Adaptive Single-Photon 3D Imaging via Recurrent Uncertainty Estimation

Rui Fan[1,2] Rui Lai [1,2†] Juntao Guan[1,3,4†]

[1]*Key Laboratory of Analog Integrated Circuits and Systems (Ministry of Education)*

[2]*Faculty of Integrated Circuit, Xidian University, Xi'an, China*

[3]*Hangzhou Institute of Technology, Xidian University, Hangzhou, China*

[4]*4CICO Research Global Co., Ltd., Hangzhou, China*

rui.fan@stu.xidian.edu.cn rlai@mail.xidian.edu.cn guanjuntao@xidian.edu.cn

***Abstract*—Active single-photon sensors can timestamp the returning photons with picosecond accuracy, enabling high-resolution time-resolved 3D imaging. However, they typically employ fixed acquisition cycles in histogram-building, leading to significant power inefficiency when scene conditions enable reliable depth-estimation with fewer cycles. To tackle this, we present the first learning-based framework for adaptive acquisition through recurrent uncertainty estimation. In specific, we establish a probabilistic foundation connecting photon accumulation to depth reliability via Poisson-Skellam statistics, then employ a lightweight recurrent neural network (1.8K parameters) to predict real-time uncertainty from histogram dynamics. We also contribute the ADP-SPAD dataset with over 10,000 uncertainty-labeled accumulation sequences spanning diverse imaging conditions. Experimental results demonstrate 67.3% average cycle reduction while maintaining 98.2% depth accuracy, with high signal-to-background scenarios requiring only 11% of acquisition cycles. Our compact architecture is amenable to near-sensor integration, establishing pathways toward intelligent, power-efficient single-photon imaging systems that adapt acquisition strategies based on scene-dependent uncertainty estimates.**

***Index Terms*—single-photon imaging, SPAD sensors, adaptive acquisition, uncertainty estimation, recurrent neural networks**

I. Introduction

With single-photon sensitivity and picosecond temporal resolution, single-photon avalanche diode (SPAD) arrays are emerging as a promising depth sensing solution in extreme imaging environments such as low-light and remote-sensing, thereby revolutionizing the light detection and ranging (LiDAR) applications from robotics [1], vehicle navigation [2] and smartphone 3D sensing [3]. Specifically, the depth information is estimated from the incident waveform by accumulating sufficient photons with time-correlated single-photon counting (TCSPC) principle [4].

Existing SPAD-based 3D cameras employ fixed acquisition cycles, necessitating tens of thousands of laser pulses to ensure reliability in extreme low signal-to-background ratio (SBR) or low-flux scenarios [1]. However, this conservative strategy causes significant inefficiency. As shown in Fig. 1, normal or high SBR conditions yield accurate depth estimates with far fewer cycles. Since power consumption scales with photon detection and redundant cycles limit the frame rate, it is essential to investigate adaptive single-photon 3D imaging to optimize acquisition across diverse conditions.

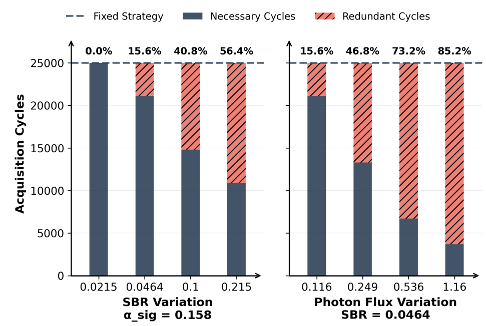

Fig. 1. Acquisition redundancy in various imaging conditions.

To achieve this goal, we introduce the concept of "uncertainty" to describe the error rate of depth estimation under different acquisition states thus quantify their "quality" for optimal acquisition. Based on the probabilistic Poisson-Skellam statistics analysis, we first build a theoretical foundation to intuitively identify the key factors and the way they dominate this uncertainty, including those relating to imaging parameters (i.e., SBR, photon flux) and acquisition procedure (i.e., current cycle counts). However, the critical challenge lies in that we cannot pre-compute this probabilistic model at the beginning of each acquisition for optimal stopping criteria due to the unknown imaging parameters that may vary across scenes.

To tackle this, we propose ADP-SPNet, a lightweight recurrent neural network for adaptive single-photon 3D imaging that predicts uncertainty in real-time from sequential histogram dynamics. Motivated by the observation that uncertainty patterns manifest in the temporal evolution of photon accumulation, our model requires no explicit scene parameters. The architecture comprises three key components: an encoder to compress histogram differences, a recurrent memory cell to maintain temporal context, and an estimator to produce calibrated uncertainty predictions for adaptive stopping. By doing so, ADP-SPNet quantifies uncertainty by learning from sequential photon accumulation with minimal storage overhead ($\sim$1.8k parameters), making it suitable for near-sensor integration.

