Transient LASSO: Transient Large-Scale Scene Reconstruction

DOMINIK SCHEUBLE*, Mercedes-Benz AG, Germany and Technical University of Darmstadt, Germany
ANDREA RAMAZZINA*, Mercedes-Benz AG, Germany and Technical University of Munich, Germany
HANNO HOLZHÜTER, Microvision Inc, USA and Leibniz University Hannover, Germany
STEFANO GASPERINI, Technical University of Munich, Germany and Munich Center for Machine Learning, Germany
STEVEN PETERS, Technical University of Darmstadt, Germany
FEDERICO TOMBARI, Technical University of Munich, Germany and Google, Switzerland
MARIO BIJELIC, Princeton University, USA and Torc, Germany
FELIX HEIDE, Princeton University, USA and Torc, USA

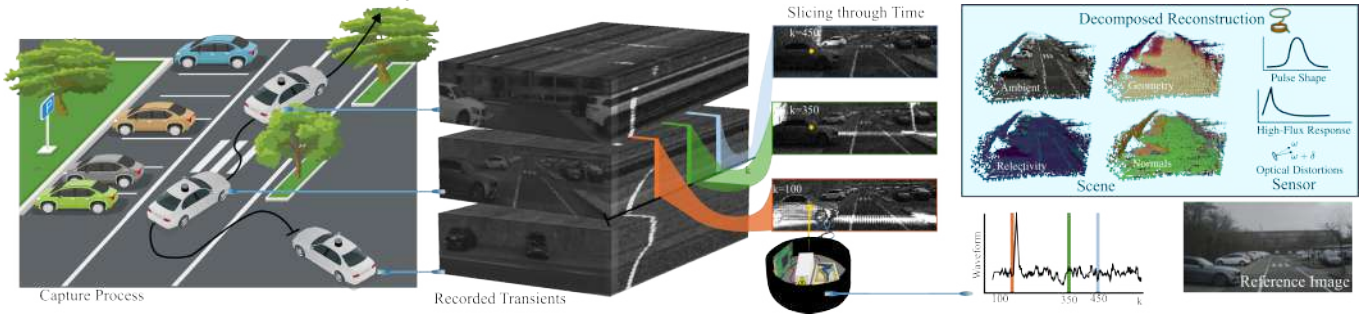

Fig. 1. **Transient LASSO.** We propose a novel neural scene reconstruction method operating on raw transient SPAD measurements from a sequence of in-the-wild large outdoor captures (left). Transient LASSO decomposes these measurements (middle) into geometry, normals, reflectivity, and the sensor parameters (right top). Our method produces transient decompositions for scenes up to 200 m, an order of magnitude larger than existing methods. Furthermore, the decomposition allows for diverse downstream applications, including accurate depth reconstruction, novel view synthesis, super-resolution, and sensor-parameter optimization.

Reconstructing the geometry and appearance of a given scene is a fundamental task in 3D computer graphics and computer vision. Recently, radiance fields have emerged as a representation of light transport in the scene, allowing, as a byproduct, also to extract 3D geometry solely from multi-view imagery. Initially designed for RGB captures, existing approaches have been extended to other sensor modalities. Among these, transient imaging — measuring the time-of-flight of light at picosecond resolution — has emerged as a promising alternative, offering rich spatio-temporal information to improve reconstruction quality from limited viewpoints and obstructed views. However, its applicability to outdoor scenarios has been highly problematic due to interference from ambient light and the different sensor behavior under high-photon-flux conditions typical of outdoor settings. Addressing this gap, we introduce Transient LASSO, a neural scene reconstruction method operating on raw transient measures of outdoor in-the-wild captures to accurately reconstruct the underlying scene geometry and properties. We demonstrate the effectiveness of our method across a variety of outdoor environments, including complex urban scenes with dense traffic and infrastructure. Finally, we also show the potential use cases of our method for downstream applications such as sensor parameter optimization.

CCS Concepts: • **Computing methodologies → Reconstruction**.

Additional Key Words and Phrases: Reconstruction, Transient Imaging, Neural Radiance Fields, LiDAR, 3D Computer Vision, Computational Imaging

ACM Reference Format:
Dominik Scheuble*, Andrea Ramazzina*, Hanno Holzhüter, Stefano Gasperini, Steven Peters, Federico Tombari, Mario Bijelic, and Felix Heide. 2025. Transient LASSO: Transient Large-Scale Scene Reconstruction. In *SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3757377.3763911

1 INTRODUCTION

Accurate reconstruction of large-scale outdoor scenes is an essential task in computer graphics and vision, and particularly crucial for the development of autonomous vehicles and drones. By inferring the geometry and appearance from multi-view captures, existing methods enable precise scene analysis, dataset generation, as well as closed-loop simulation. Radiance field-based methods, initially developed for photorealistic novel view synthesis, have since emerged as a powerful paradigm for scene reconstruction. By modeling the

* These authors contributed equally to this work.
Authors' addresses: Dominik Scheuble*, Mercedes-Benz AG, Boeblingen, Germany and Technical University of Darmstadt, Darmstadt, Germany; Andrea Ramazzina*, Mercedes-Benz AG, Boeblingen, Germany and Technical University of Munich, Munich, Germany; Hanno Holzhüter, Microvision Inc, Seattle, USA and Leibniz University Hannover, Hannover, Germany; Stefano Gasperini, Technical University of Munich, Munich, Germany and Munich Center for Machine Learning, Munich, Germany; Steven Peters, Technical University of Darmstadt, Darmstadt, Germany; Federico Tombari, Technical University of Munich, Munich, Germany and Google, Zurich, Switzerland; Mario Bijelic, Princeton University, Princeton, NJ, USA and Torc, Stuttgart, Germany; Felix Heide, Princeton University, Princeton, NJ, USA and Torc, New York, USA.

SA Conference Papers '25, December 15–18, 2025, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2137-3/2025/12
https://doi.org/10.1145/3757377.3763911

volumetric radiance and density of a scene, they enable both high-quality rendering from arbitrary viewpoints and implicit recovery of the underlying 3D geometry. This is achieved by optimizing a scene representation that maps spatial locations and viewing directions to color and opacity, enabling detailed reasoning about structure and appearance. Originally designed to work with multi-view RGB camera captures, these approaches have also been modified to work with LiDAR, Radar, or NIR captures instead. In particular, for the reconstruction of large-scale outdoor scenes, LiDAR sensing has been widely adopted thanks to its explicit distance information and high temporal resolution, either in conjunction with RGB cameras or as a standalone sensing modality. In this area, recent works [Huang et al. 2023; Tao et al. 2024; Zhang et al. 2024] introduced neural field-based approaches to model the LiDAR's internal processing to reconstruct driving scenes and enable realistic novel-view synthesis. However, these methods operate on the point cloud extracted and preprocessed by the LiDAR's digital signal processor (DSP), instead of directly using the raw transient waveform – the time-resolved intensity of the returning light pulse – captured by the sensor. This has two major drawbacks: First, it propagates any inaccuracies of the peak-finding algorithm used by the LiDAR DSP, such as missed detections, erroneous range estimates, or artifacts in multi-return scenarios; Second, these methods cannot reconstruct the raw waveform for seen or unseen positions, limiting their applicability in re-simulation and excluding their use for any algorithm relying on the raw waveform [Goudreault et al. 2023].

