2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

NeuroLiDAR: Adaptive Frame Rate Depth Sensing via Neuromorphic Event-LiDAR Fusion

Darshana Rathnayake[1], Dulanga Weerakoon[2], Meera Radhakrishnan[3] and Archan Misra[1]

***Abstract*— LiDARs are widely used for 3D depth reconstruction, but their performance is often limited by inherent hardware constraints that impose trade-offs between range, spatial resolution, and frame rate. Many LiDAR systems typically operate at low frame rates (e.g., 5-10 Hz), prioritizing long-range sensing over responsiveness to rapid scene changes. We present *NeuroLiDAR*, an adaptive depth sensing framework that achieves effective frame rates of up to ≈66 Hz by fusing temporally sparse LiDAR data with temporally dense inputs from neuromorphic event cameras. *NeuroLiDAR* integrates two components: event-based keyframe detection and event-guided depth extrapolation, to dynamically adjust the sensing rate in response to scene dynamics. To evaluate our approach, we introduce *ELiDAR*, a dataset spanning outdoor and indoor scenarios, and show that *NeuroLiDAR* reduces depth reconstruction error by ≈29% in RMSE while achieving adaptive frame rates between 27.8–47.3 Hz. Our code and dataset are available at `https://github.com/darshanakgr/neurolidar`.**

I. Introduction

LiDARs have become indispensable in automotive and industrial applications, such as autonomous driving and robotic navigation, due to their ability to provide accurate and reliable 3D depth measurements. Commercial automotive LiDARs typically operate at frame rates between 5-10 Hz, depending on the sensor's range and resolution. While increasing the frame rate could improve the temporal resolution of depth tracking, it is known to both incur substantial energy overheads and reduce the operational lifespan of the laser source (i.e., the laser diode).

This creates a fundamental tradeoff. In relatively static environments, operating a LiDAR sensor at a high frame rate consumes power unnecessarily without improving performance. In contrast, dynamic scenes with fast-moving objects or sudden vehicle manoeuvres require more frequent depth estimation than current LiDARs can provide. Ideally, a depth sensing system should operate at a low baseline LiDAR rate in static conditions, while adaptively increasing its effective frame rate when scene dynamics require it. Such an *adaptive frame rate* paradigm could simultaneously prove to be both energy-efficient and responsive.

To achieve this dual objective and enable high-fidelity depth estimation without requiring the LiDAR to operate continuously at high frame rates, we propose to integrate a complementary sensing modality, *neuromorphic event cameras*, that can capture environmental *changes* at much lower system power ($\sim$ 500 mW) and finer temporal resolution ($O(1\mu s)$). Unlike conventional frame-based cameras, which capture images at fixed intervals, event cameras generate events asynchronously, at microsecond resolution [1], whenever the incident light intensity changes principally due to motion dynamics. Our key idea is to operate the LiDAR itself to capture depth maps at a low base rate, while fusing this depth estimate with the motion dynamics captured continuously by the event camera to synthetically generate additional depth maps at appropriate intermediate time instants.

Event cameras have been shown to complement RGB cameras in tasks such as image deblurring [2], depth densification [3], [4], video frame interpolation [5], and optical flow estimation [6]. Their asynchronous nature makes them particularly well suited to fill the temporal gaps between consecutive LiDAR frames in an *adaptive* fashion, generating a higher volume of motion cues exactly when changes occur rapidly. Recent works have explored event–LiDAR fusion, primarily targeting depth densification–i.e., synthetically generating super-resolution depth estimates. For example, Li et al. [7] fused events with sparse LiDAR scans to produce up to $9.6\times$ denser point clouds, but with limited throughput (≈4 fps). Other efforts have focused on filling sparse long-range LiDAR depth maps [4], [8]. Our work is, however, the first to propose *NeuroLiDAR*, a novel depth sensing framework that *performs adaptive* ***temporal*** *super-resolution, increasing the effective depth sensing frame rate by leveraging the asynchronous, low-latency dynamics captured by an event camera*, without incurring the energy cost of high frame-rate LiDAR scanning.

Unlike prior spatial densification methods, *NeuroLiDAR* performs adaptive depth extrapolation, producing new depth frames *only* when events signal substantial changes to the scene. *NeuroLiDAR* operates in two stages: (a) a lightweight event-driven *Keyframe Detector* identifies significant scene changes relative to the most recent LiDAR depth frame (prior depth frame) and then triggers the generation of synthetic depth estimates at those key moments, while (b) a U-Net-style [9] lightweight autoencoder leverages the prior (most recently captured) reference depth frame, with a voxel grid-based event representation, to generate these depth estimates via extrapolation. This closed-loop design allows *NeuroLiDAR* to deliver high temporal fidelity only when required, while exploiting LiDAR's inherently high spatial sensing resolution. We make the following **key contributions**:

- We present ***Neuro***morphic ***LiDAR*** (*NeuroLiDAR*), a novel depth sensing system that boosts LiDAR's frame rate through *event-guided depth extrapolation*, going beyond

[1]Singapore Management University, Singapore, darshanakg.2021@smu.edu.sg, archanm@smu.edu.sg
[2]Singapore-MIT Alliance for Research and Technology Centre, Singapore, dulanga.weerakoon@smart.mit.edu
[3]University of Technology Sydney, Australia, meeralakshmi.radhakrishnan@uts.edu.au

prior densification approaches. Our system also enables *adaptive frame-rate LiDAR operation*, whereby the depth frames are generated, in a streaming fashion, only when scene dynamics warrant such generation, providing high temporal resolution without higher LiDAR scan rates or power costs.