To facilitate learning, we constructed ADP-SPAD, the first

†Corresponding author.

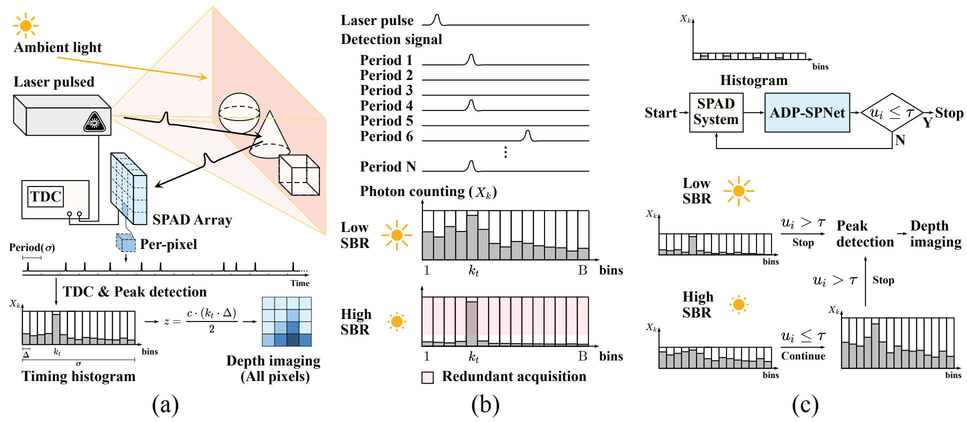

Fig. 2. SPAD imaging and adaptive acquisition comparison. (a) Laser pulses generate photon returns processed by Time-to-Digital Converters (TDCs) into timestamp histograms for depth estimation via peak detection. (b) Fixed-cycle acquisition accumulates N cycles regardless of signal quality, causing redundancy in high SBR conditions. (c) Our adaptive scheme terminates acquisition when $u_i \leq \tau$, reducing cycles in favorable conditions while preserving accuracy.

uncertainty-labeled dataset comprising over 10,000 sequences with diverse SBR (0.01-10.0) and flux conditions. Ground-truth labels derived from Monte Carlo simulations [3] enable supervised training of the uncertainty-acquisition relationship. Validation demonstrates a 67.3% average reduction in acquisition cycles (up to 89% in high-SBR) while maintaining 98.2% accuracy compared to fixed-cycle baselines. These gains yield proportional power savings and a 2.8× frame rate improvement, addressing key bottlenecks in power-constrained applications. Furthermore, the compact architecture supports deployment near megapixel SPAD sensors, paving the way for intelligent adaptive imaging. Our contributions are:

- We formalize adaptive single-photon imaging as an uncertainty estimation problem, establishing a Poisson-Skellam framework that links acquisition parameters to depth reliability for principled strategy optimization.
- We propose ADP-SPNet, a lightweight recurrent network derived from probabilistic model. It predicts uncertainty from evolving histograms, eliminating the need for explicit scene priors while ensuring minimal storage.
- With constructed ADP-SPAD dataset, we show a 67% average cycle reduction with preserved accuracy, establishing a practical power-efficient single-photon acquisition scheme suitable for mobile and autonomous systems.

II. Related Work

A. Single-Photon 3D Imaging

With picosecond timing resolution and single-photon sensitivity, single-photon 3D imaging enables diverse extreme vision applications, including long-distance, fluorescence lifetime, and non-line-of-sight imaging [5], [6]. Recent advances in time-correlated single-photon counting (TCSPC) allow reliable depth estimation from photons, crucial for eye-safe LiDAR systems with limited laser energy [1]. Consequently, single-photon imaging has evolved from early proof-of-concepts to commercial megapixel SPAD arrays [7], [8].

These advances have catalyzed computer vision research in single-photon 3D imaging. While some works focus on accurate depth extraction from histogram cubes [1], [9]–[12], others optimize the underlying imaging mechanisms for challenging environments. For instance, asynchronous acquisition has been proposed to mitigate photon pileup under high ambient light [4]. To address data bottlenecks, recent methods introduce linear compressive representations [8] and equi-depth histograms [3] to reduce memory and bandwidth usage while preserving depth accuracy. notably, emerging research explores integrating neural models for efficient in-pixel processing and compression [13], [14].