On the other hand, a parallel research direction has explored the extension of neural reconstruction methods to transient waveform captures, usually acquired with a prototype single-photon LiDAR [Klinghoffer et al. 2024; Luo et al. 2025; Malik et al. 2025, 2023]. Over the last few years, transient imaging has emerged as a promising technology, offering unique capabilities beyond conventional sensing modalities. By capturing the time-resolved response of light as it travels through the scene, transient measurements provide fine-grained spatio-temporal information that enables tasks such as perception in scattering media[Lindell and Wetzstein 2020; Zhao et al. 2021], non-line-of-sight (NLOS) imaging [Chen et al. 2020; Rapp et al. 2020a], and robust depth estimation in challenging environments [Rapp et al. 2020b; Scheuble et al. 2025]. These transient-based neural field approaches allow for the reconstruction of the raw waveform, as opposed to only the processed point cloud, showing increasing accuracy in geometry reconstruction and enabling the synthesis of the raw waveform signal for unseen views. However, these works have only been applied to prototypical sensors and in ideal, lab-like conditions, namely, close-range indoor scenarios. Yet, in real-world large-scale outdoor scenes, challenging factors such as complex ambient lighting and high-flux conditions would distort the LiDAR measurements, leading to significant artifacts.

In this work, we aim to eliminate this gap by proposing a novel neural decomposition method that operates with transient waveform captures from an automotive-grade Single Photon Avalanche Diode (SPAD) LiDAR to reconstruct real-world outdoor scenes (Fig. 1). We directly use the raw LiDAR waveforms as a supervision signal for the reconstructed waveform, which we synthesize through a physics-based transient image-formation model. To scale this approach to outdoor in-the-wild scenes, with ranges up to 80 meters, we incorporate ambient light interference, long-range beam divergence, and high-flux conditions in our model. As such, we are able to accurately reconstruct the geometry and appearance of the scene, improving over methods only using LiDAR point clouds, which do not fully leverage the information of the raw waveform, as well as other transient-based approaches, which are unfit for real-world outdoor scenes. Moreover, our physics-based rendering approach allows us to disentangle scene parameters (i.e., distance, normals, and reflectivity) from LiDAR hardware parameters such as the emitted laser power or pulse shape, enabling downstream applications such as novel data generation or hardware settings optimization.

Specifically, we introduce the following contributions:

- We propose a novel scene representation for outdoor scene reconstruction that operates directly on raw transient captures from an automotive-grade SPAD LiDAR, enabling synthesis of the full waveform signal for unseen views, as well as re-rendering of the scene under user-controlled LiDAR parameters (e.g., laser power and pulse shape) and ambient illumination settings.
- We introduce a physics-based transient image formation model that accounts for real-world phenomena such as ambient light interference, long-range beam divergence, and sensor noise, and decomposes measurements into scene geometry (i.e., occupancy and normals) and material properties (e.g., reflectivity) along with ambient illumination.
- We validate our approach on a real-world dataset captured with an automotive-grade SPAD LiDAR, where the method compares favorably in geometry and reflectivity reconstruction compared to existing transient imaging methods and conventional LiDAR methods relying solely on processed 3D point clouds.

2 RELATED WORK

LiDAR as Sensing Modality. LiDAR has emerged as a crucial component for scene understanding and 3D perception in robotics, especially autonomous driving, thanks to its ability to directly measure distance with centimeter accuracy and capture fine-grained spatial structure. Compared to classical camera sensing, LiDAR technology offers centimeter-accurate depth information of the scenes, and it is not affected by low-light and nighttime conditions [Behroozpour et al. 2017; Holzhüter et al. 2023].

Transient Imaging. Transient measures can be acquired with different devices, such as photonic mixer devices [Heide et al. 2013], streak cameras [Velten et al. 2013], or single-photon detectors (SPAD) [Malik et al. 2023]. In SPAD-based LiDAR, a transient response is obtained by firing repeated pulses at a target and compiling the detected photon timestamps into a time-resolved histogram. Thanks to its fine-grained spatio-temporal accuracy, this data has been leveraged for a variety of imaging and analysis tasks, such as perception in scattering media [Lindell and Wetzstein 2020; Zhao et al. 2021], non-line-of-sight imaging [Chen et al. 2020; Rapp et al. 2020a; Xin et al. 2019], and autonomous navigation [Rapp et al. 2020b].

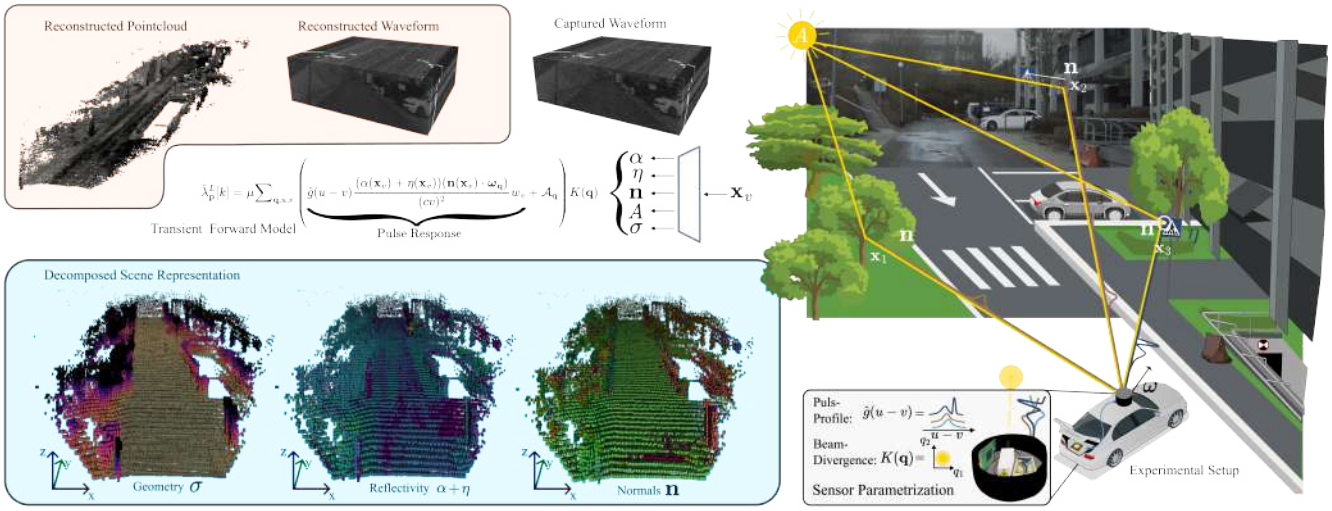

Fig. 2. **Method Overview.** Our method jointly learns scene and sensor parameters from raw LiDAR transients. On the right, we illustrate the measurement process: colored paths represent $\mathbf{x}_\tau$ with integration points $\mathbf{x}_v$, depicting how reflected light is accumulated in the measurement setup (shown below). The left side shows the method decomposition of scene and sensor properties (bottom), while the middle-left shows the core forward model used to reconstruct point clouds and waveforms on the top-left. A captured (non-decomposed) ground truth waveform is shown for comparison next to it, on the right.

Neural 3D Reconstruction. A large body of work has been recently exploring the use of neural rendering-based approaches to reconstruct a scene from posed captures, enabling accurate novel synthesis from novel viewpoints and precise 3D geometry reconstruction. At the core of these breakthroughs lie neural radiance fields [Barron et al. 2021; Chen et al. 2022; Mildenhall et al. 2020; Müller et al. 2022], which treat a scene as a continuous volumetric function of light emission. By using volumetric rendering as the imaging forward model, these methods enable smooth viewpoint interpolation. Different works have also explored alternative ways to encode these fields, including coordinate-based networks [Barron et al. 2021, 2022; Mildenhall et al. 2020; Zhang et al. 2021], dense 3D voxel grids [Chen et al. 2022; Fridovich-Keil et al. 2022; Yu et al. 2021], and hybrid schemes [Barron et al. 2023; Müller et al. 2022; Ost et al. 2025; Tancik et al. 2023]. To speed up both training and inference, several efficient implementations have been proposed [Barron et al. 2023; Chen et al. 2022; Müller et al. 2022; Yu et al. 2021]. These approaches have also been scaled to reconstruct large outdoor environments [Barron et al. 2022; Zhang et al. 2020]. However, modeling large scenes from data collected along a single vehicle trajectory remains challenging due to the narrow, trajectory-aligned viewpoints typical of automotive captures[Guo et al. 2023; Kundu et al. 2022; Liu et al. 2023; Ost et al. 2022; Ramazzina et al. 2023; Tancik et al. 2022; Turki et al. 2023; Wang et al. 2023; Yang et al. 2023].