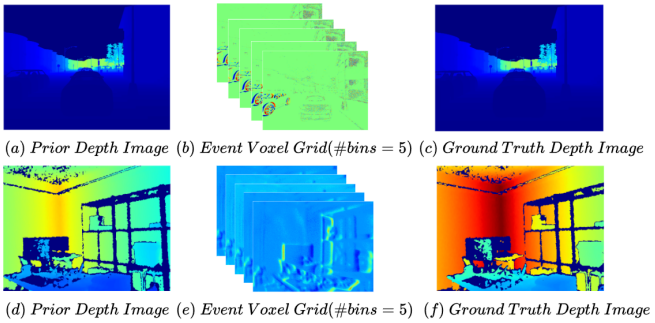
- We built a real prototype of *NeuroLiDAR*, carefully architected for high frame rate operation on embedded devices, using a commercial event camera and a LiDAR platform. We demonstrate its low-latency operation on resource-constrained embedded devices: when deployed on an NVIDIA Jetson Orin device, *NeuroLiDAR* can support an effective frame rate up to 66.67 Hz.
- To rigorously evaluate the performance of *NeuroLiDAR*, we curate a new benchmark dataset, *ELiDAR*, which consists of both a large-scale synthetic data segment generated using the CARLA simulator [10] and a smaller real-world segment collected with our *NeuroLiDAR* prototype. The synthetic segment provides synchronized depth and event data under significantly more realistic and diverse driving conditions (e.g., varying speeds, traffic densities, pedestrian activities, abrupt, and changing environmental settings), compared to prior work such as ALED [8]. Experiments on the synthetic segment show that *NeuroLiDAR* achieves ≈29% reduction in depth estimation error (RMSE) compared to a conventional LiDAR operating at 10 Hz. Subsequently, by experimentally deploying the *NeuroLiDAR* prototype on a mobile robot, we demonstrate that *NeuroLiDAR* can achieve a similar ≈28% reduction in depth estimation RMSE in an indoor setting.

II. Related Work

We discuss prior works on event-based depth sensing as well as event-based fusion approaches.

Event-Only Depth: Event cameras have found use in motion-centric tasks such as optical flow [6], gesture recognition [11], action recognition [12], SLAM [13], [14], and video super-resolution [5], largely due to their ability to capture fast dynamics with minimal motion blur. This same property makes them attractive for depth estimation in dynamic scenes. Spiking neural networks (SNNs) have been leveraged for stereo event pairs [15], which improved accuracy on MVSEC dataset [16] by ∼20% and offered 58x energy efficiency relative to ANNs. Earlier Kim et al. [13] used probabilistic filters for simultaneous localization and mapping (SLAM), though the reconstructions were semi-sparse and evaluated qualitatively. More recent systems such as EVI-SAM [14] perform robust visual–inertial fusion with dense mapping, but at modest mapping frequencies (∼7 Hz). While demonstrating the potential of purely event-based depth sensing, these works fall short of providing dense, accurate, and high frame-rate depth reconstructions.

Event–RGB Fusion for Depth Estimation: Early efforts to integrate event data with RGB cameras primarily focused on monocular depth prediction. Gehrig et al. [3] extended recurrent neural architectures to jointly process irregular event streams and RGB frames, yielding up to 30% lower mean absolute depth error than frame-based baselines. Other methods leveraged event–RGB fusion under adverse conditions: EVEN [17] enhanced RGB images with GANs before fusing them with events for robust depth estimation at night. More recently, hybrid spiking-transformer architectures [18] combined event-driven SNN feature extraction with RGB-based transformers for depth prediction, though the latency remained high (≈1 s per frame on GPUs, 171 ms on custom hardware). These works demonstrated the complementary value of events and frames, but remain limited by the range and passive stereo constraints of RGB inputs.



(a) *Prior Depth Image* (b) *Event Voxel Grid*($\#bins = 5$) (c) *Ground Truth Depth Image*

(d) *Prior Depth Image* (e) *Event Voxel Grid*($\#bins = 5$) (f) *Ground Truth Depth Image*

Fig. 1: Examples from the *ELiDAR* dataset: (a)–(c) simulated outdoor scenarios, (d)–(f) real-world indoor scenarios.

Event–LiDAR Fusion for Depth Densification: Recent works have addressed densifying sparse LiDAR depth maps using event streams. Li et al. [7] estimated depth for event pixels using neighbouring LiDAR points, yielding ≈9.6× denser point clouds and improved detection accuracy, but limited to semi-dense reconstructions at ≈4 fps. More recently, multi-stage approaches [4], [8] have been explored for spatial depth map densification using LiDAR and event data. While the work [4] reported throughput values up to 56 fps, the definition of latency was unclear, and the evaluation of reconstruction accuracy was limited to depth values of 50 m. Collectively, these methods improve the spatial resolution of LiDAR but do not target temporal adaptivity.

Our proposed *NeuroLiDAR* framework is thus motivated by the unresolved problem of adaptively increasing LiDAR's effective frame rate under fast, real-world dynamics, spanning both outdoor and indoor scenarios.

III. Background and Motivation

Reliable depth perception in dynamic environments, such as encountered in autonomous driving or robot navigation applications, requires sensors that can capture rapid scene changes. However, current commercial LiDARs remain bound by low frame rates, leaving critical gaps that demand new sensing strategies, as well as dedicated datasets to evaluate such strategies.

A. *Construction of* ELiDAR *Dataset*

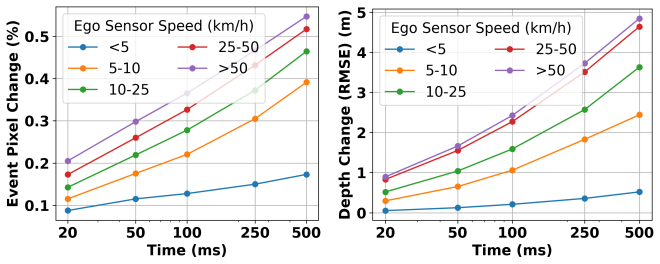
For training and evaluation of *NeuroLiDAR*, we construct a dataset termed *ELiDAR*. Our dataset contains two segments for a simulated outdoor scenario and a real-world indoor scenario. The simulated segment is generated using the CARLA simulator [10], [19], which captures diverse urban driving

scenarios with aligned LiDAR and event camera streams, recorded at 200 Hz and configured for long-range sensing (up to 200 m). Both the LiDAR and event cameras are mounted on the roof of the ego vehicle within CARLA and configured to capture at a spatial resolution of 480×640 pixels. The fields of view of the two sensors are aligned to ensure pixel-level correspondence between the depth and event modalities. The simulated split of the dataset spans varied lighting, weather, traffic, and pedestrian conditions, with ego vehicle speeds ranging from slow cruising to highway motion. This configuration allows us to generate a large scale, temporally dense dataset for training and evaluating both the keyframe detection and depth estimation models.