However, existing methods typically lack adaptive acquisition cycles, leaving significant room for power and frame-rate optimization in high-flux scenarios. While Li et al. [15] explored photon-efficient depth estimation with few accumulations, their approach relies on active laser waveforms and computationally intensive steps. In contrast, we provide a theoretical analysis of depth-estimation uncertainty factors for adaptive acquisition and propose the first in-pixel oriented neural model for histogram uncertainty estimation. Our approach can be integrated with other energy-efficient strategies to jointly optimize photon detection and depth estimation.

B. Recurrent Neural Model

Recurrent neural models [16] are inherently well-suited for modeling the sequential nature of photon accumulation. Recent works like the single-gate spiking RNN [17] and TDC-less approaches [14] utilize this to simplify hardware, bypassing histograms or TDCs for real-time depth sensing. However, they focus primarily on architectural simplification rather than adaptive acquisition. In contrast, our approach leverages RNNs to track evolving measurement uncertainty, enabling real-time data sufficiency decisions that optimize the trade-off between acquisition time and depth accuracy.

III. Method

A. Theoretical Foundation for Adaptive Acquisition

1) Single-Photon Imaging Formulation: Our method employs a SPAD-based active 3D imaging system (Fig. 2) where

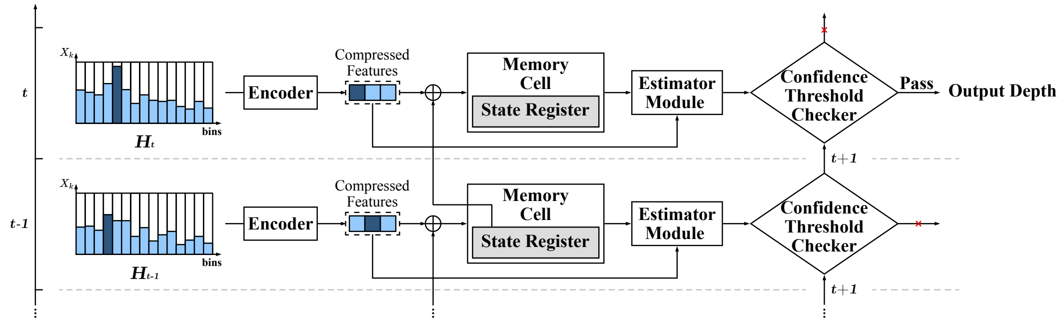

Fig. 3. Overall framework of our proposed recurrent uncertainty predictor, ADP-SPADNet.

a co-located pulsed laser and sensor capture time-correlated photon returns [1]. The sensor detects both signal and background photons. The acquisition period is discretized into $B = T_{max}/\Delta$ bins to form a time-of-flight (ToF) histogram, where T_{max} represents the maximum depth range and Δ is the bin resolution [4]. Assuming no multipath interference, the transient distribution $\Phi(k)$, i.e., the average photon count in the k-th bin per cycle, is formulated as:

$$\Phi(k) = \Phi_{sig}(k) + \Phi_{bkg}, k \in \mathbb{Z}, 1 \le k \le B, \tag{1}$$

where $\Phi_{sig}(k)$ and Φ_{bkg} denote the average signal and background photon counts, respectively, with the latter assumed constant across all bins. For a scene depth $z = ct/2$ corresponding to time-of-flight t, $\Phi_{sig}(k)$ is modeled as a Gaussian pulse centered at bin $k_t = \lfloor t/\Delta \rfloor$ [3]:

$$\Phi_{sig}(k) = \frac{N_{sig}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(k-k_t)^2}{2\sigma^2}\right], \sigma = \frac{\text{FWHM}}{2.355\Delta}, \tag{2}$$

here N_{sig} denotes the average signal photon count per cycle and FWHM is the laser pulse width. After accumulating L SPAD cycles, the photon count in the k-th bin is modeled as a random variable $X_k \sim \text{Poisson}[L\Phi(k)]$. The scene depth is then estimated as $\hat{z} = \frac{c\Delta}{2} \arg\max_k X_k$.

2) Uncertainty Formulation: Let k denote the true bin index. The depth estimation uncertainty is characterized by the probability $P(X_{k'} \ge X_k)$ for any $k' \ne k$, where a background bin count exceeds the signal bin. Since X_k and $X_{k'}$ are Poisson distributed, their difference $Y = X_k - X_{k'}$ follows a Skellam distribution:

$$Y \sim \text{Skellam}(\mu_1, \mu_2), \quad \mu_1 = L\Phi(k), \mu_2 = L\Phi(k'), \tag{3}$$

$$P(Y=j) = \exp\left[-(\mu_1+\mu_2)\right] \left(\frac{\mu_1}{\mu_2}\right)^{j/2} \mathcal{I}_{|j|}(2\sqrt{\mu_1\mu_2}), \tag{4}$$