Departing from RGB-based works, another line of work has instead focused on learning a scene representation from different sensor modalities [Borts et al. 2024; Huang et al. 2023; Ramazzina et al. 2024; Zhang et al. 2024]. In particular, recent works [Huang et al. 2023; Tao et al. 2024; Wu et al. 2024; Zhang et al. 2024; Zheng et al. 2024] have explored the use of LiDAR scans as the main supervision signal, showing accurate 3D scene reconstruction and realistic LiDAR novel view synthesis. However, these methods do not consider the raw LiDAR waveforms and instead only use the point cloud extracted from the black-box DSP[Goudreault et al. 2023] of the sensor, resulting in geometry inaccuracies and a lack of control over the scene properties.

Transient 3D Reconstruction. A parallel research stream has focused on using transient measurements as a supervision signal [Attal et al. 2021; Fujimura et al. 2023; Klinghoffer et al. 2024; Luo et al. 2025; Malik et al. 2025, 2023; Shen et al. 2021]. Some approaches leverage multi-bounce light modeling to reconstruct occluded areas [Fujimura et al. 2023; Klinghoffer et al. 2024; Shen et al. 2021]. In particular, [Fujimura et al. 2023; Shen et al. 2021] learn a neural 3D representation of an occluded object through the transients collected by the SPAD sensor pointed at a diffuse wall, while PlatoNeRF [Klinghoffer et al. 2024] uses the two-bounces signal captured from a SPAD LiDAR to reconstruct the scene geometry from a single view. Other works instead focus on rendering the full transient from novel views [Luo et al. 2025; Malik et al. 2025, 2023]. In particular, TransientNeRF and Transientangelo propose a NeRF and SDF-based representation for single object-scenes [Luo et al. 2025; Malik et al. 2023], allowing for the synthesis of the raw waveform from novel unseen views. Recently, FlyingWithPhotons [Malik et al. 2025] proposed a more general approach to handle non-coaxial light sources and account for more general light transport effects.

However, while some works [Malik et al. 2025, 2023] have shown their applicability with real captures, their implicit image formation model constrains them to simple scenes in indoor settings and prevents them from handling complex outdoor scenes. Our work aims to fill this gap by proposing the first transient-based method to reconstruct real-world scenes captured in complex urban settings. To this end, we formulate a realistic, explicit image formation model that allows us to disentangle scene and hardware parameters.

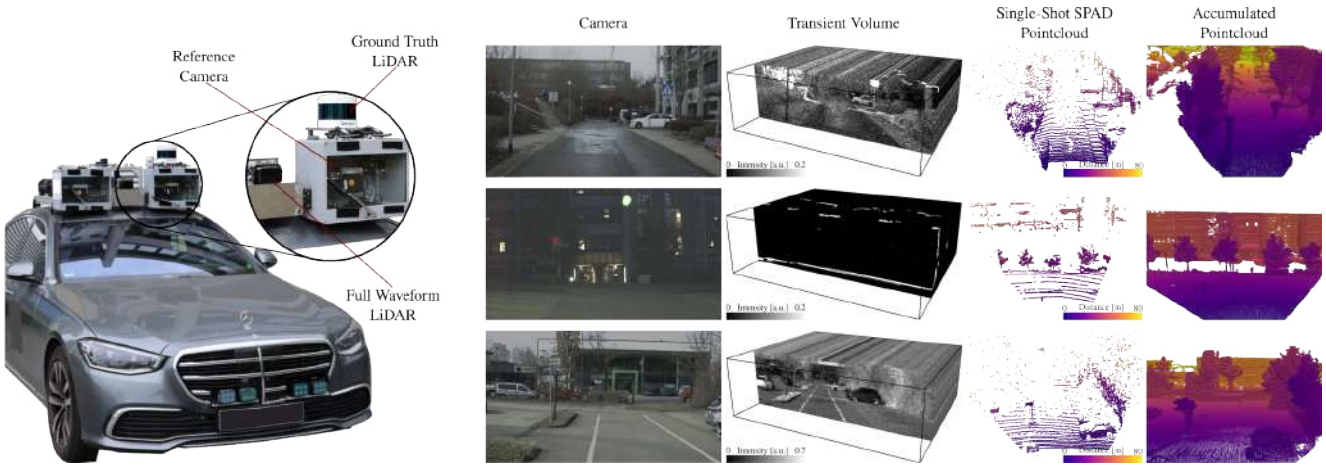

Fig. 3. **Data Capturing Setup** Left: Vehicle setup equipped with Velodyne VLS 128 LiDAR (used for ground truth acquisition), an RGB camera, and the SPAD LiDAR used to capture the transient measures. Right: example captures from our collected dataset across different urban scenarios. From left to right: RGB image, transient capture, sensor-derived point cloud, and accumulated ground truth point cloud from the dedicated Velodyne ground-truth LiDAR.

3 TRANSIENT LASSO

In this section, we introduce Transient LASSO, a novel scene representation that enables decomposing measurements into reflectivity and dense geometry using only a single driving trajectory of transient LiDAR data (see Fig. 2). Our approach is built upon a physics-driven imaging formulation that explicitly models both the back-reflected and ambient light components in the scene (Section 3.1). In particular, we introduce a novel implicit representation for the scene's geometry, material properties, and ambient light (Section 3.2). The entire representation is fitted by reconstructing the LiDAR waveform captures in the time-space domain (Section 3.4).

3.1 Transient LiDAR Forward Model

We use a LiDAR sensor with a 2D SPAD array to capture waveform measurements through time-correlated single photon counting (TCSPC) [Becker 2005; O'Connor and Phillips 1984] sensing. The system emits L consecutive laser pulses and logs photon-trigger events of the returning photon flux into temporal bins $k \in [0, \dots, M-1]$, gradually (with each repetition) constructing the observed waveform histogram ζ, where $M = 2112$ is the total number of time bins. Depending on the scene response, the SPAD operates in the low or high flux regime.

3.1.1 *Photon Flux Model.*

Following [Jarabo et al. 2017; O'Toole et al. 2014], we describe the incident photon flux reaching an infinitesimal surface element $\mathbf{q}$ as

$$\psi_{\mathbf{q}}(t) = \int_{\mathcal{T}} g(t-\tau) H_{\mathbf{q}}(\tau)\, d\tau + a_{\mathbf{q}}(t), \tag{1}$$

where $H_{\mathbf{q}}$ is the scene transient response, $a_{\mathbf{q}}$ the ambient contribution, $g(t)$ is the temporal laser pulse shape, and $\mathcal{T} = [0, M\Delta]$ the light integration time. $\Delta = 266$ps represents the width of each time bin. Please note here that for $\mathbf{q}$, we use the subscript notation $\cdot_{\mathbf{q}}$ in place of $\cdot(\mathbf{q})$ for brevity and compactness. We model the scene response $H_{\mathbf{q}}$ of a single point $\mathbf{x}_{\mathbf{q}} \in \mathbb{R}^3$ at distance d following [Rasshofer et al. 2011], i.e.,

$$H_{\mathbf{q}}(t) = \frac{\rho(\mathbf{n}\cdot\boldsymbol{\omega}_{\mathbf{q}})}{d^2}\delta\left(t - 2\frac{d}{c}\right), \tag{2}$$

where ρ is the surface reflectivity, $\mathbf{n} \in \mathbb{R}^3$ is the surface normal, $\boldsymbol{\omega} \in \mathbb{R}^3$ is the incident light direction, c is the speed of light, and $\delta(\cdot)$ is the Dirac delta function. By substituting Equation (2) in Equation (1), we obtain

$$\psi_{\mathbf{q}}(t) = \frac{\rho(\mathbf{n}\cdot\boldsymbol{\omega}_{\mathbf{q}})}{d^2} g\left(t - 2\frac{d}{c}\right) + a_{\mathbf{q}}. \tag{3}$$