The real-world segment of *ELiDAR* is collected with our *NeuroLiDAR* prototype, mounted on a robotic platform which traversed different trajectories, at varying speeds, within an indoor room environment. The LiDAR was configured to operate at 30 Hz, the maximum frame rate supported by the Intel RealSense L515. This relatively small split helps assess the generalizability of *NeuroLiDAR* from simulated outdoor scenarios to real-world indoor settings.

Figure 1 shows representative samples from the *ELiDAR* dataset, including scenes captured under both simulated outdoor and real-world indoor settings. In total, simulated splits of *ELiDAR* contain 244 separate 20s long sequences, captured under 12 predefined weather conditions. The real-world indoor split contains 15 additional 10s long sequences. *ELiDAR* fills a key gap in the availability of real-world high-frame-rate LiDAR data and provides a benchmark for both keyframe detection and adaptive depth estimation.

B. Motivation for a High Frame Rate LiDAR



(a) Event change over time (b) Depth change over time

Fig. 2: Event and depth changes with time for varying ego-sensor speeds

Commercial automotive LiDARs typically operate at frame rates between 5-10Hz, depending on the sensor's range and resolution. In principle, their frame rate can be increased, but this comes with trade-offs: higher frame rates reduce effective range and resolution, substantially increase power consumption, and may shorten the lifespan of the sensor due to limitations of the laser diode. These limitations render current commercial LiDARs inadequate for reliably capturing rapid scene changes in highly dynamic environments.

Analysis of *ELiDAR* highlights the temporal limitations of current LiDAR systems. We measured the change in depth over different time intervals and vehicle velocities (Figure 2b). For a 100 ms interval (corresponding to 10 Hz, a common frame rate for outdoor LiDARs such as Velodyne [20]), we observed an average depth change of approximately 2.4 m when the vehicle was moving at $\geq$ 50 km/h, a typical driving speed. Such changes exceed the spatial accuracy needed for reliable perception, revealing that these LiDARs are insufficient for fast-moving scenarios.

C. Motivation for an Adaptive Frame Rate LiDAR

While we have established the need to expand LiDAR capabilities to support higher frame rates, not all driving conditions require high temporal fidelity. For instance, as shown in Figure 2b, when the ego vehicle moves at $\leq$ 10 km/h, the average change in depth is $\leq$ 1 m, making high-frequency depth sampling unnecessary (compared to when the vehicle travels at $\geq$ 50 km/h). A fixed high frame rate would therefore waste energy, highlighting the need for an *adaptive frame rate LiDAR* that responds to scene dynamics.

Achieving adaptivity requires a lightweight sensing mechanism to infer scene changes without activating the high-power LiDAR sensor. High frame rate RGB cameras could serve this purpose, but their power demands[1] ($\approx$3.3 W) make them impractical for embedded deployment. In contrast, neuromorphic event cameras offer ultra-high temporal resolution ($\approx$1 MHz) at low power ($\approx$0.5 W). By asynchronously capturing pixel-level brightness changes, event cameras are particularly well-suited to directly capture motion-driven scene changes with low latency, high sensitivity, and low energy overheads. Such detected changes can, in turn, trigger explicit depth sensing by an adaptive frame rate LiDAR.

We conducted a preliminary analysis on *ELiDAR* to evaluate whether event sensor outputs can effectively capture scene dynamics (Figure 2a). Specifically, we vary the ego vehicle's speed and compute the fraction of active event pixels relative to the total (640×480) pixel count. Across increasing time windows and vehicle speeds, we consistently observe that the proportion of active event pixels rises, mimicking the trend for depth changes reported in Figure 2b. *This demonstrates that event representations can reliably capture motion dynamics in the scene and exhibit strong correlation with depth changes, making them well-suited as a triggering mechanism for adaptive depth estimation.*

IV. *NeuroLiDAR* System Design

We now outline the design goals, high-level architecture, and implementation of our *NeuroLiDAR* system.

A. Design Goals

- **Adaptive Frame Rate:** The system should dynamically adjust its frame rate in response to perceived changes in scene dynamics. In particular, *NeuroLiDAR* must support a maximum effective frame rate that exceeds the native frame rate of the baseline LiDAR sensor.
- **Low Computational Complexity:** To sustain high effective frame rates when needed, the depth estimation process

[1]https://www.shodensha.co.th/product/183316-173555/chu30

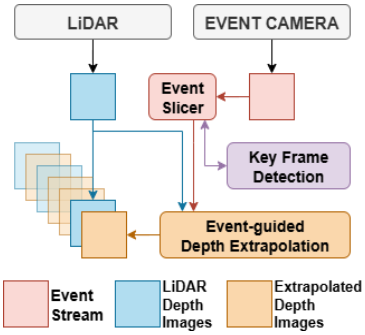


Fig. 3: High-Level Architecture

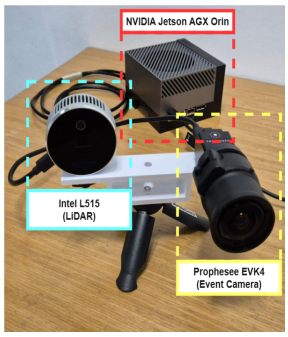


Fig. 4: *NeuroLiDAR*'s System Implementation

must be lightweight enough for real-time execution on an embedded device, such as the Jetson platform. Specifically, the inference latency l must satisfy $l \leq 1/F$, where F is *NeuroLiDAR*'s current operating frame rate.

- **Reduced Depth Extrapolation Error:** Despite being lightweight, the model should minimize depth extrapolation error across a range of frame rates, ensuring robustness without sacrificing accuracy.

