Lidar Waveforms are Worth 40x128x33 Words

Dominik Scheuble[1,4] Hanno Holzhüter[2] Steven Peters[4] Mario Bijelic[3,5] Felix Heide[3,5]
[1]Mercedes-Benz AG [2]MicroVision [3]Torc Robotics [4]TU Darmstadt [5]Princeton University

Abstract

Lidar has become crucial for autonomous driving, providing high-resolution 3D scans that are key for accurate scene understanding. To this end, lidar sensors measure the time-resolved full waveforms from the returning laser light, which a subsequent digital signal processor (DSP) converts to point clouds by identifying peaks in the waveform. Conventional automotive lidar DSPs process each waveform individually, ignoring potentially valuable context from neighboring waveforms. As a result, lidar point clouds are prone to artifacts from low signal-to-noise ratio (SNR) regions, highly reflective objects, and environmental conditions like fog. While leveraging neighboring waveforms is investigated extensively in transient imaging, applications remain limited to scientific or experimental hardware. In this work, we propose a learned DSP that directly processes full waveforms using a transformer architecture, leveraging features from adjacent waveforms to generate high-fidelity multi-echo point clouds. To assess our method, we capture data in real-world driving scenarios and a weather chamber with a conventional automotive lidar. Trained on synthetic and real data, the method improves Chamfer distance by 32cm and 20cm compared to conventional peak finding and existing transient imaging approaches, respectively. This translates to maximum range improvements of up to 17m in fog and 14m in nominal real-world conditions.

1. Introduction

Scanning lidar sensors have become a cornerstone sensing modality in autonomous driving [5, 9, 70], offering high spatial and temporal resolution for accurate 3D scene understanding. Recent advancements in lidar technology, combining solid-state scanning with single-photon avalanche diodes (SPADs) [48], have substantially reduced both cost and form factor, enabling widespread adoption across different types of autonomous vehicles [26]. These sensors operate by emitting multiple laser pulses into the scene to acquire the time-resolved full waveform of the returning light. Subsequently, a digital signal processor (DSP) pipeline detects peaks in the waveforms and returns a 3D point cloud. Currently, automotive lidar manufacturers rely on analytic DSP pipelines that process each measured waveform individually [2, 31]. These pipelines are fine-tuned through a combination of visual inspection and quality metrics, enabling accurate reconstruction in high signal-to-noise ratio (SNR) areas. However, they struggle in challenging conditions, such as object discontinuities - where diverging beams interact with multiple surfaces (Fig. 1a)) - or distant low-reflectivity targets (Fig. 1b)), both of which cause conventional DSPs to miss peaks. As shown in Fig. 1c), retroreflective materials generate false positive points from blooming [36, 80], distorted waveforms resulting in false distance estimates [52], and multipath effects from reflections off the sensor cover glass. Scattering media, such as fog [4, 8, 19, 30], rain [8, 16, 82], and snow [20, 30, 34], cause falsely detected peaks while heavily attenuating actual object peaks, see Fig. 1d).

To tackle these adverse conditions, waveform shape features can be exploited to overcome the limitations of conventional DSPs. Moreover, processing waveforms jointly, instead of individually, enables DSPs to leverage neighborhood information. Spatial neighborhoods for 3D reconstruction of waveform data have been used extensively in transient imaging, where existing work can be roughly grouped into optimization-based and learned data-driven approaches. Optimization-based approaches reconstruct scenes by formulating a waveform lidar imaging model and solving the inverse problem [63]. To this end, they derive complex imaging models to account for different effects such as multiple returns [69, 75], pulse broadening [74], or pile-up [23]. To solve the ill-posed inverse problem, they rely heavily on scene priors or regularization terms [63]. However, their model assumptions and priors fail on large outdoor scenes due to complex scene geometry and fluctuating scene parameters caused by, e.g., environmental effects. Furthermore, their long runtimes, up to many hours [51, 75], prevent a real-time application. On the other hand, existing learned data-driven methods [39, 55, 73, 76] operate only on patches instead of the full spatial resolution in order to handle the uncompressed full waveform data. This limits their ability to capture global context and introduces long run times. Moreover, they reconstruct only a single distance per waveform,

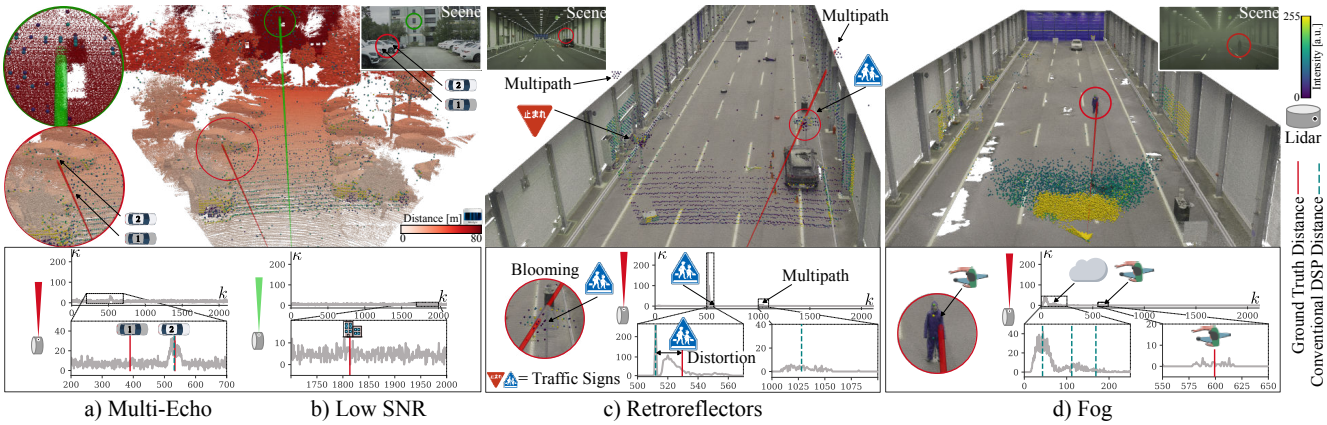

Figure 1. **Sensing Challenges for Automotive SPAD Lidar.** Intensity-coded point clouds from conventional on-device processing in a) multi-echo, b) low SNR, c) retroreflective, and d) foggy conditions, compared to ground truth. While conventional lidar DSPs process waveforms κ individually, joint long-range transient waveform processing can mitigate these challenging conditions by exploiting their waveform shape features and adjacency information. Ground truth is obtained from a) accumulated Velodyne scans for outdoor scenarios, and (b,c) RGB-colored Leica scans for scenarios in the weather chamber.

which is often insufficient for automotive lidar applications that frequently encounter multiple returns [45]. Common to all existing transient imaging methods is their reliance on scientific and experimental hardware with a limited field of view (FoV), suited for controlled lab environments with few objects [40, 59, 71] or extremely focused, long-range applications over many kilometers [37, 46, 53]. In contrast, autonomous driving demands larger FoVs, ranges up to 200m, and real-time capability.

Addressing these challenges, we introduce a novel neural DSP that learns to generate high-fidelity point clouds from waveforms of a production-grade full waveform lidar (FWL). Inspired by the success of transformers [14, 81], we design our neural DSP with a transformer architecture that processes all waveforms simultaneously, enabling it to capture both temporal and spatial context. We demonstrate that our proposed DSP is capable of mitigating environmental effects, such as fog, blooming, and distance distortions, which are challenging for traditional methods in both outdoor driving scenarios and a weather chamber with distances of up to 85 m. Additionally, we find that our neural DSP can be easily extended to generate super-resolution point clouds by exploiting waveform cues, outperforming traditional point cloud-based super-resolution methods. Specifically, we make the following contributions

- We devise a neural LiDAR DSP leveraging adjacent waveform features for improved peak detection, extendable to predict super-resolved point clouds with 9× more points.
- We introduce a realistic FWL simulation model for CARLA [13] to generate training data at large-scale.
- We validate our method using a real-world FWL dataset we captured, observing a 32 cm Chamfer distance improvement in 3D reconstruction and a up to 17 m increase in sensor range compared to prior approaches.