We split ρ into its diffuse component α and specular component η. Due to the co-location of emitter and receiver, we consider the specular component only for retroreflective materials. We also express the ambient component as a function of the diffuse albedo α and ambient light contribution A, that is $a = \alpha A$, yielding

$$\psi_{\mathbf{q}}(t) = \frac{(\alpha+\eta)(\mathbf{n}\cdot\boldsymbol{\omega}_{\mathbf{p}})}{d^2} g\left(t - 2\frac{d}{c}\right) + \alpha A. \tag{4}$$

3.1.2 *Sensor Response.*

Over long distances, the emitted light beam diverges and might hit multiple surfaces from distinct objects. Thus, the recorded waveform is not necessarily unimodal but can have multiple peaks. The total amount of the incident photon flux reaching a SPAD sensor pixel $\mathbf{p}$ with surface area $Q_{\mathbf{p}}$ can be modeled as a 2D surface integral

$$\tilde{\psi}_{\mathbf{p}}(t) = \int_{Q_{\mathbf{p}}} K(\mathbf{q}-\mathbf{p})\psi_{\mathbf{q}} d\mathbf{q}, \tag{5}$$

where K is the spatial Gaussian beam profile, describing the beam divergence, with highest intensity at the center $\mathbf{p}$ and an exponential falloff to the sides, dq is the infinitesimal sensor surface element, and q denotes a point on that surface. In practice, this is implemented as a strided 2D convolution.

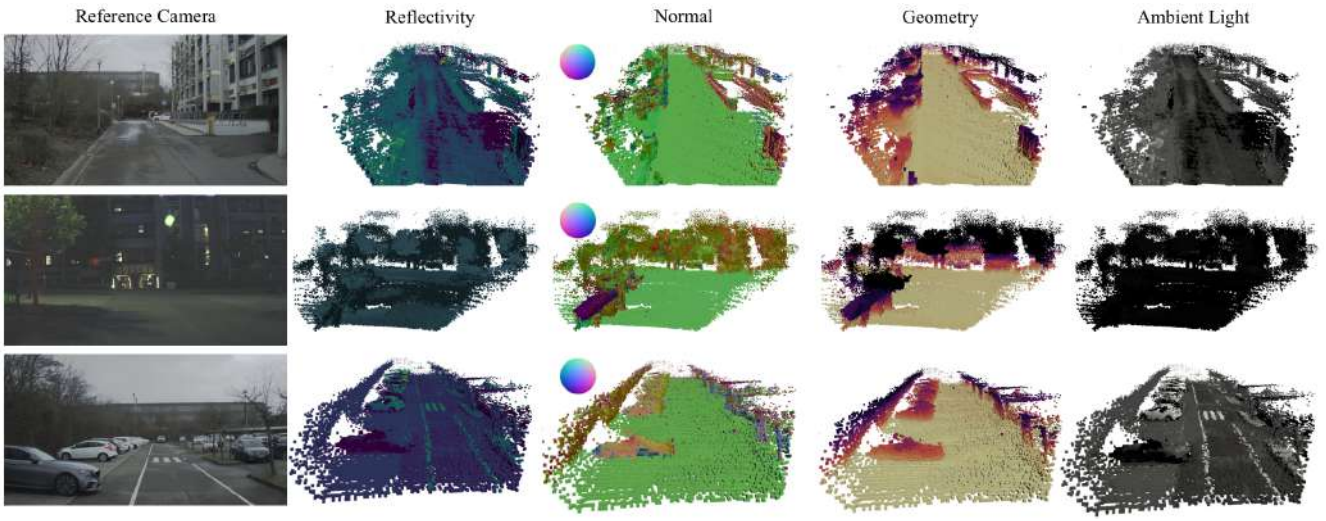

Fig. 4. **Learned Scene Decompositions.** The method disentangles the scene (left) into geometry σ and normals $\mathbf{n}$ from the material properties (reflectivity) and ambient light component A. Our approach accurately recovers fine-grained details, including the correct road and car surface normals, the complex geometry of trees and roadside structures, and the high reflectivity of road markings and traffic signs.

We model the photon detection rate λ for the pixel $\mathbf{p} \in (i, j)$ and bin k as the quantized version of the photon flux ψ, that is

$$\lambda_{\mathbf{p}}[k] = \mu \int_{k\Delta}^{(k+1)\Delta} \tilde{\psi}_{\mathbf{p}}(t)\, dt, \tag{6}$$

where $\mu \in [0, 1]$ denotes the probability of detecting an incoming photon. The complete transient image formation model is hence

$$\lambda_{\mathbf{p}}[k] = \mu \int_{Q_{\mathrm{p}}} K(\mathbf{q}) \left(\int_{k\Delta}^{(k+1)\Delta} \left(\int_{\mathcal{T}} g(t-\tau) H_{\mathbf{q}}(\tau) d\tau \right) + a_{\mathbf{q}}(t)\, dt \right) d\mathbf{q} \tag{7}$$

Finally, the measured waveform $\zeta_{\mathbf{p}}$ can be described as a random variable with respect to $\lambda_{\mathbf{p}}$, which varies for low-flux and high-flux conditions. In low-flux conditions, the response $\zeta^L_{\mathbf{p}}$ can be modeled as a per-bin-independent Poisson distribution[Pediredla et al. 2018; Rapp and Goyal 2017] of $\lambda_{\mathbf{p}}$, that is

$$\zeta^L_{\mathbf{p}}[k] = \mathrm{Poisson}(\lambda_{\mathbf{p}}[k]), \tag{8}$$

allowing us to use the observed sensor signal as unbiased supervision for our model. However, in real-world outdoor scenarios, the low-flux assumption does not always hold, as reflections of traffic signs or other retroreflective material occur frequently. In these cases, SPADs have a heavily distorted response to the incident light. Due to SPAD dead time, the sensor captures a highly distorted pile-up version of the incident flux where the peak is shifted to earlier times [Heide et al. 2018]. In these cases, the transient response $\zeta^H_{\mathbf{p}}$ can be approximated with a distortion kernel h, such that

$$\zeta^H_{\mathbf{p}}[k] = \mathcal{P}\left(\int_{k\Delta}^{(k+1)\Delta} h(t; \Theta) \delta\left(t - 2\frac{d}{c} \right) + a_{\mathbf{p}}\, dt \right) \tag{9}$$

where Θ represents hardware-specific parameters and $\mathcal{P}$ denotes a distribution that applies dedicated noise to the high-flux peak and regular Poisson noise to the ambient component of the waveform. We will detail the description of $\mathcal{P}$ in the Supplementary Material. Since our sensor is operated in free-running mode [Gupta et al. 2019; Ingle et al. 2019], this prevents distortions from high ambient flux as observed by other works, see e.g., [Heide et al. 2018; Ingensand and Kahlmann 2005].