B. System Architecture

We develop *NeuroLiDAR* by coupling an event camera with a base, low-frame-rate LiDAR. The event stream captured by the neuromorphic camera serves two complementary roles: (a) identifying the moments when new depth frames should be generated (i.e., the reference LiDAR depth frames should be augmented via extrapolation), and (b) providing motion priors for the actual depth extrapolation process. Figure 3 presents the high-level system architecture, comprising three key components:

1) **Event Slicer:** Processes the raw event stream into structured event representation(s) that can be used by both the keyframe detection and depth extrapolation modules.
2) **Keyframe Detection:** Determines whether an event slice corresponds to a significant scene change and, if so, triggers the depth extrapolation module via Event Slicer.
3) **Event-guided Depth Extrapolation:** Combines the most recent LiDAR depth image with accumulated events since its capture to generate an extrapolated depth image.

Keyframe detection and depth extrapolation are performed by two separate DNN models, both running concurrently on a representative low-power device (Jetson AGX Orin), resulting in an adaptive frame rate *NeuroLiDAR*. In addition, to balance accuracy vs. computational efficiency, *NeuroLiDAR* uses two distinct event representations: (i) the computationally cheaper, less descriptive event frame representation for the keyframe detector, and (ii) the more descriptive voxel grid representation [21] for Event-guided Depth Extrapolation.

C. System Implementation

We implement the *NeuroLiDAR* prototype (illustrated in Figure 4) using an Intel RealSense L515 [22] as the low-frame-rate LiDAR sensor and a Prophesee EVK4 [23] as the event camera. The keyframe detection and depth extrapolation models are deployed on an NVIDIA Jetson AGX Orin [24]. To support real-time operation with reduced energy consumption, both models are quantized to 16-bit FP precision and executed using TensorRT.

V. Adaptive Event-Guided Depth Extrapolation

We now detail *NeuroLiDAR*'s key functions. For a formal description, let the depth frame captured at time t be denoted as $D_t \in \mathcal{R}^{H\times W}$, where H and W represent the height and width of the depth map, respectively. We define an asynchronous event stream within a time interval of $(t, t+\Delta)$ as $\mathcal{E}_{(t,t+\Delta)} = \{e_i = (p_i, x_i, y_i, t_i)\}_{i=0}^{N}$ where $p_i \in \{-1, 1\}$ indicates the polarity corresponding to a brightness increase or decrease, $\{x_i, y_i\} \in \mathcal{R}^{H\times W}$ denotes the pixel coordinates, and t_i denotes the timestamp.

A. Event Slicer

The event slicer accumulates asynchronously arriving events and converts them into structured event representations. Concretely, for an asynchronous event stream $\mathcal{E}_{(t,t+\Delta)}$, an event frame $\mathcal{E}^F(x, y)$ (consumed by the Event Slicer) is defined as follows:

$$\mathcal{E}^F(x, y) = \sum_{e_i \in \mathcal{E}_{(t,t+\Delta)}} p_i \mathbf{1}\{x = x_i, y = y_i\} \tag{1}$$

In our design, we set $\Delta = 20$ms and pass this *event frame* representation to the keyframe detection model, which processes the frames and triggers the depth extrapolator whenever it identifies a *significant change* in the scene.

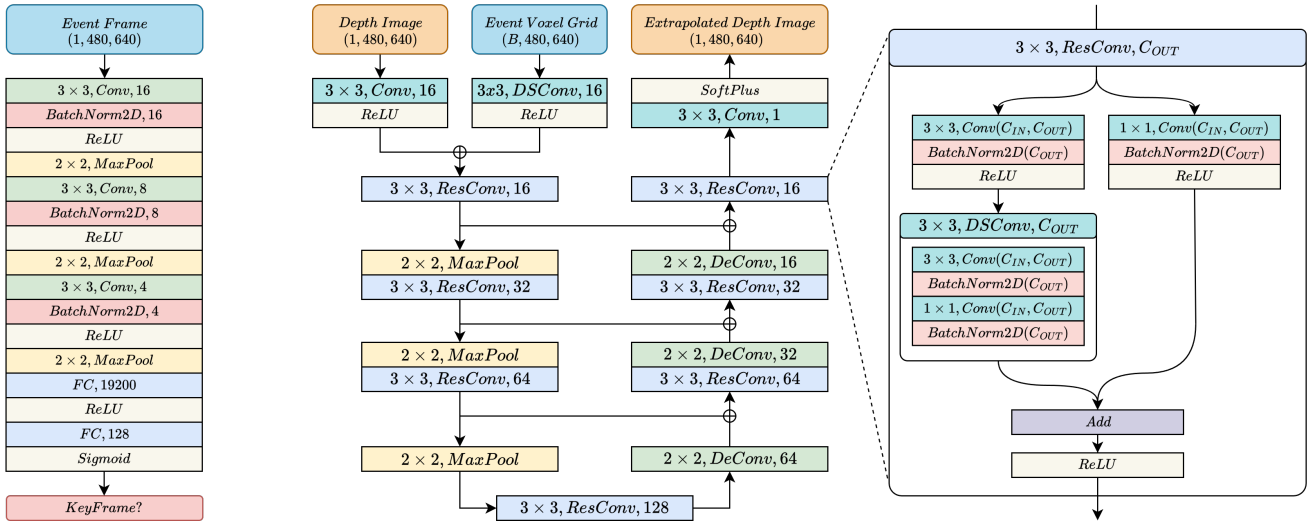
Assuming that a keyframe is detected at $t = t+t_1$, we then obtain an event voxel $\mathcal{E}^V(b, x, y)$ (consumed by the Event-guided Depth Extrapolation component) where $b \in [0, B)$ for $B = 5$ as follows.

$$\mathcal{E}^V(b, x, y) = \sum_{e_i \in \mathcal{E}_{(t,t+t_1)}} p_i \mathbf{1}\{b = \left\lfloor B \cdot \tfrac{t_i - t}{t_1} \right\rfloor, x = x_i, y = y_i\} \tag{2}$$

B. Keyframe Detection Model

A core feature of *NeuroLiDAR* is its ability to adaptively regulate frame rate based on scene dynamics, delivering high temporal fidelity in fast-changing environments while conserving energy when the environment is largely static. To enable this, we design a lightweight keyframe detection model that monitors incoming event streams and identifies moments of significant change, such as ego-motion or the appearance of moving objects (e.g., vehicles, pedestrians). Only when such keyframes are detected is the depth extrapolation module triggered, ensuring that new depth frames are generated precisely when needed rather than periodically.