2. Related Work

Although existing transient imaging methods are designed for experimental hardware, lidar DSPs can be seen in the context of transient imaging. Tab. 1 compares our work to related transient imaging methods.

Classical DSPs. Conventional automotive lidar DSPs run on-device with limited memory, significantly compressing waveforms to reduce transfer rates [42, 82]. These DSP pipelines are often proprietary but typically involve matched filtering [63, 79] and peak or leading edge detection [31] on individual waveforms to estimate distances. Spatial effects like blooming are corrected at the point cloud level [15]. While efficient, these pipelines often fail in challenging low SNR or scattering conditions as they do not consider neighboring waveforms. Conversely, optimization-based transient imaging methods effectively reconstruct distant objects [37, 46, 53], low-SNR scenes [32, 51] or improve reconstruction in scattering media [22, 74]. These methods construct imaging models and then solve the inverse problem using maximum likelihood estimation [77] or Monte Carlo processing [6] to reconstruct the scene. However, these methods are computationally expensive and unsuitable for real-time automotive sensing, as integrating various complex scene properties and sensor non-linearities becomes increasingly intractable. Additionally, they rely on hand-crafted spatial regularizers [21, 69], superpixels [60] or point cloud denoising [75] to integrate spatial context. In contrast, our method is real-time capable and integrates global 3D context.

Learned DSPs can be divided into multimodal approaches [39, 73] requiring additional image data and unimodal approaches requiring solely waveform data [55, 56, 76]. Multimodal approaches compromise the independence of lidar and camera, eliminating the redundancy needed for robust perception in autonomous driving. Existing learned DSPs

DSP	Conv.	Opti. [60, 74]	Multi. [39, 73]	Uni. [55, 56]	Ours
Data-Driven	✗	✗	✓	✓	✓
Neighborhood	✗	✓	✓	✓	✓
Multiple Echoes	✓	(✓)	✗	✗	✓
Full Res.	✓	✓	✗	✗	✓
Real-Time	✓	✗	(✓)	(✓)	✓
Sensing Scenario					
Low-SNR	✗	✓	✓	✓	✓
Scattering	✗	(✓)	(✓)	(✓)	✓
Blooming	✗	✗	(✓)	(✓)	✓
Long-Range	✓	✓	(✓)	✓	✓

Table 1. **Transient Imaging and Lidar DSPs**. Each criterion is fully ✓, partially (✓), or not ✗ met. We compare conventional (Conv.), optimization-based (Opti.), learned multimodal (Multi.), and unimodal (Uni.) DSPs with the proposed method.

function as waveform denoising methods, supervising with a pseudo ground truth waveform using a Kullback-Leibler (KL) divergence loss and reconstructing distance with an argmax operation. This limits them to a single distance per waveform and lacks supervision in sky regions where no objects are present. In contrast, the proposed DSP treats reconstruction as a classification problem, distinguishing between empty and occupied areas, followed by distance regression. This enables multiple peak detection and proper handling of sky regions. While unimodal DSPs in [56, 76] also consider multiple returns, they assume only one peak is valid, which is reasonable for kilometer-range applications where the beam divergence of a single emitted beam causes returns on multiple receiver pixels, but is not suitable for autonomous driving.

Full Waveform Lidar Sensors find various applications in, e.g., airborne systems for geoscience [24], archaeology [12], and forestry [57, 83], and non-line-of-sight imaging [47, 62, 68]. The benefit of FWL data in adverse weather conditions, where conventionally estimated point clouds suffer from scattering, has been investigated by [58, 66, 82]. Recent works also demonstrate that it is possible to recover neural radiance fields from FWL data [33, 43, 44]. The FWL in our work acquires waveforms by recording timestamps of SPAD trigger events in histograms via Time-Correlated Single Photon Counting (TCSPC) [1, 10, 49] across multiple pulse emission cycles. Following prior work, the SPADs in our sensor operate in free-running mode [28, 72, 85] with asynchronous acquisition [18, 78], enabling multiple photon detections per cycle. Unlike first-photon methods [32] that deactivate after initial detection until the next cycle, the free-running mode allows detection of photons from objects even after initial triggers from high ambient light, preventing pile-up distortions mentioned in other works [23, 27].

3. Sensing Forward Model

In this work, we employ a sensor with a 2D SPAD array capturing full waveform data via TCSPC sensing.

General Forward Model. We model full waveform sensing using the transient photon flux ψ incident on the sensor following [29, 50]. For an idealized pixel (i, j) illuminating a single opaque point, the photon flux is given as

$$\psi_{i,j}(t) = \int_0^\infty g(\tau) H_{i,j}(t-\tau) d\tau + a_{i,j}(t), \tag{1}$$

where $g(t)$ is the temporal shape of the laser pulse, $H_{i,j}$ is the transient scene response and $a_{i,j}(t)$ accounts for ambient photon flux. The lidar illuminates the scene with N sequential laser pulses g and records trigger events from arriving photons within a time bin $k \in [0..T-1]$ to build up the measured waveform histogram κ, where T denotes the number of temporal bins. Building on Eq. (1), the photon detection rate λ in bin k for N pulses can then be modeled as the quantized version of the photon flux by

$$\lambda_{i,j}[k] = \mu \int_{k\Delta}^{(k+1)\Delta} \psi_{i,j}(t) dt, \tag{2}$$

where Δ denotes the temporal bin width and $\mu \in [0, 1]$ is the photon detection probability [63]. The captured waveform κ is modeled as a random variable drawn from the probability distribution P characterizing the sensor response as

$$\kappa_{i,j}[k] \sim P(\lambda_{i,j}), \tag{3}$$

that is an integer number representing the number of trigger events per k. Since the distribution P depends on the photon rates across all temporal bins, the sensor response inherently models temporal correlations between neighboring bins [23, 64]. In the following, we analyze both low and high-flux conditions before introducing a model for fog, see Fig. 2.

Low-Flux Transients. In the low-flux regime, we adopt the multi-peak model from Goudreault *et al.* [17] to simulate diverging laser beams. We discretize the scene into a grid of points at distances $\mathbf{d} \in \mathbb{R}^{aH \times aW}$ by ray-tracing with the view direction for every pixel (i, j), where a denotes an integer upsampling factor. Following Rasshofer *et al.* [65], the scene response of a single point object at $d_{i,j}$ is given by

$$H_{i,j}(t) = \frac{\rho_{i,j}}{4 d_{i_j}^2} \delta \left(t - 2 \frac{d_{i,j}}{\mathrm{c}} \right), \tag{4}$$

where $\rho_{i,j}$ denotes the incident angle dependent reflectivity, c the speed of light, and δ a Dirac pulse. Substituting Eq. (4) into Eq. (1) yields the photon flux

$$\psi_{i,j}(t) = \frac{\rho_{i,j}}{4 d_{i_j}^2} g \left(t - 2 \frac{d_{i,j}}{\mathrm{c}} \right) + a_{i,j}(t). \tag{5}$$

As indicated by Fig. 2f), due to beam divergence, the pulse g can hit multiple objects, producing a multiple-peak waveform. Following [17], this is modeled as a linear combination