3.2 Implicit Field Representation

We represent the scene as a neural field $f : \{\mathbf{x}\} \rightarrow \{\sigma, \mathbf{n}, \alpha, \eta, A\}$ mapping each point in space $\mathbf{x}$ to its volumetric density σ, normal $\mathbf{n}$, albedo α, retroreflectivity η and ambient light A. We employ two multi-headed neural fields, f_g to model the geometry, and f_a for the appearance. Both fields share a spatial embedding χ via multi-resolution hash encoding $\mathcal{H}_e$ [Müller et al. 2022]:

$$\begin{aligned}
f_{g_o} &: \{\chi\} \longrightarrow \{\sigma\} && \text{Vol. Density} \\
f_{g_n} &: \{\mathbf{x}, \chi\} \rightarrow \{\mathbf{n}\} && \text{Normal} \\
f_{a_\alpha} &: \{\mathbf{x}, \chi\} \rightarrow \{\alpha\} && \text{Albedo} \\
f_{a_\eta} &: \{\mathbf{x}, \chi\} \rightarrow \{\eta\} && \text{Retroreflectivity} \\
f_{a_A} &: \{\mathbf{x}, \chi\} \rightarrow \{A\} && \text{Ambient Component} \\
\text{with } f_e &: \{\mathcal{H}_e(\mathbf{x})\} \rightarrow \{\chi\} && \text{Spatial Embedding}
\end{aligned}$$

Here, normal, ambient, and reflectivity are conditioned on volumetric embeddings χ via the field f_e. We use the geometry field f_g to model the volumetric density σ and surface normal $\mathbf{n}$, and the appearance field f_a to represent the diffusive albedo α and the ambient light A. Moreover, we employ a proposal sampler f_P [Barron et al. 2022] for efficient sampling. Further details on architecture are provided in the Supplementary Material.

3.3 Transient Field Learning

We decompose the scene from a set of N transient captures with corresponding poses. For each pixel $\mathbf{p}$, we reconstruct the noise-free transient waveform $\lambda^L_{\mathbf{p}}$ for low-flux conditions, through time-resolved volume rendering using the model described in Section 3.1 along an integrated ray $\mathbf{x}(\tau)$ and write in short $\mathbf{x}_\tau$. Specifically, we

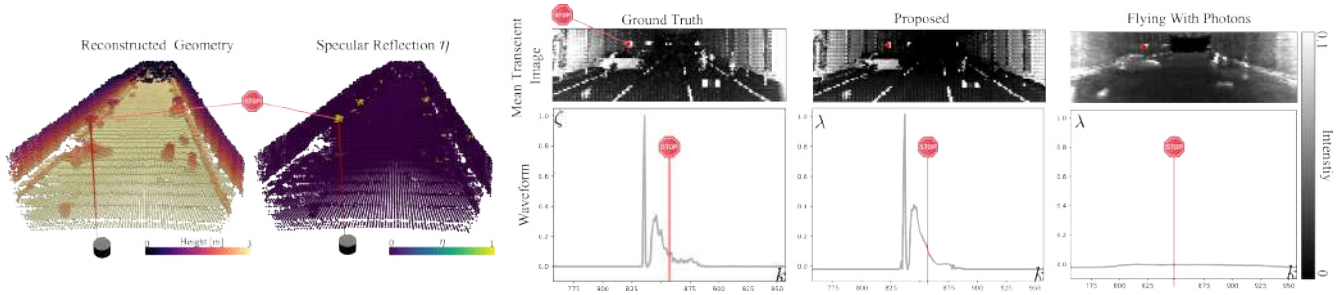

Fig. 5. **Example of high-flux transient components of a retro-reflective traffic sign.** From left to right, we show the reconstructed geometry, the predicted high-flux response, the captured signal, our proposed method (Transient LASSO), and the baseline (Flying with Photons[2025]). The top row presents the average across all waveform bins k, while the bottom row displays the response of a single pixel located on the red retro-reflective stop sign. Notably, only our method successfully reconstructs the high-flux component due to its tailored handling.

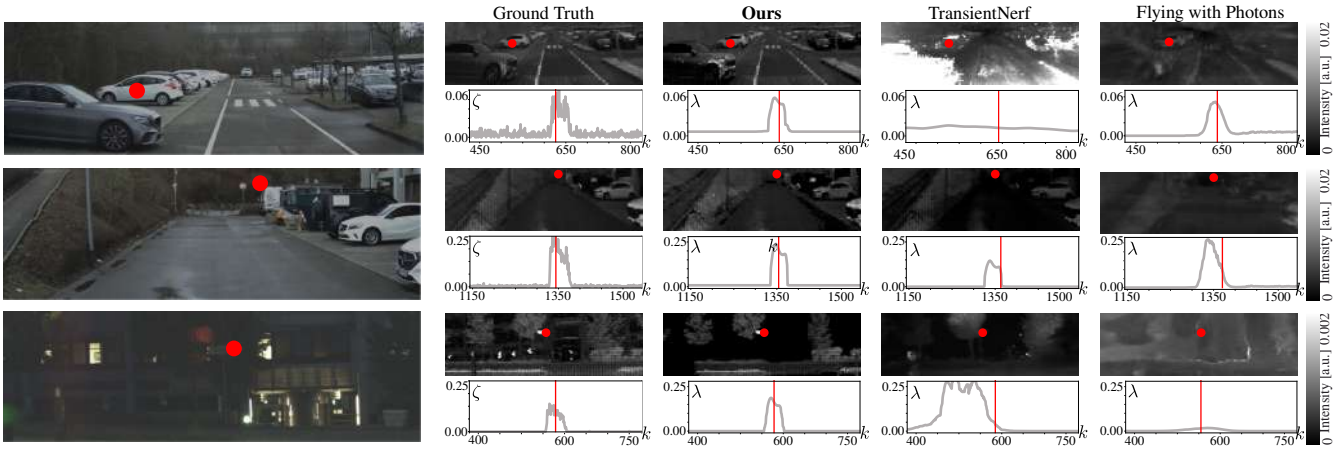

Fig. 6. **Example of the nominal scene response for in-the-wild conditions.** From left to right, we show an image of the scene, the captured signal, our proposed method (Transient LASSO), and the baseline methods (TransientNerf[2023] and Flying with Photons[2025]). The top row shows the average across all waveform bins k, while the bottom row displays the response of a single pixel located in the image. All methods suppress the noise present in the raw data and recover the underlying signal λ. Our approach demonstrates improved fidelity in reconstructing the nominal response due to the accurate forward model.

model the scene response as

$$H_{\mathbf{q}}(\tau) = T^2(\tau)\sigma(\mathbf{x}_\tau)\frac{(\alpha(\mathbf{x}_\tau)+\eta(\mathbf{x}_\tau))\,(\mathbf{n}(\mathbf{x}_\tau)\cdot\boldsymbol{\omega}_q)}{(c\tau)^2}, \tag{10}$$

and the response $\mathcal{A}_q$ to the ambient light as

$$\mathcal{A}_{\mathbf{q}} = \int_0^\infty \alpha(\mathbf{x}_\tau)A(\mathbf{x}_\tau)T(\tau)\sigma(\mathbf{x}_\tau)d\tau \tag{11}$$

For both cases, we assume exponential scattering with the transmittance as in [Mildenhall et al. 2020],

$$T(\tau) = \exp\left(\int_{t_0}^{\tau}\sigma(\mathbf{x}(t))\,dt\right). \tag{12}$$

Here, $T^2(\tau)$ is the squared transmittance accounting for both the incident and return path for the active scene response. For each pixel p and each sampled ray p in the field of view, we cast the ray $\mathbf{x}_\tau$ from its origin $\mathbf{o}_\mathbf{p}$ into the scene with viewing direction $\boldsymbol{\omega}_q$.

$$\mathbf{x}_\tau = \mathbf{o}_\mathbf{p} + c\tau(\boldsymbol{\omega}_q + \vartheta_\mathbf{p}). \tag{13}$$

We learn an offset ϑ initialized at the beginning to zero for each sensor pixel. This offset allows the model to account for systematic deviations in the scanning mechanism, effectively correcting for spatial distortions introduced by optical or mechanical imperfections, which are challenging to calibrate due to the low resolution and large field of view. This leads to the total integral being

$$\hat{\lambda}^L_\mathbf{p}[k] = \mu\int_{\mathcal{Q}_\mathbf{p}} K(\mathbf{q})\int_{k\Delta}^{(k+1)\Delta}\mathcal{A}_\mathbf{q} + \int_{\mathcal{T}} g(t-\tau)H_q(\mathbf{x}_\tau)d\tau dt\,d\mathbf{q}, \tag{14}$$