*1) **Characterizing a Keyframe**:* We define a keyframe as one that captures a significant scene change, typically arising in outdoor driving from dynamic objects such as vehicles and pedestrians and in indoor environments from unexpected motion of human occupants and other objects (e.g., robots, machinery). To operationalize this, we extract contextual cues from event sub-streams and use them to train our DNN



(a) *Keyframe Detection Model* (b) *Depth Extrapolation Model*

Fig. 5: Neural network architectures of keyframe detection and depth extrapolation network with expanded view of ResConv

model. We employ a threshold-based strategy guided by three indicators: ego-sensor speed, distance to surrounding objects, and the appearance of new objects in the field of view. For ground truth labelling, a keyframe is defined when any of the following conditions are satisfied:

- The ego sensor speed exceeds 10 m/s (=36 km/h).
- The shortest distance to a surrounding object (such as a vehicle or pedestrian) falls below 8 meters.
- A new object enters the sensor's field of view.

*2) **Architecture of the Keyframe Detection Model**:* We formulate keyframe detection as a binary classification problem, where each event frame is classified as either a keyframe or a non keyframe. We design a lightweight three-layer convolutional neural network (illustrated in Figure 5(a)) that takes an event frame defined in equation 1 as input and outputs a binary classification result. The model is trained using the binary cross entropy loss function for 20 epochs with the NAdam optimizer, batch size of 64, and a learning rate of 1×10^{-4}. The dataset contains 24,720 training and 11,220 test samples, randomly drawn from 114 training and 50 test sequences in the simulated outdoor segment.

C. Event-guided Depth Extrapolation

Once the keyframe detector predicts a significant change in the scene at time $t = t_1$, we then activate the event-guided depth extrapolation model. This model takes two inputs: (a) a prior depth image (D_t), which is the most recent depth image captured by the LiDAR, and (b) the event voxel representation from Event Slicer (Equation 2) in the time interval (t, t_1). The events provide motion information relative to the prior depth image and guide the extrapolation task. To satisfy the real-time requirements of our system, we design the model architecture to be lightweight and consistent with the keyframe detection model, ensuring that each depth frame is predicted before the arrival of the next frame. The model is a streamlined U-Net architecture, illustrated in Figure 5(b), composed of convolution, max-pooling, and transposed convolution operations followed by ReLU activations. The final output layer, which generates the depth image, applies a SoftPlus activation. The architecture follows an encoder–decoder structure that employs max-pooling and transposed convolution to downsample and upsample the spatial dimensions of the feature maps. The intermediate convolutional layers consist of two parallel convolution operations (3x3 and 1x1): the 1×1 convolution layer extracts features from the input in parallel to 3×3 convolution, and its output feature maps are added to those of the depthwise separable convolution before the final activation.

For training, we adopt a combination of loss functions $(\mathcal{L}_{depth}, \mathcal{L}_{grad}, \mathcal{L}_{norm}, \mathcal{L}_{SSIM})$, employed in METER [25]. For $\mathcal{L}_{depth}$, we use MSE loss instead of L1, with weighting factors $\lambda_2 = 10, \lambda_3 = 0.01, \lambda_4 = 1$. Our model is trained for 50 epochs using the NAdam optimizer with an initial learning rate of 1×10^{-3} and a cosine annealing scheduler. The training dataset consists of 22,543 training samples and 5,597 test samples, randomly drawn from 80 training sequences and 16 test sequences within the simulated outdoor segment.

VI. Results

We now evaluate the *NeuroLiDAR*'s performance on both the simulated outdoor and real indoor segments of *ELiDAR*.

A. Evaluation Metrics

Following prior work on depth estimation [26], we adopt standard depth evaluation metrics to benchmark the performance of *NeuroLiDAR*. In addition to accuracy based measures, we also evaluate system level efficiency by reporting the inference latency and, retrospectively, the maximum effective frame rate that *NeuroLiDAR* can achieve on an NVIDIA Jetson AGX Orin. To benchmark depth extrapolation accuracy, we use the following standard metrics:

1) RMSE: $\sqrt{\frac{1}{|D|}\sum_{d\in D}(d-\hat{d})^2}$

2) log RMSE: $\sqrt{\frac{1}{|D|}\sum_{d\in D}(\log d - \log \hat{d})^2}$

TABLE I: *NeuroLiDAR*'s keyframe detection accuracy (*ELiDAR* Simulated)

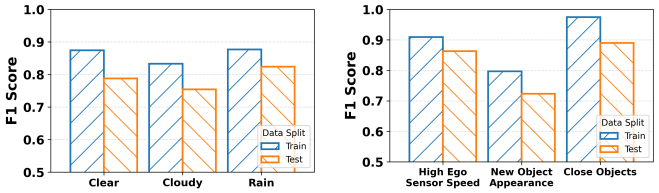
Split	F1-Score	Precision	Recall
Train	0.888	0.893	0.884
Test	0.798	0.825	0.773

3) AbsRel: $\frac{1}{|D|}\sum_{d\in D}\frac{|d-\hat{d}|}{d}$

4) SQRel: $\frac{1}{|D|}\sum_{d\in D}\frac{(d-\hat{d})^2}{d}$

5) δ: % of $\hat{d}$ s.t. $\max\left(\frac{d}{\hat{d}},\frac{\hat{d}}{d}\right) < \delta$, for $\delta \in 1.25, 1.25^2, 1.25^3$

Additionally, we adopt standard classification metrics: Precision, Recall, and F1-score, to evaluate the efficacy of the keyframe detection model. Unless specified otherwise, all evaluations of *NeuroLiDAR* are conducted on the synthetic outdoor driving segment of *ELiDAR*.