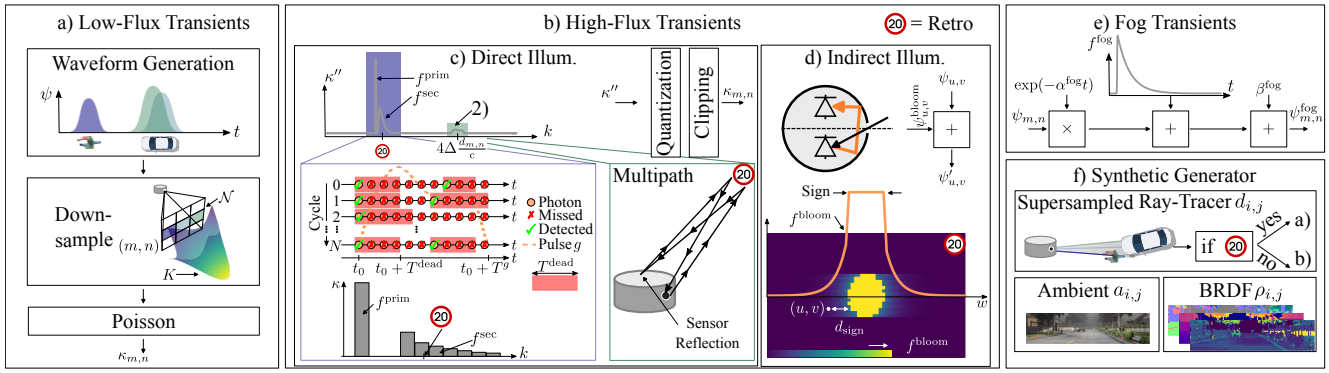

Figure 2. **Transient Waveform Lidar Forward Model.** Full waveform lidar sensing is modeled by distinguishing between low-flux (a) and high-flux returns from retroreflectors (b). In low-flux conditions, a multi-peak waveform $\kappa_{m,n}$ at pixel (m, n) is simulated from a supersampled scene [17]. For pixels illuminating retroreflectors (c), waveform distortions and multipath effects from sensor reflections are considered to model high-flux FWL sensing. Pixels (u, v) that do not illuminate retroreflectors directly (d) might be affected by blooming effects, due to sensor-internal scattering. Realistic fog transients (e) are simulated by adapting [19, 25]. For scene geometry, reflectivity, and ambient light, we extract a supersampled scene, denoted with indices (i, j), from CARLA (f) to enable training data generation.

of neighboring transients from a supersampled scene. Specifically, for a pixel (m, n) in the $H \times W$ SPAD array, the flux $\psi_{m,n}$ is computed by downsampling transients in the spatial neighborhood around (m, n), such that

$$\psi_{m,n}(t) = \sum_{i,j \in \mathcal{N}(m,n)} K_{i,j} \psi_{i,j}(t), \tag{6}$$

where $\mathcal{N}$ defines the $a \times a$ spatial neighborhood and $K_{i,j}$ the spatial intensity profile of the beam.

In low-flux conditions, when the number of expected photons per pulse is smaller than one [32], each bin k of waveform $\kappa_{m,n}$ can be modeled independently as a Poisson random variable [54, 60, 63], simplifying Eq. (3) to

$$\kappa_{m,n}[k] \sim \text{Poisson}\left(\mu \int_{k\Delta}^{(k+1)\Delta} \psi_{m,n}(t) dt\right). \tag{7}$$

High-Flux Transients. In high-flux conditions, bins k are no longer independent due to SPAD dead time [23, 61]. We model high-flux returns from retroreflective materials, such as traffic signs. Other high-flux effects like pile-up from strong ambient light (see [23]) are mitigated by operating our free-running SPADs in asynchronous mode [18, 78]. As shown in Fig. 2c), high-flux returns immediately trigger events at t_0 in early segments of the returning pulse g, inducing a steep primary peak f^{prim}. After a brief dead time T^{dead}, the SPADs are reactivated and detect subsequent photons. Since $T^{\text{dead}} \approx$ 3-5ns is shorter than the pulse length $T^{\text{g}} \approx$ 10ns, photons immediately trigger additional events from time $t_0 + T^{\text{dead}}$ onward, forming a secondary peak f^{sec}. However, detector non-idealities, such as slow voltage ramp-up after quenching [11], prevent SPAD recovery in every cycle, reducing detection probability for photons arriving during $t \in [t_0 + T^{\text{dead}}, t_0 + T^{g}]$. Moreover, the secondary peak decays due to a combination of varying dead times of the SPADs per pixel and fall-off of the emitted pulse in later segments. We model the primary peak f^{prim} as Gaussian and the secondary peak f^{sec} with an exponentially modified Gaussian distribution, such that

$$\kappa'_{m,n}[k] = f^{\text{prim}}(k, d_{m,n}, \theta^{\text{prim}}) + f^{\text{sec}}(k, d_{m,n}, \theta^{\text{sec}}), \tag{8}$$

where θ^{prim} and θ^{sec} parameterize the shape of both pulses. We refer to the Supplementary Material for more details.

With high-flux returns, we observe returns that reflect off the sensor front cover, travel back to the object, and register as a multipath peak at double the distance, as shown in Fig. 2c). To model multipath effects, Eq. (8) is extended as

$$\kappa''_{m,n}[k] = \kappa'_{m,n}[k] + \frac{r}{d^2_{m,n}} g'\left(k - 4\Delta \frac{d_{m,n}}{\text{c}}\right), \tag{9}$$

where r denotes the reflectivity of the retroreflector and g' the quantized pulse by Δ. Integer quantization and clipping to model saturation of $\kappa''_{m,n}$ yields the final $\kappa_{m,n}$.

Highly reflective targets affect not only directly illuminated pixels, but also neighboring ones, as photons scatter within the sensor and illuminate adjacent SPADs (see Fig. 2d). This effect, known as blooming, generates false positive points around high-reflectance areas with exponentially decaying intensity as the distance from the target increases [15]. To model blooming, for each pixel (u, v) not illuminating a retroreflector, the point $\mathbf{p}^{\text{bloom}}_{u,v}$ at the intersection of the ray and the plane containing the high-reflectance region is determined. Then, the Euclidean distance $d^{\text{sign}}_{u,v}$ between $\mathbf{p}^{\text{bloom}}_{u,v}$ and the closest point on the retroreflective target along line u is computed. Due to the line-wise readout of our FWL, blooming effects are independent for each line. Finally, the flux from indirect illumination generating the blooming is computed as

$$\psi^{\text{bloom}}_{u,v}(t) = f^{\text{bloom}}(d^{\text{sign}}_{u,v}) g\left(t - 2\frac{\|\mathbf{p}^{\text{bloom}}_{u,v}\|}{\text{c}}\right), \tag{10}$$

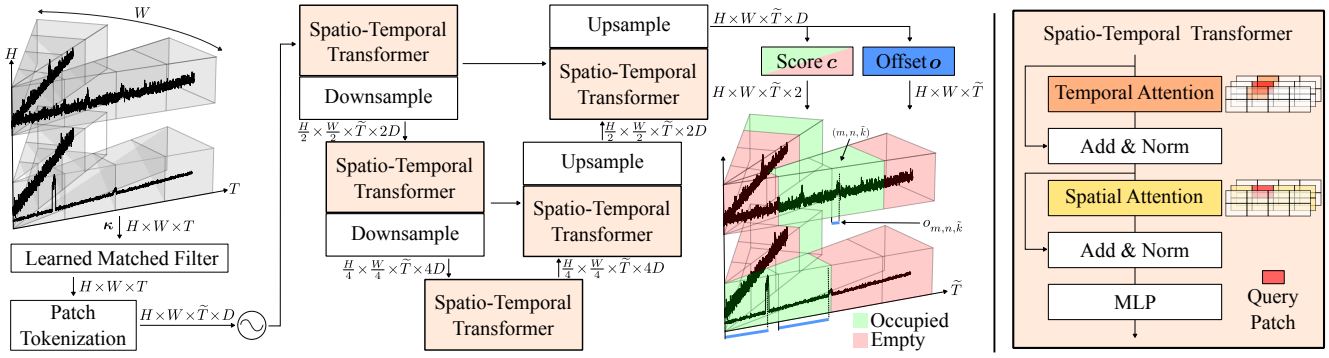

Figure 3. **Neural DSP**. Waveforms $\boldsymbol{\kappa}$ are divided along the temporal dimension into $\widetilde{T}$ patches. Waveform patches are then processed with spatio-temporal transformers in an U-Net architecture. The spatio-temporal transformer performs first temporal attention for each pixel individually and then uses a SWIN approach to compute spatial attention. The neural DSP predicts a classification score, classifying whether a patch is empty or occupied, and then regresses a distance to locate the peak within the patch. By thresholding with the classification score, the neural DSP is able to output a multi-peak point cloud at inference time.

where $f^{\text{bloom}} = B\exp(-\alpha^{\text{bloom}} d^{\text{sign}}_{u,v})$ with B and α^{bloom} describe the decaying blooming effect. The total flux $\psi'_{u,v}$ is computed by adding the blooming flux to the scene flux $\psi_{u,v}$, and Eq. (7) is used to simulate FWL sensing.

