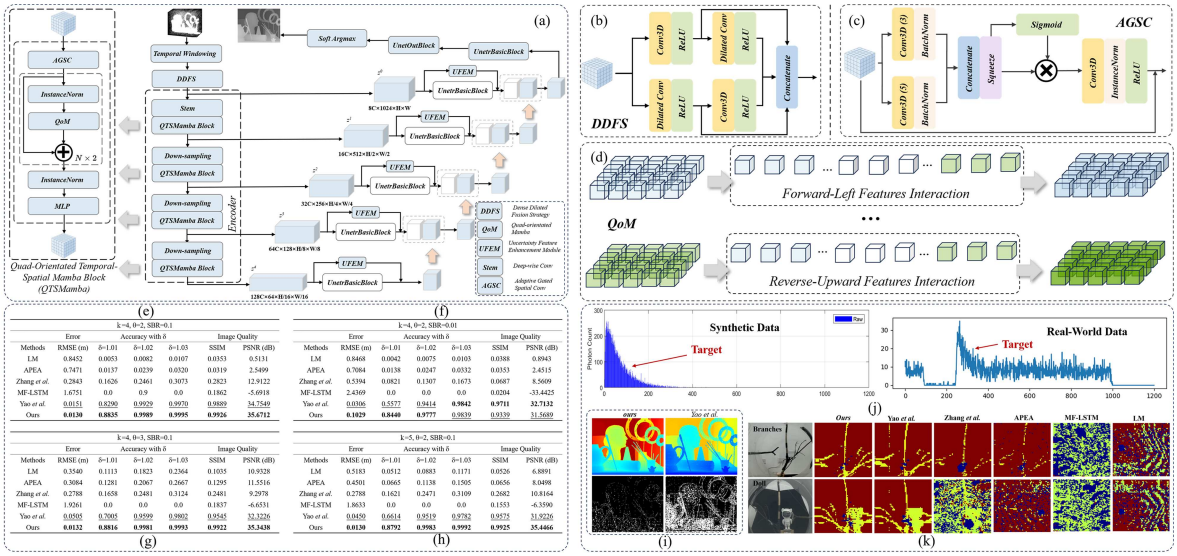

Figure 1 (Color online) (a) An overview of the proposed FogMaNet; (b) DDFS; (c) AGSC; (d) QoM; comparison of methods under four simulated fog conditions on the Middlebury datasets, bold and underlined values indicate the best and second-best performance, respectively: (e) $k = 4$, $\theta = 2$, SBR $= 0.1$, (f) $k = 4$, $\theta = 2$, SBR $= 0.01$, (g) $k = 4$, $\theta = 3$, SBR $= 0.1$, (h) $k = 5$, $\theta = 2$, SBR $= 0.1$; (i) comparison of predicted depth maps and binary error maps for the Middlebury Art scene under simulated fog (Gamma distribution with $k = 4$, $\theta = 2$, SBR $= 0.1$); (j) synthetic data and real-world data; (k) comparison of reconstructed depth maps on real-world scenes.

erwise be weakened by flattening; it extracts multi-scale spatial cues using parallel depth-wise 3D convolutions and generates a sigmoid-gated spatial weight map to emphasize informative regions and suppress fog-induced artifacts, with a residual pathway for stable training, as shown in Figure 1(c). QTSMamba is the core spatial-temporal modeling block: it reshapes the 3D feature volume into sequences along four orientations in Figure 1(d), applies Mamba2 state-space modeling to each direction with linear complexity, and fuses the multi-directional outputs via learnable adaptive weights followed by a residual connection, thereby capturing long-range temporal and spatial dependencies without quadratic attention cost. Finally, UFEM (uncertainty-aware feature enhancement module) mitigates severe back-scatter induced ambiguity by estimating a normalized feature-level uncertainty map and attenuating uncertain regions while preserving high-confidence responses, again with residual reinforcement, which reduces depth bias and structural collapse under heavy fog.

Experiment and result. We train FogMaNet mainly on large-scale synthetic foggy photon-count histograms to cover extreme conditions in Figure 1(j). Specifically, we run Monte Carlo simulations for a discrete set of fog levels parameterized by meteorological visibility V, convert each V to an extinction coefficient via the Koschmieder relation $\beta = 3.912/V$, and fit a Gamma model $\text{Gamma}(k, \theta)$ to the simulated back-scatter returns to obtain an explicit calibration $V \leftrightarrow (k, \theta)$. After this one-time calibration, we generate training data efficiently by sampling fog back-scatter directly from the fitted (k, θ) at the desired V instead of per-photon Monte Carlo, and combining it with an attenuated target return and Poisson shot noise.

Figures 1(e)–(h) summarize the quantitative evaluation on the Middlebury dataset under four simulated fog settings with increasing scattering severity. Across all fog levels, FogMaNet consistently achieves the best overall accuracy and the lowest reconstruction error, while exhibiting only minor degradation as fog becomes denser, indicating strong robustness beyond the training distribution. To highlight scene-level behavior, Figure 1(i) visualizes the Middlebury "Art" case. Compared with Yao et al., FogMaNet preserves sharper depth discontinuities and more complete object surfaces, and the corresponding error map shows substantially fewer incorrect regions, particularly in areas where fog backscatter typically triggers structural collapse. Finally, Figure 1(k) reports real foggy single-photon LiDAR reconstructions collected in a laboratory-controlled environment. Although no ground-truth depth is available, FogMaNet produces noticeably cleaner depth maps with clearer object contours and fewer back-scatter induced spurious points, demonstrating that the proposed architecture transfers effectively from simulation to real-world fog conditions.

Conclusion. In summary, we introduced FogMaNet, a Mamba2-driven fog-aware depth reconstruction framework for single-photon LiDAR that combines efficient long-range spatial-temporal modeling (QTSMamba) with structure-preserving spatial gating (AGSC) and uncertainty-guided refinement (UFEM). Evaluations on Middlebury-based fog synthesis at multiple densities, and real foggy captures consistently show that FogMaNet delivers more accurate and visually stable depth reconstructions than representative baselines, particularly in the extreme low-SBR regime dominated by back-scatter. Future work will prioritize broader outdoor generalization via real fog data and domain adaptation.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. U21A20478, 6247610) and Fundamental Research Funds for the Central Universities (Grant No. 2024ZYGXZR062).

References

1 Pu Y, Zhu C, Yao G, et al. Computational imaging based on single-photon detection: a survey. Artif Intell Rev, 2025, 58: 251
2 Wang H, Deng X, Wei M, et al. Depth image reconstruction in heterogeneous foggy environments through single-photon imaging. IEEE Trans Instrum Meas, 2025, 74: 1–17
3 Pu Y, Zhu C, Yao G, et al. Mamba-unet-depth: enhancing long-range dependency for photon-efficient imaging. IEEE Trans Circuits Syst II, 2025, 72: 1123–1127
4 Zhang Y, Li S, Sun J, et al. Three-dimensional single-photon imaging through realistic fog in an outdoor environment during the day. Opt Express, 2022, 30: 34497–34509
5 Guo A, Chen X, Dong F Y, et al. A 22-nm 64-kB lightning-like hybrid computing-in-memory macro with a compressed adder tree and analog-storage quantizers for transformer and CNNs. Sci China Inf Sci, 2025, 68: 222401