(a) Different weather conditions (b) Keyframe characteristics

Fig. 6: Keyframe detection under different conditions

B. Performance Evaluation of Keyframe Detection Model

Table I summarizes the performance of the keyframe detection model across the simulated segment of *ELiDAR* dataset. On the test dataset, our model achieves an F1-score of 79.8%. The temporal gap between a false positive and either a true positive or the next actual LiDAR depth image is observed to be low, varying between 41.7–88.6 ms, underscoring that the keyframe detection module can reliably identify the optimal moments to trigger depth estimation (with a fairly tight error bound) while remaining lightweight and computationally efficient. Importantly, this efficiency ensures that extrapolated depth frames can be produced in real time, thereby maintaining the high and adaptive frame rates required to handle fast and dynamic motion sequences.

In Figures 6a and 6b, we further analyze the F1-score of keyframe detection under different weather conditions and motion/object dynamics. Our model achieves strong performance in detecting high-speed motion and nearby objects, reaching F1-scores of 86% and 89% respectively. The performance slightly decreases when identifying newly appearing objects (F1-score=72%). This reduction may stem from the insufficient temporal span across frames. Nevertheless, the lightweight keyframe detector consistently delivers F1-scores between 75.4% and 82.8% across diverse weather conditions, demonstrating both robustness and effectiveness.

C. Performance Evaluation of Depth Extrapolation Model

We compare baseline extrapolation methods with our event-guided depth extrapolation model in Table II. The effective frame rate F of *NeuroLiDAR* is also shown, where extrapolation raises the base LiDAR rate from $F/2$ to F. To ensure fair evaluation under fixed conditions, we test across fixed frame rates ranging from 2–50 Hz. In the adaptive (*Adap.*) setting, we randomly sample frames from the dataset at varying frame rates, with the model again extrapolating a single intermediate frame. For both fixed and adaptive settings, we compare our depth extrapolation model against a *Repeat* baseline, which simply reuses the sampled depth frame at $F/2$ as the extrapolated frame. Across all metrics, our event-guided extrapolation consistently outperforms this baseline, demonstrating robustness under both static and adaptive frame rate conditions, e.g., comparing *Adap.* setting, our approach has 31.43% lower RMSE than *Repeat*.

TABLE II: Depth estimation error for different frame rates: *NeuroLiDAR* vs. other baselines

Method	FPS (F)	RMSE ↓	Log RMSE ↓	ABS Rel ↓	SQ Rel ↓	$\delta<$ 1.250 ↑	$\delta<$ 1.562 ↑	$\delta<$ 1.953 ↑
Repeat	2	12.211	3.487	0.139	4.521	0.860	0.911	0.933
	5	9.820	2.980	0.092	2.603	0.910	0.940	0.954
	10	8.382	2.373	0.063	1.884	0.941	0.960	0.970
	20	6.904	2.016	**0.045**	1.341	0.958	0.971	0.978
	50	4.976	1.486	**0.026**	0.697	**0.975**	0.984	0.987
	Adap.	8.414	2.455	0.072	2.187	0.929	0.953	0.964
Ours	2	**8.422**	**0.528**	**0.109**	**2.075**	**0.885**	**0.944**	**0.967**
	5	**6.385**	**0.394**	**0.074**	**1.010**	**0.929**	**0.967**	**0.981**
	10	**5.542**	**0.292**	**0.055**	**0.689**	**0.953**	**0.979**	**0.988**
	20	**4.747**	**0.237**	0.046	**0.500**	**0.963**	**0.984**	**0.991**
	50	**3.883**	**0.185**	0.036	**0.325**	0.974	**0.989**	**0.994**
	Adap.	**5.769**	**0.325**	**0.063**	**0.910**	**0.941**	**0.972**	**0.984**

TABLE III: Ablation results on input type, model variants, and loss functions.

Method	RMSE ↓	Log RMSE ↓	ABS Rel ↓	SQ Rel ↓	$\delta<$ 1.25 ↑	$\delta<$ 1.56 ↑	$\delta<$ 1.95 ↑
Input-based Ablations							
w/o Event	7.799	0.364	0.103	1.693	0.911	0.952	0.967
Depth + Ev. frames	6.161	0.332	0.071	1.082	0.933	0.968	0.981
Model Ablations							
Data concat	5.936	**0.315**	0.073	0.954	0.936	0.970	0.983
w/o skip	5.935	0.392	0.079	0.985	0.931	0.969	0.983
Ours	**5.769**	0.325	**0.063**	**0.910**	**0.941**	**0.972**	**0.984**

D. Ablation Studies for Event-Guided Depth Extrapolation

Table III presents a detailed ablation study of *NeuroLiDAR*'s depth extrapolation model, considering two dimensions: (a) input-based and (b) model-based. Below, we describe the variants and their effects.

- **w/o Event:** The event encoder is removed, leaving the model without motion priors for extrapolation. This leads to a substantial degradation in performance across all metrics, increasing RMSE by +2.03 compared to the full model.
- **Depth + Ev. frames:** Instead of voxel-grid representation, we use event frames (the same representation employed in the keyframe detection model). While this representation provides a coarse motion prior, it lacks the fine-grained temporal resolution of voxel grids, resulting in moderately higher errors, with RMSE increasing by +0.392.
- **Data concat:** Our primary model encodes events and depth inputs separately, followed by feature-level fusion. This variant concatenates the events and depth as a single input and passes it through a single encoder. This results in a performance drop (RMSE +0.167), showing the importance of specialized encoders for different modalities.