Scattering Transients. Existing work [20, 65] finds that waveforms in fog exhibit exponential decay compared to clear-weather waveforms plus an initial broad peak caused by scattering from suspended droplets. Following [22, 25], this initial scattering peak can be effectively modeled using an exponentially modified Gaussian distribution f^{fog}. Thus, the flux $\psi^{\text{fog}}_{m,n}(t)$ in foggy conditions can be written as

$$\psi^{\text{fog}}_{m,n}(t) = \exp(-\alpha^{\text{fog}} t)\psi_{m,n}(t) + f^{\text{fog}}(t, \theta^{\text{fog}}) + \beta^{\text{fog}}, \quad (11)$$

where α^{fog} quantifies the fog thickness, θ^{fog} parameterizes the shape of the initial scattering peak, and β^{fog} is a small offset to model multiple scattering, which improves waveform realism. FWL sensing is then modeled with Eq. (7) to yield the foggy waveform.

4. Neural DSP

To leverage full waveform data in a learning-based approach, we devise a neural DSP that processes all waveforms $\boldsymbol{\kappa} \in \mathbb{R}^{H\times W\times T}$ in unison, where T is the number of temporal bins. We treat peak finding as a classification problem by dividing the temporal axis into $\widetilde{T} = 33$ patches and classifying whether each patch is occupied or empty. For occupied patches, we regress a distance offset $\boldsymbol{o} \in \mathbb{R}^{H\times W\times\widetilde{T}}$ to determine the final distance as shown in Fig. 3. At inference time, our neural DSP outputs a multi-peak point cloud by thresholding with the classification score $\boldsymbol{c} \in \mathbb{R}^{H\times W\times\widetilde{T}\times 2}$. We describe the architecture in Sec. 4.1, an extension for superresolution in Sec. 4.2, and training details in Sec. 4.3.

4.1. Spatio-Temporal Waveform Transformers

We first apply matched filtering [79] to increase SNR, processing $\boldsymbol{\kappa}$ with a 1D temporal convolution. The waveform is then divided into $\widetilde{T}$ patches and tokenized to feature vectors of size $D = 32$. The feature embeddings are augmented with sinusoidal positional encodings along the temporal axis and processed in an U-Net architecture [7] using spatio-temporal transformers as elementary building blocks.

To handle the large number of patches, our spatio-temporal transformer decomposes attention into temporal and spatial attention [3], as illustrated in Fig. 3. Temporal attention is computed between every patch of a specific pixel. Spatial attention uses a shifted window (SWIN) approach [41], operating on non-overlapping local windows. Window positions are shifted between consecutive layers to enable cross-window interactions. Given the non-square spatial resolution of $\boldsymbol{\kappa}$, we adopt non-square 2x4 attention windows following [84]. After adding features from residual connections and layer normalization, the resulting patch features are processed with a multi-layer perceptron (MLP).

Downsampling employs patch merging layers [41], concatenating features from 2x2 neighborhoods and projecting them to $2D$-dimensional embeddings. The encoder applies two downsampling steps, each halving the spatial resolution to capture global context, followed by a spatio-temporal transformer at the bottleneck. For upsampling, the decoder reverses this process using patch unmerging layers [84] to restore the original spatial resolution.

We employ two prediction heads that process the patch embeddings through linear projections: classification scores $\boldsymbol{c}$ with softmax activation, and offsets $\boldsymbol{o}$ with sigmoid activation, constraining offsets within patches. The final distance for an occupied patch defined by index $(m, n, \tilde{k})$, where $\tilde{k} \in [0..\widetilde{T} - 1]$, can be determined as

$$d_{m,n,\tilde{k}} = \tilde{k}\frac{T}{\widetilde{T}} + o_{m,n,\tilde{k}}\frac{T}{\widetilde{T}}. \quad (12)$$

4.2. Learning Super-Resolution from Transients

Current lidar sensors suffer from lower resolution compared to, e.g., cameras. However, the full waveform contains help-

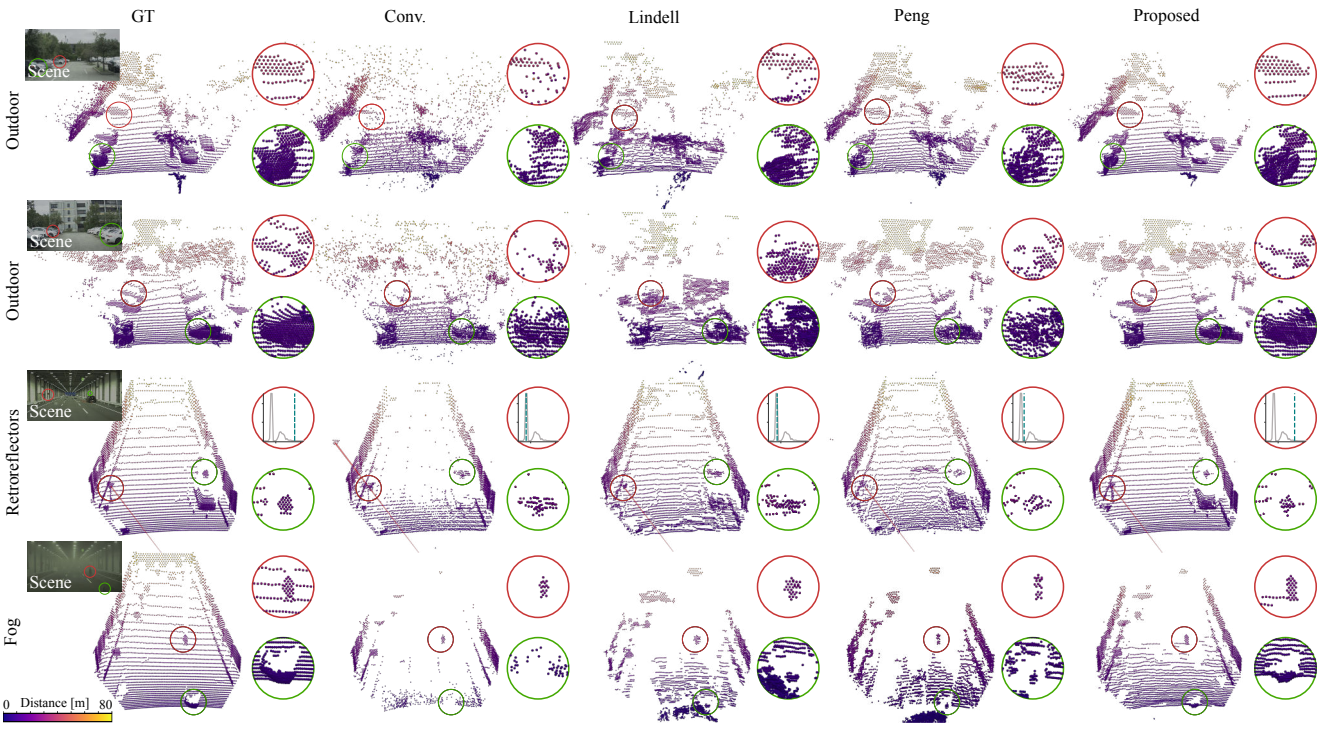

Figure 4. **Qualitative Evaluation on Clear and Foggy Real-World Data.** Conventional (Conv.) DSPs struggle in low SNR regions, see, e.g., dark car or windows in the first two rows. The transient imaging methods Lindell *et al*. [39] and Peng *et al*. [56] improve reconstruction in low SNR areas, but miss high-frequency details in contrast to the neural DSP (zoom-ins). Furthermore, our neural DSP learns to correct distance distortions and blooming from retroreflectors (zoom-ins in third row) and is the only method to deliver satisfactory results in fog.