We approximate the beam integral as a linear combination of rays sampled in $\mathcal{Q}_\mathbf{p}$. At the same time, we numerically estimate the path integral via numerical quadrature as in [Mildenhall et al. 2020], by sampling positions $\mathbf{x}_v$ along ray $\mathbf{x}_\tau$ We query the geometry and appearance fields f_g, f_a to obtain the volume density σ, surface normal $\mathbf{n}$, material properties α, η, and ambient component A. The transient response can then be computed as

$$\hat{\lambda}^L_\mathbf{p}[k] = \mu\sum_{\mathbf{q},u,v}\left[\text{Pulse}_{u,v,\mathbf{q}}\,w_v + \mathcal{A}_\mathbf{q}\right]K(\mathbf{q}),\ \text{with} \tag{15}$$

$$\text{Pulse}_{u,v,\mathbf{q}} = \hat{g}(u-v)\frac{(\alpha(\mathbf{x}_v)+\eta(\mathbf{x}_v))(\mathbf{n}(\mathbf{x}_v)\cdot\boldsymbol{\omega}_\mathbf{q})}{(cv)^2}\ \text{and} \tag{16}$$

$$w_v = \exp\left(-2\Sigma^v_{l=0}\sigma_l\delta_l\right)(1-\exp(-\sigma_v\delta_v)), \tag{17}$$

$$\mathcal{A}_\mathbf{q} = \sum_v \alpha(\mathbf{x}_v)A(\mathbf{x}_v)\bar{w}_v \tag{18}$$

$$\bar{w}_v = \exp\left(-\Sigma^v_{l=0}\sigma_l\delta_l\right)(1-\exp(-\sigma_v\delta_v)). \tag{19}$$

In practice, we first compute the pulse response for the full waveform ($H_{\mathbf{q}}$) and then convolve it with the laser pulse. This can be efficiently done via a 1D convolution, with the kernel $\hat{g}$, representing the discretized pulse shape. We keep $\hat{g}$ as optimizable to improve over the initial estimate and make up for calibration inaccuracies and manufacturing differences, causing a deviation in pulse shape. If the estimated photon flux $\hat{\lambda}_{\mathbf{p}}[k]$ exceeds a threshold B_T, we consider the full transient to be high-flux. In this case, we approximate Equation (9) as a 1D convolution between a learned kernel $\hat{h}$ and the weight distribution $w[k] \quad \text{for } k = 0, ..., M-1$. Then, the total reconstructed (noise-free) transient $\hat{\lambda}^T_{\mathbf{p}}$ becomes

$$\hat{\lambda}^T_{\mathbf{p}} = \begin{cases} \hat{\lambda}^L_{\mathbf{p}}, & \text{if } \hat{\lambda}^L_{\mathbf{p}}[k] < B_T \ \forall k \in \{0, ..., M-1\} \\ \hat{\lambda}^H_{\mathbf{p}} = (\hat{h} * w) + \mathcal{A}_{\mathbf{p}}, & \text{otherwise.} \end{cases} \tag{20}$$

3.4 Training Supervision

We fit our neural representation by supervising the predicted full waveform with a reconstruction loss $\mathcal{L}_b$ using the observed transient measurements, together with a partial pseudo-ground truth depth and normal supervision via $\mathcal{L}_g$. We also regularize the learned ambient component A through $\mathcal{L}_a$, specular component η with $\mathcal{L}_h$ and normals $\mathbf{n}$ through $\mathcal{L}_n$. All loss components comprising the total loss function are described in the following.

Reconstruction Loss. We supervise the reconstructed transients with the ground truth captures $\zeta_{\mathbf{p}}[k]$

$$\mathcal{L}_b = \sum_{\mathbf{p},k} (1-s)||\hat{\lambda}^L_{\mathbf{p}}[k] - \zeta_{\mathbf{p}}[k]||_2 + s||\hat{\lambda}^H_{\mathbf{p}}[k] - \zeta_{\mathbf{p}}[k]||_2, \tag{21}$$

where the mask s indicates the high-flux transients, obtained with a heuristic detailed in the Supplementary Material.

Geometry Supervision. We also supervise the density and normal by extracting a pseudo-ground truth normal $\tilde{\mathbf{n}}$ and depth $\tilde{d}$ from the transient measurements. While these estimates are noisy and can only be extracted for a subset of the measured transients, they promote the correct disentanglement of geometry and appearance

$$\mathcal{L}_g = \sum_{\mathbf{p} \in P_D} ||\hat{d}_{\mathbf{p}} - \tilde{d}_{\mathbf{p}}||_2 + \sum_{\mathbf{p} \in P_n} ||\hat{\mathbf{n}}_{\mathbf{p}} - \tilde{\mathbf{n}}_{\mathbf{p}}||_2, \tag{22}$$

where P_D and P_n indicate the pixels where the estimated pseudo-ground truth depth and normal are available, and $\hat{d}_{\mathbf{p}}$ and $\hat{\mathbf{n}}_{\mathbf{p}}$ are the estimated depth and normal for a pixel $\mathbf{p}$.

Normal Consistency. For each sampled point $\mathbf{x}$, we enforce a consistency between the predicted normal $\mathbf{n}(\mathbf{x})$ and the density gradient $\dot{\mathbf{n}} = -\nabla\mathbf{x}/||\nabla\mathbf{x}||$, as well as penalizing normals that are backfacing the LiDAR as

$$\mathcal{L}_n = \sum_{\mathbf{x}} \left(\|\mathbf{n}(\mathbf{x}) - \dot{\mathbf{n}}(\mathbf{x})\|_2 + \max(0, \mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\omega})^2 \right). \tag{23}$$

Retroreflector Loss. We also encourage a high specular component η for the surfaces responsible for high-flux transients:

$$\mathcal{L}_h = -\sum_p s_p \sum_v w_v \eta_v. \tag{24}$$

Ambient Light Consistency. To avoid the ambient light component from overfitting the scene, we enforce it to be locally smooth, that is

$$\mathcal{L}_a = \sum_{\mathbf{x}} ||A(\mathbf{x}) - A(\mathbf{x} + \epsilon_{\mathbf{x}})||_2. \tag{25}$$

The neighborhood parameter ϵ_x is set high at the beginning of the training and gradually decreased.

Total Training Loss. The final optimization objective is

$$\mathcal{L} = v_1\mathcal{L}_b + v_2\mathcal{L}_g + v_3\mathcal{L}_n + v_4\mathcal{L}_a + v_5\mathcal{L}_h. \tag{26}$$

The hyperparameters $v_{1,..,5}$ are provided in the Supplement.

4 IMPLEMENTATION DETAILS

We train our model for 200,000 steps, using a batch size of 64 transients. We use the Adam optimizer, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate of 1×10^{-3} for the networks f_g, f_a, f_e, f_p, and 1×10^{-4} for the optimizable parameters $\vartheta_{\mathbf{p}}, \hat{\mathbf{g}}, \hat{h}$. We train a scene on a single A100 for approximately 6 hours.