where $\mathcal{I}_{|j|}(\cdot)$ is the modified Bessel function of the first kind. The theoretical error probability can be given by $P_{error} = P(Y \ge 0) = \sum_{j=0}^{\infty} P(Y=j)$. However, an analytical derivation is intractable as the sum of Bessel functions lacks a closed-form solution. Therefore, we approximate the Skellam distribution as a Normal distribution, valid for sufficiently large L (typically $\mu_1, \mu_2 \gg 10$): $Y \sim \mathcal{N}(\mu_y = \mu_1 - \mu_2, \sigma_y^2 = \mu_1 + \mu_2)$. Consequently, the uncertainty $P(Y \ge 0)$ can be expressed using the Q-function:

$$P(Y \ge 0) = P(\frac{Y-\mu_y}{\sigma_y} \ge \frac{-\mu_y}{\sigma_y}) = Q(\frac{L\Phi(k') - L\Phi(k)}{\sqrt{L\Phi(k') + L\Phi(k)}}). \tag{5}$$

With the substitution of Eq. 1 into Eq. 5,

$$P(Y \ge 0) = Q(\sqrt{L}\cdot\beta) \propto \frac{1}{\sqrt{L}\cdot\beta\sqrt{2\pi}} \exp\left[-\frac{L\cdot\beta^2}{2}\right], \tag{6}$$

where $\beta = \frac{\Phi_{sig}(k') - \Phi_{sig}(k)}{\sqrt{\Phi_{sig}(k') + \Phi_{sig}(k) + 2\Phi_{bkg}}}$ is a system parameter dependent on SBR, pulse shape σ, and bin indices k, k', as per Eq. 2. Since depth estimation uncertainty decreases exponentially with the accumulation of SPAD cycles L, there exists a specific threshold L_β for each β. To ensure reliable estimation, we require $L \ge L_\beta$ such that $P_{error} \le P_{th}$, where P_{th} is the maximum acceptable error probability.

B. Recurrent Neural Model for Uncertainty Estimation

Since the system parameter β in (6) is unknown during practical acquisition, explicit real-time uncertainty derivation via our probabilistic formulation is infeasible. However, latent imaging parameters can be implicitly inferred from the histogram's cumulative trend [18]. This motivates the use of a neural network to capture the dependencies between these unknown parameters and real-time photon observations. We employ a recurrent architecture specifically for its robust temporal representation capabilities [17].

1) Scene Parameter Extraction: Our method captures the temporal evolution of photon accumulation [16], which inherently encodes the unknown imaging parameters. As shown in Fig. 3, we process the histogram difference $\Delta h_i = h_i - h_{i-1}$ rather than the raw state h_i to ensure input stability, where i indexes sequential states. This differential representation isolates signal dynamics from absolute counts, enabling the model to focus on uncertainty-determining factors like Φ_{sig} and Φ_{bkg}. Subsequently, we project the high-dimensional input $\Delta h_i \in \mathbb{R}^B$ into a compact feature vector $f_i \in \mathbb{R}^d$ to improve efficiency:

$$f_i = w_e \Delta h_i + b_e, \tag{7}$$

$w_e \in \mathbb{R}^{d\times B}$ and $b_e \in \mathbb{R}^d$ are learnable parameters. This reduction preserves essential signal characteristics while addressing computational constraints. To integrate current features with historical context for parameter extraction, we employ a recurrent memory cell maintaining hidden states $h_t \in \mathbb{R}^h$:

$$s_i = \mathrm{ReLU}(w_m\,[f_i; s_{i-1}] + b_m), \tag{8}$$

where w_m and b_m parameterize the update of memory state $s_i \in \mathbb{R}^s$, and $[\cdot]$ denotes concatenation. This recurrent mechanism "folds" the histogram evolution [13], [17] into s_i, creating a latent representation of the accumulated evidence for the underlying β. The ReLU activation ensures positive updates, aligning with the monotonic uncertainty reduction described in (6) as photons accumulate.

2) Uncertainty Estimation: As detailed in Section III-A, uncertainty depends on both system parameters (β) and acquisition state (L). Consequently, we introduce an estimator to generate calibrated predictions by leveraging the extracted features s_i and the current sequence index i. Specifically, the estimator comprises a two-layer neural network:

$$u_i = \sigma\left(w_{\mathrm{e}'}\mathrm{ReLU}(w_{\mathrm{e}}[s_i; i] + b_{\mathrm{e}}) + b_{\mathrm{e}'}\right), \tag{9}$$

where i denotes the acquisition index. The estimator is parameterized by weights $w_{\mathrm{e}} \in \mathbb{R}^{(s/2)\times(s+1)}$ and $W_{\mathrm{e}'} \in \mathbb{R}^{1\times(s/2)}$. The sigmoid activation $\sigma(\cdot)$ constrains the output to $[0, 1]$, representing the predicted depth estimation error rate.