ful information to artificially increase the spatial resolution without the need for more pixels. For example, multiple peaks in the waveform suggest that a higher-resolution lidar with smaller beam divergence would resolve distinct points. Conversely, a single strong peak indicates that a higher-resolution system would detect multiple points at similar distances. To exploit these super-resolution (SR) waveform cues, we extend our neural DSP by appending an additional spatio-temporal transformer and an upsampling block to the U-Net decoder. This allows to effectively disentangle Eq. (6) to render a SR feature map of size $aH \times aW \times \widetilde{T} \times D$ upsampled by $a = 3$. We then apply classification and offset heads to produce the final SR point cloud.

4.3. Training

We supervise patch classification with a focal loss [38] to counter the imbalance of occupied and empty patches as

$$\mathcal{L}_{\text{class}} = \sum_{i=0}^{1} \alpha (1 - [\boldsymbol{c}]_i)^{\gamma} \, [\boldsymbol{c}_{\text{gt}}]_i \log \left([\boldsymbol{c}]_i \right), \qquad (13)$$

where $\alpha = 0.25$ and $\gamma = 2$ are default hyperparameters, $\boldsymbol{c}_{\text{gt}}$ are the ground truth binary labels, $[\boldsymbol{c}]_0$ and $[\boldsymbol{c}]_1$ denote the probability that a patch is occupied or empty, respectively. The offset $\boldsymbol{o}$ is supervised with a L1 loss

$$\mathcal{L}_{\text{offset}} = \left| \mathbf{m} \odot (\boldsymbol{o}_{\text{gt}} - \boldsymbol{o}) \right|_1, \qquad (14)$$

where $\boldsymbol{o}_{\text{gt}}$ are the ground truth offsets computed by rewriting Eq. (12) and $\mathbf{m}$ is a boolean mask denoting all patches including a ground truth scene point.

We train on synthetic data and fine-tune on real data. We use Eq. (9) and Eq. (10) to augment additional traffic signs, as well as Eq. (11) for fog into the real data. Note that the neural DSP is only trained on synthetic fog and has never seen real-world fog. More details on training and augmentation are provided in the Supplementary Material.

5. Evaluation

We evaluate our method against conventional DSPs and state-of-the-art learned transient imaging approaches on synthetic data with perfect ground truth and real-world measurements in clear and foggy conditions. For synthetic data, we follow [17, 67] and integrate our forward model into CARLA, see Fig. 2f). For real data, we configure a production-grade Microvision MOVIA™ FWL to have a resolution of $H = 40$ by $W = 128$ over a 15°(vertical) by 60°(horizontal) FoV. Waveforms are recorded with a temporal resolution of 266 ps over $T = 2112$ bins, allowing for a range of 85 m. As shown in Fig. 1a), we capture data in outdoor scenarios - using accumulated Velodyne VLS-128 point clouds as ground truth - for fine-tuning and testing. Additional test data is captured in both clear and foggy conditions in a weather chamber, with dense Leica scans as ground truth.

	SNR	0 – ∞		0 – 2		2 – 4	
	Method	CD [m] ↓	Recall [%] ↑	CD [m] ↓	Recall [%] ↑	CD [m] ↓	Recall [%] ↑
Synthetic	Conv.	0.648	81.15	2.355	37.81	0.682	88.99
	Lindell [39]	1.146	71.27	2.301	44.70	1.036	76.33
	Peng [56]	0.749	78.18	1.249	58.13	0.898	82.08
	Proposed	**0.397**	**87.79**	**1.140**	**64.07**	**0.421**	**91.81**
Clear	Conv.	0.626	81.52	1.294	52.65	0.402	94.74
	Lindell [39]	1.098	80.69	1.793	65.77	0.702	81.57
	Peng [56]	0.505	88.87	0.659	78.79	0.501	86.74
	Proposed	**0.307**	**92.57**	**0.428**	**85.18**	**0.266**	**95.44**
Fog	Conv.	2.199	36.62	2.912	23.13	0.539	94.15
	Lindell [39]	4.711	40.93	5.293	31.99	2.165	73.68
	Peng [56]	2.326	68.27	2.700	62.30	0.528	88.43
	Proposed	**1.110**	**73.78**	**1.265**	**68.43**	**0.499**	**94.35**

Table 2. **Quantitative Results.** Existing transient imaging methods [39, 56] improve reconstruction in low SNR (0–2) scenarios. At high SNR (2–4), their denoising strategy offers no advantage over conventional peak finding (Conv.). The proposed neural DSP outperforms baselines across all SNR ranges and excels in fog.

	Conv.	Lindell[39]	Peng[56]	Proposed
Clear	43.38	30.27	51.44	**65.27**
Fog	28.80	28.41	39.55	**56.71**

Table 3. **Maximum Range (↑) [m]**: On clear and foggy real-world data, the neural DSP outperforms the best baseline by 14m and 17m maximum range in clear and foggy conditions, respectively.

As the ground truth may contain multiple valid distances per pixel, we evaluate on the point cloud level using Chamfer Distance (CD) and Recall as metrics. To assess the benefit in challenging conditions, we evaluate two SNR ranges: difficult low-SNR conditions (0-2) and easier high-SNR conditions (2-4). Following commercial sensors, we further introduce a maximum range metric on low-reflective targets. We use low-SNR points (0–2) as these targets, bin them by distance, calculate distance-binned Recall, and define maximum range as the farthest distance achieving 50% Recall. Details on the simulator, datasets, and metrics are provided in the Supplementary Material

5.1. Baseline DSPs

Since commercial on-device DSPs are proprietary (though we compare against commercial lidar sensors below), we implement a conventional ray-based pipeline following [17]. This pipeline employs matched filtering, ambient light removal, and peak detection to identify up to four peaks per waveform above a specified threshold. In foggy conditions, we only use the last peak [4] and adjust the threshold. Furthermore, we compare against two learned transient imaging methods: Lindell *et al*. [39] and Peng *et al*. [56]. For Lindell *et al*., we use their method without RGB camera input. As discussed earlier, both methods struggle in sky regions. Since 26% of pixels in our dataset correspond to sky regions, this would hinder a fair comparison. Hence, we add a lightweight convolutional network, operating along the temporal dimension of the final feature map, to predict a sky mask and supervise their denoising only on occupied pixels.