5 DATASET

To assess the proposed algorithm, we collect a diverse set of 10 sequences in day and night conditions, in urban outdoor scenes, and in a controlled environment where environmental parameters such as illumination can be set accurately. To this end, we equipped a test vehicle with a sensor setup, as shown in Figure 3, comprising an automotive-grade RGB camera (OnSemi AR0230) and two automotive-grade production-grade LiDAR sensors, namely Velodyne VLS128, and a SPAD LiDAR (MicroVision Movia-L). Analogous to [Scheuble et al. 2025], the SPAD LiDAR captures 40×128 full waveform histograms with 2112 time bins per frame. Every waveform results from a multiple-pulse emissions TCSPC measurement performed with a vertical-cavity surface-emitting laser (VCSEL) array in conjunction with the 2D SPAD array. This enables simultaneous collection of full waveforms from all VCSEL-SPAD combinations. As ground truth for the outdoor scenes, we extract a dense point cloud by slow-driving across the scene and aggregate the Velodyne LiDAR scans, as detailed in the Supplementary Material. For the controlled environment, we instead rely on a Leica ScanStation P30 laser scanner with 1.2mm range accuracy. We project the ground truth scans into the SPAD LiDAR field-of-view by considering its scan pattern. Each of the 10 scenes comprises 100 captures, recorded over 200 seconds in suburban areas. For novel view synthesis (section 6.2), every fifth capture is held out during training and used as the test set, following [Huang et al. 2023; Tao et al. 2024].

6 APPLICATIONS

In this section, we validate the proposed method using the collected dataset described in Section 5. Specifically, in Section 6.1 we assess the 3D geometry reconstruction quality of the proposed method

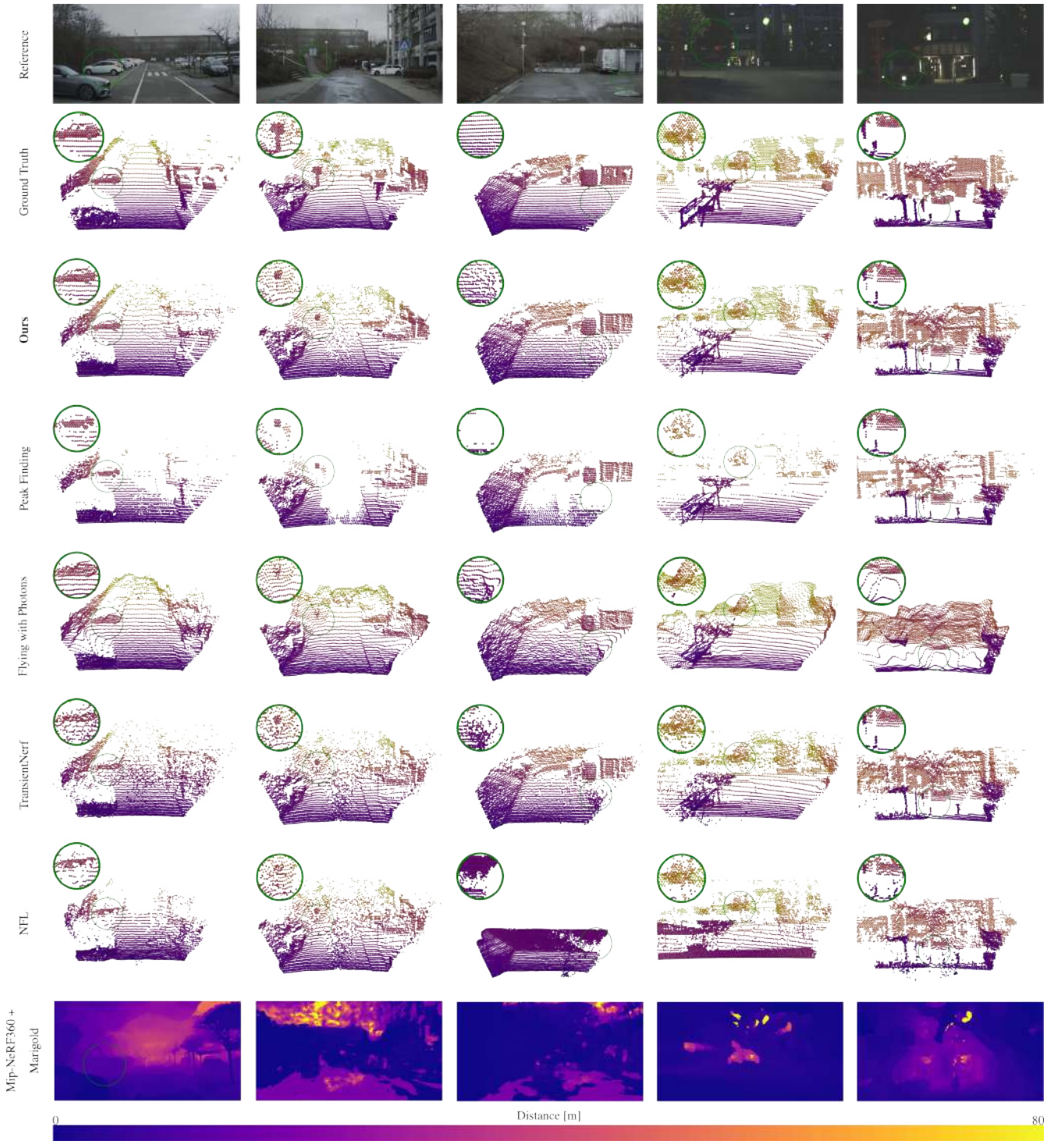

Fig. 7. **Qualitative comparison of predicted point clouds.** For our proposed method (LASSO), the sensor's internal peak-finding, and recent methods by [Malik et al. 2025], [Malik et al. 2023], [Huang et al. 2023], and [Barron et al. 2021] using [Ke et al. 2024] for depth supervision. Each column presents a different scene (five in total), with green zoom-in regions highlighting key differences. Note that the image-based methods fail to reconstruct the scene due to inherent scale ambiguity and limited viewpoints from a driving trajectory. Among all methods, only our approach consistently reconstructs the full scene—from the vehicle and traffic sign on the left, to the wet ground, cluttered treetops, and distant small structures like the light pole on the right.

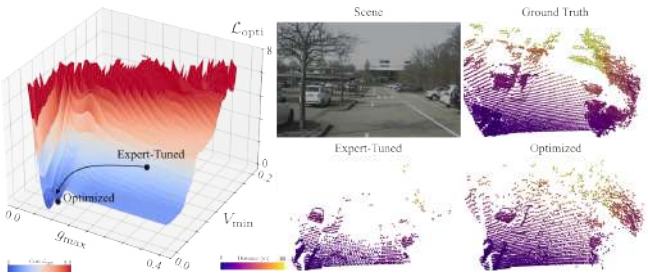

Fig. 8. **Lidar In-the-loop Optimization with Transient LASSO.** We optimize DSP parameters using the learned scene reconstruction to produce waveforms as input to a lidar DSP. The optimization landscape $\mathcal{L}_{\text{opti}}$ is shown on the left, illustrating how it varies with changes in parameters $V\text{min}$ and g_{max}. The differentiable, decomposed representation enables the use of gradient-based methods to optimize peakfinding[Goudreault et al. 2023], enhancing the point cloud captured from the example scene (top middle). The optimized DSP parameters (bottom right) significantly outperform the expert-selected values (bottom middle), resulting in a more complete and measurable reconstruction of the scene (top right).

against state-of-the-art approaches, while in Section 6.2, we investigate its performance for full-waveform rendering from unseen views. Finally, in Section 6.3, we leverage the disentanglement to optimize the LiDAR parameters (Section 6.3).

We compare our models with state-of-the-art neural scene reconstruction using different sensing modalities. In particular, we use Gaussian Splatting [Kerbl et al. 2023] and Mip-NeRF 360[Barron et al. 2022] + Marigold[Ke et al. 2024] for reconstruction from RGB images, NFL[Huang et al. 2023] and LiDAR-NeRF[Tao et al. 2024] for LiDAR pointcloud scans, as well as Transient NeRF[Malik et al. 2023] and Flying with Photons[Malik et al. 2025] for raw transient measures. We also include in the comparison the result obtained by running a common DSP approach[Goudreault et al. 2023] to extract a pointcloud from the transient measure (i.e. "Sensor-Derived").