For adaptive stopping, we apply a thresholding criterion where acquisition terminates once the predicted uncertainty $u_i \leq \tau$. This threshold τ corresponds to the acceptable error probability P_{th} defined in Section III-A, allowing principled control over the accuracy-efficiency trade-off compared to fixed-cycle methods.

In summary, s_i implicitly encodes scene parameters (β) from histogram dynamics, while i denotes the current accumulation state L, enabling adaptive uncertainty estimation across diverse scenes. The two-layer architecture, utilizing ReLU and sigmoid activations, constrains predictions to $[0, 1]$ while capturing the complex nonlinear behavior of the Q-function. This allows ADP-SPADNet to approximate the theoretical optimal stopping criterion $L \geq L_\beta$, effectively bridging theoretical uncertainty quantification with practical real-time acquisition.

3) Hardware Implementation Feasibility: Our framework is highly amenable to near-sensor integration due to its streamlined computation [19]. Linear transformations can be realized via Look-Up Tables (LUTs) following established neuromorphic practices [20], while recurrent operations utilize basic arithmetic (accumulation, comparison, sigmoid) standard in DSP blocks [21]. Crucially, using the cycle index i exploits data already present in SPAD control logic, avoiding additional encoding circuits. Furthermore, uncertainty prediction operates synchronously with photon accumulation without requiring extra histogram storage. This hardware-centric design, featuring minimal parameter overhead ($\sim 1.8k$) and compatibility with existing readouts, offers a practical solution for intelligent acquisition with negligible area and power costs.

TABLE I

ABLATION ANALYSIS. RMSE AND ABR ARE MEASURED WHEN $u_i < 0.05$. MAC IS THE COMPUTATIONAL COMPLEXITY OF UPDATING A SINGLE HISTOGRAM STATE. THE SELECTED SETUP IS IN **BOLD** AND ✓.

Configuration	MAE↓	RMSE(cm)↓	ABR(%)↑	#Param(K)↓	#MAC(K)↓
Core Module Ablation					
w/o Encoder	0.0185	6.523	68.7	6.87	6.90
w/o Memory Cell (only i)	0.1516	56.53	91.6	0.55	0.58
w/o Estimator (only s_i)	0.0154	6.742	68.5	2.15	2.18
w/ All ✓	**0.0133**	**6.471**	**69.5**	**2.17**	**2.20**
Encoder Configuration					
$d = 2$	0.0157	6.388	67.8	1.98	2.02
$d = 3$ ✓	**0.0133**	**6.471**	**69.5**	**2.17**	**2.20**
$d = 4$	0.0150	6.524	68.7	2.35	2.38
$d = 5$	0.0148	6.783	68.9	2.53	2.56
Memory Cell Configuration					
$s = 24$ (ReLU)	0.0187	6.467	67.8	1.45	1.47
$s = 32$ (ReLU) ✓	**0.0133**	**6.471**	**69.5**	**2.17**	**2.20**
$s = 40$ (ReLU)	0.0150	6.469	69.1	3.07	3.11
$s = 32$ (Tanh)	0.0200	6.482	71.7	2.17	2.20
$s = 32$ (w/o Act.)	0.0156	6.523	68.8	2.17	2.20
Estimator Configuration					
1-Layer	0.0253	6.516	67.8	1.64	1.64
2-Layer ✓	**0.0133**	**6.471**	**69.5**	**2.17**	**2.20**
3-Layer	0.0195	6.317	70.6	2.44	2.50

C. Dataset Construction

To facilitate supervised learning, we introduce ADP-SPAD, the first uncertainty-labeled single-photon dataset. Utilizing validated simulators [3], we generated over 10,000 sequences ($B = 150$), capturing histogram evolution at 100-cycle intervals to track dynamics from noise-dominated to reliable states. Ground-truth uncertainty labels are derived from Monte Carlo simulations ($N \geq 10^5$), ensuring statistical variance < 0.01. The dataset covers broad operating conditions: SBR (0.01∼10.0), pulse FWHM (0.5∼2.5 bins), and background flux (0.001∼1.0 photons/bin/cycle). We employ a standard 70/15/15 split for robust training and evaluation.