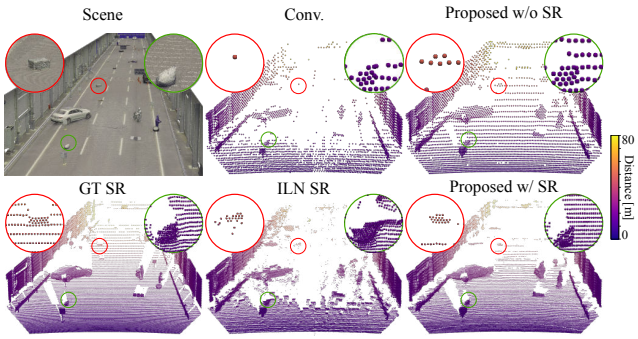

Figure 5. **Qualitative Super-Resolution (SR) Results.** Extending our neural DSP to render SR point clouds is beneficial for distinguishing small hazardous items from the road, see zoom-ins for first row compared to SR results in the second row. The best-performing baseline ILN [35] suffers from artifacts and flying pixels.

5.2. Assessment

Reported in Tab. 2, our neural DSP improves point cloud reconstruction on synthetic data, reducing CD by 63.22% over conventional DSPs. Analyzing waveforms jointly is particularly effective in low-SNR conditions, where all transient imaging methods, including our neural DSP, achieve higher Recall than the conventional DSP. However, Lindell *et al*. and Peng *et al*. produce less accurate reconstructions, especially at high SNR, where their denoising approach offers no advantage over conventional peakfinding.

The results on synthetic data are consistent with those from the real clear-weather test set. As shown in Fig. 4 (first row), all transient imaging methods reconstruct the low-reflectivity black car missed by the conventional DSP. However, only our neural DSP captures the car's shape with high-frequency details. This aligns with the second row, where our neural DSP provides detailed reconstructions at low SNR, accurately reconstructing the car's side windows. In contrast, Lindell *et al*. and Peng *et al*. reconstruct low-frequency scene trends, but their distance predictions exhibit more noise. Furthermore, as shown in the third row, the proposed neural DSP corrects distance distortions and multipath effects caused by retroreflectors, challenging for baseline DSPs. Our classification approach differentiates between blooming and actual target peaks, as seen e.g. for the rightmost traffic sign. In contrast, the denoising strategy of Lindell *et al*. and Peng *et al*. struggles to fully suppress false peaks, resulting in excess points around the traffic sign. Overall, the proposed neural DSP improves CD by 32 cm over conventional peakfinding and by 20 cm over the best-performing baseline by Peng *et al*. As CD merely penalizes missed or false points over distance errors, this highlights our method's strength: suppressing false peaks and recovering hard-to-detect ones, rather than refining distances.

In foggy conditions, existing transient imaging methods fail to suppress false peaks, see Fig. 4 (fourth row). While their higher Recall in Tab. 2 illustrates the benefit of joint processing, both methods suffer from severe backscatter.

Real Data	Pre-Training	High-Flux	CD [m] ↓	Recall [%] ↑
✗	✓	✓	0.535	85.64
✓	✗	✗	0.443	90.99
✓	✓	✗	0.384	91.98
✓	✓	✓	**0.307**	**92.57**

Table 4. **Imaging Model Ablations.** Point cloud quality decreases on the clear test set when simulation components are withheld during training, confirming the accuracy of the forward model.

	Synthetic		Real	
	CD [m] ↓	Recall [%] ↑	CD [m] ↓	Recall [%] ↑
Tulip [84]	0.552	89.76	0.678	86.93
ILN [35]	0.387	88.12	0.413	87.52
Proposed	**0.264**	**92.81**	**0.273**	**94.47**

Table 5. **Quantitative Super-Resolution Evaluation**. Our neural DSP can leverage waveform information to render super-resolution (SR) point clouds, allowing for improved reconstruction performance compared to range image-based SR methods.

The proposed neural DSP reconstructs both the pedestrian at 30m and the low-reflective bumper, reducing CD by 1.08m compared to the best-performing baseline.

These improvements translate to an increase in maximum range of 13.8m and 17.2m compared to the best-performing baseline in clear and foggy conditions, as shown by Tab. 3.

Our method processes waveforms in a single pass, achieving real-time rates of 68 Hz on unoptimized PyTorch code (sensor at 15 Hz). Peng *et al*. and Lindell *et al*. require multiple passes, reducing rates to 12 Hz and 5 Hz, respectively.

5.3. Validation of Forward Model

To validate the accuracy and effectiveness of our forward model, we ablate different components of the simulation during training, and measure test performance. As shown in Tab. 4, using only simulated data still achieves 0.535m CD on the clear test set (comparable to the best baseline), while removing the simulation altogether increases CD by 44.3%, confirming its necessity. Omitting the high-flux simulation increases CD by 25.1% as predicted point clouds suffer from blooming. Removing the fog simulation reduces accuracy to 2.311m CD (+108.2%) on the fog test as the neural DSP is unable to suppress false scattering peaks.

5.4. Lidar Super-Resolution

To assess our neural DSP for lidar SR, we compare against state-of-the-art methods ILN [35] and Tulip [84], training all methods to reconstruct point clouds with upsampling factor $a = 3$. On both synthetic and real-world data, we find that the use of waveform data significantly improves reconstruction compared to baseline methods, see Tab. 5. As the baseline methods ingest point clouds with only the strongest measured peak per pixel, they cannot leverage multiple peak cues. These cues are beneficial for detecting small hazardous items on the road, as illustrated by the qualitative findings in Fig. 5. Regular DSPs (first row) produce few points on small objects, making them hard to distinguish from the road, while the proposed method (second row) allows clear distinction. The best-performing baseline ILN suffers from a significant number of flying pixels, whereas the neural SR DSP yields a high-frequency SR scene reconstruction.

Distance accuracy [m] ↓ compared to dense GT scan in weather chamber				
Sensor	Proposed	Aeva Aeries II	Luminar Iris	Velodyne VLS-128
Clear	**0.056**	0.071	0.067	0.035
Fog	**0.074**	0.126	0.166	0.103

Table 6. **Comparison with Commercial Lidars and GT Velodyne.** Our proposed neural DSP compares favorably to commercially available lidars, especially in fog.

5.5. Comparison with Automotive Lidar Sensors

We compare the point clouds produced by the neural DSP against two commercially mass-produced lidar sensors: Aeva Aeries II and Luminar Iris. The Velodyne VLS-128 targets ground-truthing applications and operates at a different cost level. Tab. 6 shows distance accuracy comparisons to the dense Leica ground truth scan in the weather chamber. In clear conditions, the neural DSP yields higher accuracy than the commercial sensors, second only to the Velodyne used to train the network. In fog, the neural DSP is capable of suppressing scatter points and delivers the best distance accuracy, outperforming all sensors, including the Velodyne. Qualitative results are reported in the Supplement.

6. Conclusion

We propose a neural DSP that learns high-fidelity multi-echo point clouds from lidar waveforms by extracting waveform features from adjacent rays. To train it, we devise a realistic sensor forward model simulating low-, high-flux, and foggy transients. In clear weather, our neural DSP improves point cloud quality by 32 cm and 20 cm CD over conventional DSPs and transient imaging methods. Further, our integrated super-resolution improves CD by 50% over the best-performing baseline while allowing for a ninefold increase in returned points. In fog, the Neural DSP reduces CD by over a meter, translating to 17m maximum range improvements, compared to the best-performing baseline. While our approach builds on current production-grade lidar sensors, limited on-device compute prevents deployment of deep neural networks, requiring offloading to a host. Bandwidth constraints further limit the frame rate, though this may be alleviated in future hardware. As our methods outperform conventional DSPs, particularly under challenging conditions, we hope this work encourages LiDAR manufacturers to make future sensing pipelines more programmable, enabling end-to-end scene understanding directly on-device.

Acknowledgements This project has been supported by the German Ministry BMWE within the project "NXT GEN AI METHODS". Felix Heide was supported by an NSF CAREER Award (2047359), a Packard Foundation Fellowship, a Sloan Research Fellowship, a Sony Young Faculty Award, a Project X Innovation Award, a Amazon Science Research Award, and a Bosch Research Award.