TABLE IV: End-to-end depth estimation comparison

Method	RMSE ↓	Log RMSE ↓	ABS Rel ↓	SQ Rel ↓	$\delta <$ 1.25 ↑	$\delta <$ 1.56 ↑	$\delta <$ 1.95 ↑
Repeat	6.226	1.912	**0.045**	1.210	0.956	0.971	0.956
Linear	8.248	2.758	0.075	1.992	0.926	0.952	0.963
Exponential	25.131	3.589	0.454	10.683	0.255	0.520	0.622
NeuroLiDAR	**4.416**	**0.245**	0.047	**0.457**	**0.962**	**0.984**	**0.991**

• **w/o skip:** This variant measures the impact of 1×1 convolution layer as a skip connection in ResConv Blocks. By removing it, the performance drops across all metrics, with RMSE increasing by +0.166, confirming their role in preserving spatial detail during depth extrapolation.

E. End-to-End System Evaluation

We now conduct an end-to-end system evaluation, including both the keyframe detection and depth extrapolation tasks on the test sequences of *ELiDAR*. In this experiment, we set the low baseline frame rate of LiDAR to 10 Hz. We first compare the performance against several baselines (reported in Table IV):

• **Repeat:** Most recent depth image is repeated as the extrapolated frame at the keyframe instance.

• **Linear:** Assuming the depth varies linearly over time, we regress the pixel-wise depth value temporally with a linear function ($d_{t+\Delta} = ad_t + b$), where both a and b coefficients are determined by previous depth images

• **Exponential:** Similarly, we assume depth varies as an exponential function ($d_{t+\Delta} = ae^{\Delta d_t}$), where a is estimated with the last two consecutive depth images.

When compared to these baselines, *NeuroLiDAR* reduces the RMSE by ≈1.81 (relative to the repeat baseline), achieving lower error across the evaluation metrics, while reaching a maximum frame rate of **47.30 Hz** (a **4.7**× increase over the standard 10 Hz LiDAR). On the *ELiDAR* dataset, *NeuroLiDAR* adapts its frame rate between **27.81 Hz** and **47.30 Hz**, with a mean effective frame rate of **40.58 Hz**.

In Figure 7, we further analyze the end-to-end latency of the system, which is determined by three main components: the event slicer, the keyframe detector, and the event-guided depth extrapolator. For a voxel-grid with a time window of 20 ms, *NeuroLiDAR* can theoretically operate within 15 ms (supporting up to 66.67 Hz), which is $\sim 6.67\times$ the operating frequency of the LiDAR (10 Hz). While the latencies of the keyframe detector (2.81 ms) and the event-guided depth extrapolator (9.31 ms) remain constant across different voxel sizes, the latency of the event slicer depends on the event window size–i.e., the duration until the keyframe detector signals a significant scene change. A larger window size such as 500 ms means that the *NeuroLiDAR* needs to operate only at 2 Hz due to a largely static scene. Even then, end-to-end latency of the system is **79.6 ms**, which is far lower than the required 500 ms time window. *NeuroLiDAR* can thus operate at a lower effective frame rate in static scenes, thereby allowing more time for frame extrapolation.

F. Generalizability

In addition to evaluation on the synthetic outdoor *ELiDAR* dataset, we demonstrate the generalizability of *NeuroLiDAR*

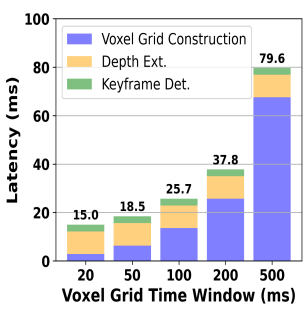


Fig. 7: End-to-end latency including keyframe detection, voxel-grid construction, and depth extrapolation.

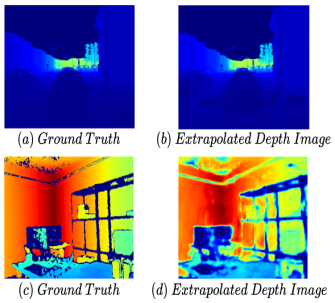


Fig. 8: Ground truth vs. extrapolated depth images: (a–b) simulated outdoor, (c–d) real-world indoor.

TABLE V: *NeuroLiDAR* depth estimation error for the real-world indoor scenario

Method	RMSE ↓	Log RMSE ↓	ABS Rel ↓	SQ Rel ↓	$\delta <$ 1.25 ↑	$\delta <$ 1.56 ↑	$\delta <$ 1.95 ↑
Repeat	0.819	4.213	0.068	0.207	**0.939**	0.942	0.944
NeuroLiDAR	**0.532**	**0.316**	**0.066**	**0.087**	0.903	**0.948**	**0.970**

for the real indoor dataset, collected using our prototype (Figure 4). Table V presents the resulting accuracy of *NeuroLiDAR*'s event-guided depth extrapolation model. For this, we finetuned the depth extrapolation model on the training split of the real-world subset of *ELiDAR* for 50 epochs with a learning rate of 1×10^{-4} by freezing the encoder's parameters and scaling the sensor's range from 200m to 9m.

Testing involved adaptively sampling depth frames at effective frame rates of up to 15 Hz and extrapolating a single intermediate frame from the test split of the real-world indoor dataset. Compared to the *Repeat* baseline, *NeuroLiDAR* achieves substantial gains across most accuracy metrics, including a 28.61% reduction in RMSE. While these results demonstrate promising generalization to indoor environments, additional large-scale evaluation in indoor environments such as warehouses and factory floors will be pursued as future work. We do not assess the generalizability of the keyframe detection module, as this scenario differs fundamentally from outdoor driving conditions.

VII. Discussion

Increasing the Spatial Fidelity of LiDARs: *NeuroLiDAR* has primarily focused on improving the temporal resolution of LiDAR sensing by adaptively extrapolating depth frames. While this effectively increases the frame rate up to 65+ Hz, the challenge of simultaneously enhancing spatial fidelity, i.e., increasing the point cloud resolution of LiDAR outputs, remains open. We speculate that the same DNN-based depth extrapolation framework presented here can be extended to address spatial resolution, applying techniques analogous to the use of U-Net-based autoencoders for event-guided depth densification [4], [8]. In future, we plan to explore how U-Net autoencoders can be adapted to jointly support both high temporal and spatial fidelity within a single network.