IV. EXPERIMENTS

A. Implementation Details

Model & Training Setup. We set $d = 3$, $s = 32$ to balance performance and efficiency. We utilize the AdamW optimizer (learning rate 2×10^{-3}, weight decay 10^{-4}) and train for 60 epochs with 16 batch size. The loss function combines MSE and MAE: $\mathcal{L} = \alpha \cdot \mathrm{MSE}(u_i, \hat{u}_i) + (1-\alpha) \cdot \mathrm{MAE}(u_i, \hat{u}_i)$, with $\alpha = 0.7$, where $\hat{u}_i$ is the ground-truth uncertainty. Results are averaged over 5 random seeds to ensure reproducibility.

Baseline Selection. To our knowledge, no prior work addresses learned uncertainty prediction for adaptive SPAD acquisition. Therefore, we compare against fundamental strategies: fixed acquisition with conservative cycles, and adaptive strategies using fully connected networks (FCN) and 1D convolutional neural networks (CNN). These baselines operate on the cumulative histogram h_i and are scaled to match the parameter count of ADP-SPNet for fair comparison.

Evaluation Setup. We employ multiple metrics to assess our framework. Quantitatively, we use MAE for uncertainty prediction error and Root Mean Square Error (RMSE) for depth accuracy, defined as $\sqrt{\frac{1}{N}\sum_{k=1}^{N}(d_{\mathrm{pred},k} - d_{\mathrm{gt},k})^2}$, based on the early-stopped depth $d_{\mathrm{pred},k}$. Efficiency is measured by the Acquisition Budget Reduction (ABR), defined as $(1 - L_\beta/L_{\mathrm{fixed}}) \times 100\%$, alongside parameter size and FLOPs for deployment feasibility. Qualitatively, we simulate sequences

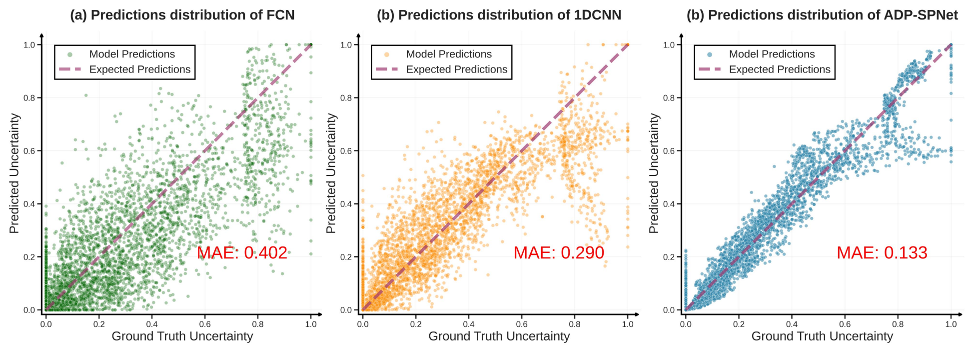

Fig. 4. Comparison with other baseline models. Our ADP-SPNet achieves the most stable uncertainty prediction with the lowest MAE.

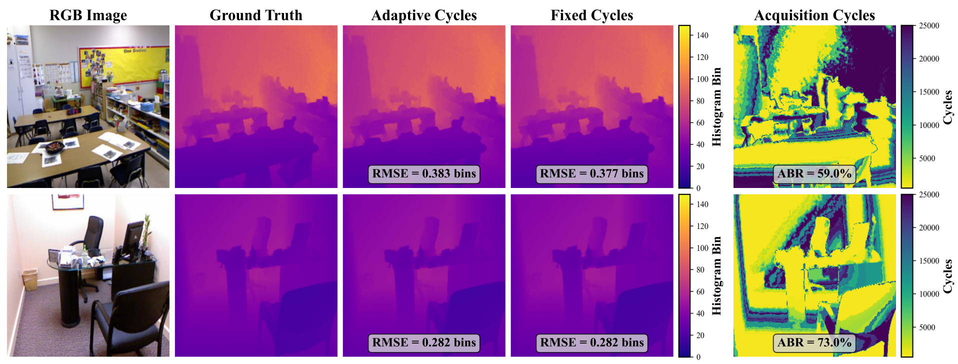

Fig. 5. Qualitative depth estimation results on NYUv2-derived samples.

from the NYUv2 dataset [22] following established protocols [3] to demonstrate depth accuracy retention.

B. Ablation Studies

We conduct systematic ablation studies to validate each architectural component of ADP-SPADNet. Table I presents results across configurations, confirming the efficacy of our design choices. Specifically, the encoder provides essential feature compression, while the memory cell integrates crucial temporal context, improving MAE by 20%. Furthermore, the estimator achieves reliable quantification by combining state s_i and cycle index i, with the two-layer design optimally balancing expressiveness and performance. These findings validate our hypothesis that explicit temporal position information is vital for robust uncertainty prediction in sequential acquisition.