References

[1] Wolfgang Becker. *Advanced Time-Correlated Single Photon Counting Techniques*. Springer, Berlin, Heidelberg, 2005. 3

[2] Behnam Behroozpour, Phillip A. M. Sandborn, Ming C. Wu, and Bernhard E. Boser. Lidar System Architectures and Circuits. *IEEE Communications Magazine*, 55(10):135–142, 2017. Conference Name: IEEE Communications Magazine. 1

[3] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *ICML*, page 4, 2021. 5

[4] Mario Bijelic, Tobias Gruber, and Werner Ritter. A benchmark for LiDAR sensors in fog: Is detection breaking down? In *IEEE Intelligent Vehicles Symposium (IV)*, 2018. 1, 7

[5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[6] Gerald Buller and Andrew Wallace. Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition. *IEEE Journal of selected topics in quantum electronics*, 13(4):1006–1015, 2007. 2

[7] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 5

[8] Alexander Carballo, Jacob Lambert, Abraham Monrroy Cano, David Robert Wong, Patiphon Narksri, Yuki Kitsukawa, Eijiro Takeuchi, Shinpei Kato, and Kazuya Takeda. LIBRE: The multiple 3D LiDAR dataset. In *IEEE Intelligent Vehicles Symposium (IV)*, 2020. 1

[9] M. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays. Argoverse: 3D tracking and forecasting with rich maps. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[10] E. Charbon. Introduction to time-of-flight imaging. In *SENSORS, 2014 IEEE*, pages 610–613, 2014. 3

[11] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Appl. Opt.*, 35(12):1956–1976, 1996. 4

[12] Michael Doneus and Christian Briese. Full-waveform airborne laser scanning as a tool for archaeological reconnaissance. *BAR International Series*, 1568:99, 2006. 3

[13] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *1st Annual Conference on Robot Learning*, 2017. 2

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021. 2

[15] Karl Christoph Goedel, Holger Maris Gilbergs, and Johannes Richter. Computer unit for a lidar device, and lidar device, 2023. US2023092128A1. 2, 4

[16] Christopher Goodin, Daniel Carruth, Matthew Doude, and Christopher Hudson. Predicting the influence of rain on LiDAR in ADAS. *Electronics*, 8, 2019. 1

[17] Felix Goudreault, Dominik Scheuble, Mario Bijelic, Nicolas Robidoux, and Felix Heide. Lidar-in-the-loop hyperparameter optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13404–13414, 2023. 3, 4, 6, 7

[18] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3d imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7909–7918, 2019. 3, 4

[19] Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 1, 4

[20] Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. LiDAR Snowfall Simulation for Robust 3D Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 5

[21] Abderrahim Halimi, Yoann Altmann, Aongus McCarthy, Ximing Ren, Rachael Tobin, Gerald S Buller, and Steve McLaughlin. Restoration of intensity and depth images constructed using sparse single-photon data. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 86–90. IEEE, 2016. 2

[22] Felix Heide, Lei Xiao, Andreas Kolb, Matthias B. Hullin, and Wolfgang Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Opt. Express*, 22(21):26338–26350, 2014. 2, 5

[23] Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. Sub-picosecond photon-efficient 3d imaging using single-photon sensors. *Scientific reports*, 8(1):1–8, 2018. 1, 3, 4

[24] Michelle A Hofton, Jean-Bernard Minster, and J Bryan Blair. Decomposition of laser altimeter waveforms. *IEEE Transactions on geoscience and remote sensing*, 38(4):1989–1996, 2000. 3

[25] Hanno Holzhüter. Method and device for optical distance measurement, 2024. US Patent 11,892,569. 4, 5

[26] Hanno Holzhüter, Jörn Bödewadt, Shima Bayesteh, Andreas Aschinger, and Holger Blume. Technical concepts of automotive lidar sensors: a review. *Optical Engineering*, 62(3): 031213–031213, 2023. 1

[27] Hilmar Ingensand and Timo Kahlmann. Single-photon detection for high precision ranging – a trade-off study. In *1st Range Imaging Research Day Proceedings*, pages 33–42. Hilmar Ingensand and Timo Kahlmann, 2005. 3

[28] Atul Ingle, Andreas Velten, and Mohit Gupta. High flux passive imaging with single-photon sensors. In *Proceedings*

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6760–6769, 2019. 3

[29] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65–79, 2017. 3

[30] Maria Jokela, Matti Kutila, and Pasi Pyykönen. Testing and validation of automotive point-cloud sensors in adverse weather conditions. *Applied Sciences*, 9, 2019. 1

[31] B Jutzi and U Stilla. Laser pulse analysis for reconstruction and classification of urban objects. *International archives of photogrammetry remote sensing and spatial information sciences*, 34(3/W8):151–156, 2003. 1, 2

[32] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco NC Wong, Jeffrey H Shapiro, and Vivek K Goyal. First-photon imaging. *Science*, 343(6166): 58–61, 2014. 2, 3, 4

[33] Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, and Rakesh Ranjan. Platonerf: 3d reconstruction in plato's cave via single-view two-bounce lidar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14565–14574, 2024. 3

[34] Matti Kutila, Pasi Pyykönen, Maria Jokela, Tobias Gruber, Mario Bijelic, and Werner Ritter. Benchmarking automotive LiDAR performance in arctic conditions. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2020. 1

[35] Youngsun Kwon, Minhyuk Sung, and Sung-Eui Yoon. Implicit lidar network: Lidar super-resolution via interpolation weight prediction. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8424–8430. IEEE, 2022. 7, 8

[36] Jacob Lambert, Alexander Carballo, Abraham Monrroy Cano, Patiphon Narksri, David Wong, Eijiro Takeuchi, and Kazuya Takeda. Performance analysis of 10 models of 3d lidars for automated driving. *IEEE Access*, 8:131699–131722, 2020. 1

[37] Zheng-Ping Li, Jun-Tian Ye, Xin Huang, Peng-Yu Jiang, Yuan Cao, Yu Hong, Chao Yu, Jun Zhang, Qiang Zhang, Cheng-Zhi Peng, Feihu Xu, and Jian-Wei Pan. Single-photon imaging over 200km. *Optica*, 8(3):344–349, 2021. 2

[38] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2020. 6

[39] David B. Lindell, Matthew O'Toole, and Gordon Wetzstein. Single-photon 3d imaging with deep sensor fusion. *ACM Trans. Graph.*, 37(4), 2018. 1, 2, 3, 6, 7

[40] David B. Lindell, Matthew O'Toole, and Gordon Wetzstein. Towards transient imaging at interactive rates with single-photon detectors. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2018. 2

[41] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5

[42] Ievgeniia Maksymova, Christian Steger, and Norbert Druml. Review of LiDAR Sensor Data Acquisition and Compression for Automotive Applications. *Proceedings*, 2(13):852, 2018. Number: 13 Publisher: Multidisciplinary Digital Publishing Institute. 2

[43] Anagh Malik, Noah Juravsky, Ryan Po, Gordon Wetzstein, Kiriakos N Kutulakos, and David B Lindell. Flying with photons: Rendering novel views of propagating light. In *European Conference on Computer Vision*, pages 333–351. Springer, 2024. 3

[44] Anagh Malik, Parsa Mirdehghan, Sotiris Nousias, Kyros Kutulakos, and David Lindell. Transient neural radiance fields for lidar view synthesis and 3d reconstruction. *Advances in Neural Information Processing Systems*, 36, 2024. 3

[45] Yunze Man, Xinshuo Weng, Prasanna Kumar Sivakumar, Matthew O'Toole, and Kris M Kitani. Multi-echo lidar for 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3763–3772, 2021. 2

[46] Aongus McCarthy, Nils J. Krichel, Nathan R. Gemmell, Ximing Ren, Michael G. Tanner, Sander N. Dorenbos, Val Zwiller, Robert H. Hadfield, and Gerald S. Buller. Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express*, 21(7):8904–8915, 2013. 2