Adaptive Bin Size for Event-Voxel Grids: To further enhance the fidelity of spatiotemporal depth extrapolation,

NeuroLiDAR can potentially utilize alternative non-uniform voxelization techniques, borrowed from point cloud compression literature, to create non-uniform event voxel grids. For example, the most recently constructed depth map may be used as a *prior* to create finer voxels in regions characterized by sharp depth variations; such approaches may be especially useful for higher depth resolution required for fine-grained robot manipulation. However, such non-uniform voxelization requires further investigation for careful optimization, as it is likely to increase the processing latency of the *Event Slicer*.

VIII. Conclusion

We presented *NeuroLiDAR*, a novel LiDAR depth sensing system that combines a standard LiDAR with a neuromorphic event camera to achieve higher temporal fidelity through adaptive frame rates supporting up to 66.67 Hz. *NeuroLiDAR* continuously uses the event stream to perceive scene changes and then triggers the depth sensing module to extrapolate depth frames at appropriate time instants. For depth extrapolation, we use a lightweight U-Net-style autoencoder to fuse a prior depth frame captured by the standard LiDAR with a motion representation from the high frame rate event camera. Evaluation performed using a new benchmark *ELiDAR* dataset, which models both diverse urban driving scenarios and a smaller, real-world indoor environment, shows that *NeuroLiDAR* reduces depth reconstruction error by ≈30% compared to a standard LiDAR operating at 10 Hz.

Acknowledgment

This work was supported in part by: 1) National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program. The Mens, Manus, and Machina (M3S) is an interdisciplinary research group (IRG) of the Singapore-MIT Alliance for Research and Technology (SMART) centre; and 2) the Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2 grant (Grant ID: T2EP20124-0055). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the granting agencies or the university.

References

[1] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, p. 154–180, 2022.

[2] L. Sun, C. Sakaridis, J. Liang, P. Sun, K. Zhang, J. Cao, Q. Jiang, K. Wang, and L. Van Gool, "Event-Based Frame Interpolation with Ad-hoc Deblurring." IEEE, June 2023, pp. 18 043–18 052.

[3] M. G. J. H.-C. Daniel Gehrig, Michelle Rüegg and D. Scaramuzza, "Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction," *IEEE Robotic and Automation Letters. (RA-L)*, 2021.

[4] M. Cui, Y. Zhu, Y. Liu, Y. Liu, G. Chen, and K. Huang, "Dense depth-map estimation based on fusion of event camera and sparse lidar," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[5] Y. Lu, Z. Wang, M. Liu, H. Wang, and L. Wang, "Learning spatial-temporal implicit neural representations for event-guided video super-resolution," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1557–1567.

[6] J. Cuadrado, U. Rançon, B. R. Cottereau, F. Barranco, and T. Masquelier, "Optical flow estimation from event-based cameras and spiking neural networks," *Frontiers in Neuroscience*, vol. 17, p. 1160034, 2023.

[7] B. Li, H. Meng, Y. Zhu, R. Song, M. Cui, G. Chen, and K. Huang, "Enhancing 3-d lidar point clouds with event-based camera," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[8] V. Brebion, J. Moreau, and F. Davoine, "Learning to estimate two dense depths from lidar and event data," in *Scandinavian Conference on Image Analysis*. Springer, 2023, pp. 517–533.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[11] J. Chen, J. Meng, X. Wang, and J. Yuan, "Dynamic graph cnn for event-camera based gesture recognition," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2020, pp. 1–5.

[12] Q. Liu, D. Xing, H. Tang, D. Ma, and G. Pan, "Event-based action recognition using motion information and spiking neural networks." in *IJCAI*, 2021, pp. 1743–1749.

[13] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3d reconstruction and 6-dof tracking with an event camera," in *European conference on computer vision*. Springer, 2016, pp. 349–364.

[14] W. Guan, P. Chen, H. Zhao, Y. Wang, and P. Lu, "Evi-sam: Robust, real-time, tightly-coupled event–visual–inertial state estimation and 3d dense mapping," *Advanced Intelligent Systems*, vol. 6, no. 12, p. 2400243, 2024.

[15] X. Wu, W. He, M. Yao, Z. Zhang, Y. Wang, B. Xu, and G. Li, "Event-based depth prediction with deep spiking neural network," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.

[16] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.

[17] P. Shi, J. Peng, J. Qiu, X. Ju, F. P. W. Lo, and B. Lo, "Even: An event-based framework for monocular depth estimation at adverse night conditions," in *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2023, pp. 1–7.

[18] S. A. Tumpa, A. Devulapally, M. Brehove, E. Kyubwa, and V. Narayanan, "Snn-ann hybrid networks for embedded multimodal monocular depth estimation," in *2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2024, pp. 198–203.

[19] D. G. Javier Hidalgo-Carrio and D. Scaramuzza, "Learning monocular dense depth from events," *IEEE International Conference on 3D Vision.(3DV)*, 2020.

[20] Ouster, "Vlp16: mid range lidar sensor," https://ouster.com/products/hardware/vlp-16, accessed:2025-09-14.

[21] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, pp. 154–180, 2020.

[22] I. Corporation, "Intel® realsense™lidar camera l515," https://www.intel.com/content/www/us/en/products/sku/201775/intel-realsense-lidar-camera-l515/specifications.html, accessed: 2025-09-14.

[23] M. by Prophesee, "Evk4: The ultra high-speed and compact, hd event-based vision evaluation kit built to endure field testing conditions." https://www.prophesee.ai/event-camera-evk4/, accessed: 2025-09-14.

[24] NVIDIA, "Nvidia jetson orin," https://www.nvidia.com/en-sg/autonomous-machines/embedded-systems/jetson-orin/, accessed: 2025-09-14.

[25] L. Papa, P. Russo, and I. Amerini, "Meter: A mobile vision transformer architecture for monocular depth estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 10, pp. 5882–5893, 2023.

[26] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow, "Digging into self-supervised monocular depth estimation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3828–3838.