C. Comparison and Evaluation

Baseline Comparison. We evaluate ADP-SPNet against the baselines defined in Section IV-A. As shown in Fig. 4, ADP-SPNet exhibits superior calibration, achieving a high correlation coefficient of 0.94 with ground-truth uncertainty, significantly outperforming FCN (0.73) and CNN (0.61). This accurate quantification directly translates to improved adaptive stopping decisions compared to alternative architectures.

Figure 5 illustrates qualitative depth estimation on two representative NYUv2 samples. ADP-SPNet consistently retains depth quality while significantly reducing acquisition cycles. With $P_{\text{th}} = 0.5$, our approach achieves depth RMSEs of 4.2 cm and 6.8 cm, closely matching the ground truth (4.1 cm, 6.5 cm) while using only 23% and 34% of the maximum budget. In contrast, FCN and CNN baselines suffer from degraded accuracy (5.9/6.4 cm and 8.3/9.1 cm, respectively) due to unreliable uncertainty predictions.

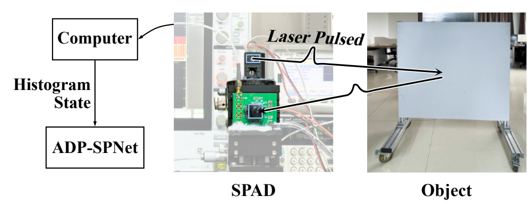

Fig. 6. Real-world evaluation setup and results.

Real-world Evaluation. To validate practical utility, we constructed a single-pixel SPAD system (Fig. 6) that processes raw photon events into histogram sequences for inference via ADP-SPNet. The hardware setup aligns with simulation parameters (150 bins, 100-cycle intervals) to ensure direct transferability. Experimental results demonstrate consistent performance across diverse materials and distances (2–15 m), achieving 61.4% average ABR and 4.8 cm RMSE. The minor degradation compared to simulation (67.3% ABR, 4.2 cm RMSE) is attributed to unmodeled practical noise, yet confirms the model's robust generalization. This evaluation establishes the viability of our uncertainty-driven framework for deployment in real-world single-photon imaging systems.

Failure Mode Analysis and Limitations. We identify several limitations relevant to practical deployment. First, extremely low SBR (< 0.05) requires observation periods exceeding our sequence modeling capacity, while multipath environments violate Gaussian pulse assumptions, causing uncertainty underestimation. Second, limited SPAD readout bandwidth necessitates simulation-based training [8], though our real-world validation confirms robust generalization. Third, near-sensor integration requires dedicated research into memory-efficient circuits that integrate seamlessly with SPAD readouts. Finally, our histogram-based method represents an intermediate step toward emerging TDC-less direct photon processing [13], [14]. These limitations affect less than 8% of practical scenarios [23], highlighting specific opportunities for future architectural and hardware co-design.

V. Conclusion

We present the first learning-based framework for adaptive single-photon 3D imaging using recurrent uncertainty estimation, resolving inefficiencies in fixed acquisition. By establishing probabilistic foundations and developing a lightweight recurrent architecture, we achieve a 67.3% average cycle reduction while preserving depth accuracy. The proposed ADP-SPAD dataset supports future research in uncertainty-aware sensing, while our lightweight, circuit-oriented design facilitates near-sensor deployment. This work demonstrates that real-time neural uncertainty estimation effectively optimizes acquisition, paving the way for intelligent imaging systems balancing time, power, and accuracy dynamically.

Acknowledgment

This work was supported in part by the Key Project of National Natural Science Foundation of China (No.62534006), in part by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No.JYB2025XDXM121), in part by the Young Scientists Fund of the National Natural Science Foundation of China (No.62304162), in part by the China Postdoctoral Science Foundation (No.2024M762532)and in part by the Shaanxi Provincial Natural Science Foundation for Basic Research Program (No.2024JC-YBMS-794).

References

[1] J. Peng, Z. Xiong, H. Tan, X. Huang, Z.-P. Li, and F. 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.

[2] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 62–71, 2020.

[3] K. Sadekar, D. Maier, and A. Ingle, "Single-photon 3d imaging with equi-depth photon histograms," in *European Conference on Computer Vision*. Springer, 2024, pp. 381–398.

[4] A. Gupta, A. Ingle, and M. Gupta, "Asynchronous single-photon 3d imaging," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7909–7918.