[47] Ji Hyun Nam, Eric Brandt, Sebastian Bauer, Xiaochun Liu, Marco Renna, Alberto Tosi, Eftychios Sifakis, and Andreas Velten. Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second. *Nature communications*, 12(1):6526, 2021. 3

[48] C. Niclass, A. Rochas, P. Besse, and E. Charbon. Design and characterization of a cmos 3-d image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits*, 40(9):1847–1854, 2005. 1

[49] Desmond V. O'Connor and David Phillips. *Time-Correlated Single Photon Counting*. Academic Press, 1984. 3

[50] Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Transactions on Graphics (ToG)*, 33(4):1–11, 2014. 3

[51] Matthew O'Toole, Felix Heide, David B. Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing Transient Images from Single-Photon Sensors. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2289–2297, Honolulu, HI, 2017. IEEE. 1, 2

[52] Angus Pacala and Mark Frichtl. Accurate photo detector measurements for lidar. Ouster, Inc., 2021. Published patent application US11209544B2. 1

[53] Agata M Pawlikowska, Abderrahim Halimi, Robert A Lamb, and Gerald S Buller. Single-photon three-dimensional imaging at up to 10 kilometers range. *Optics express*, 25(10): 11919–11931, 2017. 2

[54] Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal Processing Based Pile-up Compensation for Gated Single-Photon Avalanche Diodes, 2018. arXiv:1806.07437 [physics]. 4

[55] 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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 225–241. Springer, 2020. 1, 2, 3

[56] 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*, 45(4):4180–4197, 2023. 2, 3, 6, 7

[57] Åsa Persson, U Söderman, J Töpel, and Simon Ahlberg. Visualization and analysis of full-waveform airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3/W19):103–108, 2005. 3

[58] Martin Pfennigbauer, Clifford Wolf, Josef Weinkopf, and Andreas Ullrich. Online waveform processing for demanding target situations. In *Laser Radar Technology and Applications XIX; and Atmospheric Propagation XI*, page 90800. SPIE, 2014. 3

[59] Ryan Po, Adithya Pediredla, and Ioannis Gkioulekas. Adaptive gating for single-photon 3d imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16354–16363, 2022. 2

[60] Joshua Rapp and Vivek K Goyal. A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Transactions on Computational Imaging*, 3 (3):445–459, 2017. 2, 3, 4

[61] Joshua Rapp, Yanting Ma, Robin M. A. Dawson, and Vivek K Goyal. Dead time compensation for high-flux ranging. *IEEE Transactions on Signal Processing*, 67(13):3471–3486, 2019. 4

[62] Joshua Rapp, Charles Saunders, Julián Tachella, John Murray-Bruce, Yoann Altmann, Jean-Yves Tourneret, Stephen McLaughlin, Robin MA Dawson, Franco NC Wong, and Vivek K Goyal. Seeing around corners with edge-resolved transient imaging. *Nature communications*, 11(1):5929, 2020. 3

[63] Joshua Rapp, Julian Tachella, Yoann Altmann, Stephen McLaughlin, and Vivek K Goyal. Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances. *IEEE Signal Processing Magazine*, 37(4):62–71, 2020. 1, 2, 3, 4

[64] Joshua Rapp, Yanting Ma, Robin MA Dawson, and Vivek K Goyal. High-flux single-photon lidar. *Optica*, 8(1):30–39, 2021. 3

[65] Ralph H Rasshofer, Martin Spies, and Hans Spies. Influences of weather phenomena on automotive laser radar systems. *Advances in radio science*, 9:49–60, 2011. 3, 5

[66] Guy Satat, Matthew Tancik, and Ramesh Raskar. Towards photography through realistic fog. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, 2018. 3

[67] Dominik Scheuble, Chenyang Lei, Seung-Hwan Baek, Mario Bijelic, and Felix Heide. Polarization wavefront lidar: Learning large scene reconstruction from polarized wavefronts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21241–21250, 2024. 6

[68] Siyuan Shen, Suan Xia, Xingyue Peng, Ziyu Wang, Yingsheng Zhu, Shiying Li, and Jingyi Yu. Holi-1-to-3: Transient-enhanced holistic image-to-3d generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12, 2024. 3

[69] Dongeek Shin, Feihu Xu, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K Goyal. Computational multi-depth single-photon imaging. *Opt. Express*, 24(3):1873–1888, 2016. 1, 2

[70] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[71] Qilin Sun, Xiong Dun, Yifan Peng, and Wolfgang Heidrich. Depth and transient imaging with compressive spad array cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 273–282, 2018. 2

[72] Xin Sun, Hu Yan, Hongcun He, Xiangshun Kong, Chen Mao, and Feng Yan. Comparative analysis of free-running and gating imaging modes of spad sensors. In *Photonics*, page 721. MDPI, 2024. 3

[73] Zhanghao Sun, David B Lindell, Olav Solgaard, and Gordon Wetzstein. SPADnet: deep RGB-SPAD sensor fusion assisted by monocular depth estimation. *Optics Express*, 28 (10):14948–14962, 2020. 1, 2, 3

[74] Julián Tachella, Yoann Altmann, Steve McLaughlin, and J-Y Tourneret. 3d reconstruction using single-photon lidar data exploiting the widths of the returns. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7815–7819. IEEE, 2019. 1, 2, 3

[75] Julián Tachella, Yoann Altmann, Nicolas Mellado, Aongus McCarthy, Rachael Tobin, Gerald S Buller, Jean-Yves Tourneret, and Stephen McLaughlin. Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. *Nature communications*, 10(1):4984, 2019. 1, 2

[76] Hao Tan, Jiayong Peng, Zhiwei Xiong, Dong Liu, Xin Huang, Zheng-Ping Li, Yu Hong, and Feihu Xu. Deep learning based single-photon 3d imaging with multiple returns. In *2020 International Conference on 3D Vision (3DV)*, pages 1196–1205. IEEE, 2020. 1, 2, 3

[77] Rachael Tobin, Abderrahim Halimi, Aongus McCarthy, Martin Laurenzis, Frank Christnacher, and Gerald S Buller. Three-dimensional single-photon imaging through obscurants. *Optics express*, 27(4):4590–4611, 2019. 2

[78] Alessandro Tontini. *Advanced techniques for SPAD-based CMOS d-ToF systems*. PhD thesis, Università degli studi di Trento, 2024. 3, 4

[79] George Turin. An introduction to matched filters. *IRE transactions on Information theory*, 6(3):311–329, 1960. 2, 5

[80] Satarupa Uttarkabat, Sarath Appukuttan, Kwanit Gupta, Satyajit Nayak, and Patitapaban Palo. Bloomnet: Perception of blooming effect in adas using synthetic lidar point cloud data. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 1886–1892, 2024. 1

[81] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 2

[82] Andrew Wallace, Abderrahim Halimi, and Gerald Buller. Full waveform LiDAR for adverse weather conditions. *IEEE Transactions on Vehicular Technology*, 69, 2020. 1, 2, 3

[83] Michael A Wulder, Joanne C White, Ross F Nelson, Erik Næsset, Hans Ole Ørka, Nicholas C Coops, Thomas Hilker, Christopher W Bater, and Terje Gobakken. Lidar sampling for large-area forest characterization: A review. *Remote sensing of environment*, 121:196–209, 2012. 3

[84] Bin Yang, Patrick Pfreundschuh, Roland Siegwart, Marco Hutter, Peyman Moghadam, and Vaishakh Patil. Tulip: Transformer for upsampling of lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15354–15364, 2024. 5, 8

[85] Chao Yu, Qi Xu, and Jun Zhang. Recent advances in ingaas/inp single-photon detectors. *Measurement Science and Technology*, 35(12):122003, 2024. 3