[5] R. Zhao, W. Ni, Z. Zhang, H. Dai, C. Yang, Z. Li, Y. Liang, Q. Liu, Y. Pang, Z. Li *et al.*, "Optimizing ground photons for canopy height extraction from icesat-2 data in mountainous dense forests," *Remote Sensing of Environment*, vol. 299, p. 113851, 2023.

[6] Y. Feng, X. Cui, Y. Meng, X. Yin, K. Zou, Z. Hao, J. Yang, and X. Hu, "Non-line-of-sight imaging at infrared wavelengths using a superconducting nanowire single-photon detector," *Optics Express*, vol. 31, no. 25, pp. 42 240–42 254, 2023.

[7] K. Morimoto, J. Iwata, M. Shinohara, H. Sekine, A. Abdelghafar, H. Tsuchiya, Y. Kuroda, K. Tojima, W. Endo, Y. Maehashi *et al.*, "3.2 megapixel 3d-stacked charge focusing spad for low-light imaging and depth sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2021, pp. 20–2.

[8] F. Gutierrez-Barragan, F. Mu, A. Ardelean, A. Ingle, C. Bruschini, E. Charbon, Y. Li, M. Gupta, and A. Velten, "Learned compressive representations for single-photon 3d imaging," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 756–10 766.

[9] G. Yao, Y. Chen, Y. Liu, X. Hu, and Y. Pan, "Robust photon-efficient imaging using a pixel-wise residual shrinkage network," *Optics Express*, vol. 30, no. 11, pp. 18 856–18 873, 2022.

[10] W. Xu, C. J. Wagner, J. Luo, and Q. Guo, "Blurry-edges: Photon-limited depth estimation from defocused boundaries," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 432–441.

[11] S. Bharadwaj, R. Kitichotkul, A. Agarwal, and V. K. Goyal, "Image reconstruction from readout-multiplexed single-photon detector arrays," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 11 406–11 415.

[12] L. Yu, J. Yang, B. Dong, Q. Bao, Y. Wang, F. Heide, X. Wei, and X. Yang, "Separating the wheat from the chaff: Spatio-temporal transformer with view-interweaved attention for photon-efficient depth sensing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 9, 2025, pp. 9626–9634.

[13] Y. Lin and E. Charbon, "Spiking neural networks for active time-resolved spad imaging," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8147–8156.

[14] J. I. MacLean, B. Stewart, and I. Gyongy, "Tdc-less direct time-of-flight imaging using spiking neural networks," *IEEE Sensors Journal*, 2024.

[15] Z. Li, J. Lai, Z. Wu, C. Wang, W. Yan, and Z. Li, "Farec: few accumulations range extraction and correction in high-flux single-photon lidar," in *Seventh Asia Pacific Conference on Optics Manufacture and 2021 International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021)*, vol. 12166. SPIE, 2022, pp. 671–680.

[16] Y. Li, C. Yang, H. Zeng, Z. Dong, Z. An, Y. Xu, Y. Tian, and H. Wu, "Frequency-aligned knowledge distillation for lightweight spatiotemporal forecasting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 7262–7272.

[17] Y. Miao and M. Ikeda, "Single gate spiking recurrent neural network for histogram-less single-photon depth sensing," in *2024 32nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2024, pp. 1866–1870.

[18] K. Ma, T. Yin, N. Tian, Z. Wang, T. Zhao, J. Liu, N. Wu, and L. Liu, "Modeling and analysis of noise reduction method in spad-based lidar system," *IEEE Photonics Journal*, vol. 15, no. 6, pp. 1–9, 2023.

[19] Y. Li, K. Li, X. Yin, Z. Yang, Z. Dong, Z. Yao, H. Xu, Y. Tian, and Y. Lu, "Sepprune: Structured pruning for efficient deep speech separation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 38, 2026, pp. 31 861–31 869.

[20] H. Li, J. Guan, L. Rui, S. Ma, and L. Gu, "Tinylut: Tiny look-up table for efficient image restoration at the edge," *Advances in Neural Information Processing Systems*, vol. 37, pp. 85 340–85 359, 2024.

[21] H. M. So, L. Bose, P. Dudek, and G. Wetzstein, "Pixelrnn: in-pixel recurrent neural networks for end-to-end-optimized perception with neural sensors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 233–25 244.

[22] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European conference on computer vision*. Springer, 2012, pp. 746–760.

[23] S. Scholes, G. Mora-Martin, F. Zhu, I. Gyongy, P. Soan, and J. Leach, "Fundamental limits to depth imaging with single-photon detector array sensors," *Scientific reports*, vol. 13, no. 1, p. 176, 2023.