6.1 3D Scene Reconstruction

We evaluate the 3D reconstruction quality of our proposed framework, reporting quantitative and qualitative comparison results (Figure 7). The quantitative comparison is performed by comparing the ground truth (described in Section 5) with a pointcloud of the scene obtained by querying the fitted neural scene representation. As metrics, we use the Chamfer Distance (CD), Distance Accuracy (one-sided CD), as well as Recall on a pointcloud level. As can be seen in Table 1, our proposed method outperforms competing baselines operating on both raw (waveform) [Malik et al. 2025, 2023] or processed (extracted pointclouds) [Huang et al. 2023; Tao et al. 2024] and RGB signal. Compared to NFL[Huang et al. 2023], our approach directly operates on the raw waveform LiDAR signal, leveraging important information lost when extracting the pointcloud, where typical data compression rates of 1000 can be seen. On the other hand, compared to competing methods operating on transient measures [Malik et al. 2025, 2023], our approach is the only one capable of correctly modeling complex outdoor scenes, thanks to the disentanglement between scene geometry, surface reflectivity and ambient light, as exemplified in Figure 4.

Table 1. **Quantitative Results for 3D Scene Reconstruction.** We compare with existing methods using three different modalities, namely Images (RGB), Point clouds, and Transients. Our method achieves the best performance across all compared approaches.

Method	Modality	CD [m] ↓	Dist. Acc. [m] ↓	Recall [%] ↑
Gaussian Splatting [2023]	RGB	2.357	1.018	35.23
Mip-NeRF 360 [2021] + Marigold [2024]	RGB	1.872	0.970	40.50
LiDAR-NeRF [2024]	Point Cloud	1.379	0.704	40.26
NFL [2023]	Point Cloud	1.827	0.485	41.35
TransientNeRF [2023]	Transient	0.837	0.500	56.88
Flying-With-Photons [2025]	Transient	0.826	0.493	56.72
Ours	Transient	**0.642**	**0.359**	**60.12**

Table 2. **Quantitative Results for Novel View Synthesis on Held-out Poses.** We compare our method to [2023] and [2025]. In addition, we show an ablation with our approach, removing either the ambient modeling or the optimization of sensor intrinsic, as the pulse estimation by removing respective modules. Our approach improvements across ablations and exceeds reference methods by **10%** PSNR and **51%** SSIM.

Method	PSNR [dB] ↑	SSIM ↑
TransientNeRF [2023]	39.15	0.373
Flying-With-Photons [2025]	50.02	0.600
Ours w/o ambient	49.67	0.515
Ours w/o pulse optimization	52.37	0.884
Ours	**55.13**	**0.912**

6.2 Novel View Synthesis

We also evaluate our method's capability to render transients for novel unseen views. At the unseen held-out poses, we evaluate PSNR between the predicted waveform $\hat{\lambda}^T_{\mathbf{p}}$ and the recorded signal $\zeta_{\mathbf{p}}$ following [Malik et al. 2023], as well as SSIM between the 2D time-integrated transient images obtained by averaging waveforms across temporal bins [Godard et al. 2019]. As reported quantitatively in Table 2, our model outperforms the next-best method[Malik et al. 2025] by more than 10% dB in PSNR. When evaluating on SSIM, the gap increases to 51.9%, underscoring the superiority of our model in recovering finer structures and spatial details. This can be attributed to the fact that all previous transient-based neural reconstruction approaches are inherently limited by their scene modeling and cannot correctly represent the ambient light contribution, the normal-incidence angle effect, or the high-flux regime cases. This can also be observed in Figures 5 and 6, where the transients rendered by Transient-NeRF and Flying-with-Photons cannot correctly handle the ambient light component and high-flux occurrences, resulting in distorted and blurry renderings.

6.3 LASSO-in-the-Loop Parameter Optimization

Recently, [Goudreault et al. 2023] have demonstrated the benefits of optimizing LiDAR sensor and DSP processing parameters directly for downstream applications. However, since LiDAR sensors typically restrict access to these parameters, they validated their method in a simulated environment. In contrast, our disentanglement of scene and sensor parameters enables direct optimization on captured real-world data. To demonstrate this, we use a conventional argmax peakfinding algorithm [Behroozpour et al. 2017; Goudreault

et al. 2023] with a noise floor threshold $V_{\min}$ to extract point clouds from waveforms ζ. We optimize laser power $g_{\max}$ and threshold $V_{\min}$ for a high-fidelity reconstruction using LASSO in-the-loop.

Transient LASSO enables rerendering with a continuous parameterization for parameters, such as pulse shape, laser intensity, and gain. This continuous parameterization yields a smooth and differentiable loss landscape (see Figure 8), enabling optimization of sensor and DSP hyperparameters using gradient-based methods. As a result, our method supports principled, end-to-end sensor design and configuration based on task-specific objectives. To demonstrate this, we minimize the following cost function

$$\mathcal{L}_{\text{opti}} = c_1 \text{CD}\left(\boldsymbol{O}\left(g_{\max}, V_{\min}\right), \boldsymbol{O}_{\text{GT}}\right) + c_2 g_{\max}^2, \tag{27}$$

where CD is Chamfer Distance between extracted point cloud $\boldsymbol{O}$ and ground truth $\mathbf{O}_{\text{GT}}$. $c_1 = 1$ and $c_2 = 0.2$ are scalar weights. Details for peak finding are provided in the Supplementary Material.

We evaluate the optimization on a sunny scene, as shown in Figure 8. Here, a lower $V_{\min}$ increases range, but introduces additional clutter points from ambient light; raising laser power reduces clutter, but incurs economic penalties as described in Eq. (27). Leveraging our method, we can use a stochastic or gradient-based optimizer to find an optimized set of parameters $g_{\max}$ and $V_{\min}$ that yields an improved scene reconstruction of 1.169m CD (21.47%) over the expert-tuned parameters (1.420m CD).

7 CONCLUSION

In this work, we introduce Transient LASSO as a novel neural reconstruction framework that leverages the rich spatio-temporal cues of raw transient measurements for large-scale outdoor scene reconstruction. As a test-time optimization method, the method hinges on accurate modeling of the sensor intrinsic parameters, namely pulse shape, optics distortions, and high-photon-flux response, as well as the scene ambient illumination, reflectance, and normals. Our method is capable of decomposing accurate geometry and appearance even in cluttered urban environments where existing methods struggle, see Fig. 7. Beyond high-fidelity geometry and reflectance, Transient LASSO enables downstream tasks such as transient-guided super-resolution and automatic sensor-parameter optimization. We hope that releasing our method will catalyze further research on transient imaging and encourage the adoption of time-resolved sensors in real-world graphics and vision applications for outdoor long-range scenes.

ACKNOWLEDGMENTS

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. Dominik Scheuble and Andrea Ramazzina have been supported by the German Ministry BMWE within the project “NXT GEN AI METHODS”. We thank Alex Zuber for his support.

REFERENCES

Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O'Toole. 2021. Törf: Time-of-flight radiance fields for dynamic scene view synthesis. *Advances in neural information processing systems* 34 (2021), 26289–26301.

Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields . In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 5835–5844. https://doi.org/10.1109/ICCV48922.2021.00580

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 5460–5469. https://doi.org/10.1109/CVPR52688.2022.00539

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields . In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 19640–19648. https://doi.org/10.1109/ICCV51070.2023.01804

Wolfgang Becker. 2005. *Advanced Time-Correlated Single Photon Counting Techniques*. Springer, Berlin, Heidelberg, Heidelberg, BW, Germany. https://doi.org/10.1007/3-540-28882-1

Behnam Behroozpour, Phillip AM Sandborn, Ming C Wu, and Bernhard E Boser. 2017. Lidar system architectures and circuits. *IEEE Communications Magazine* 55, 10 (2017), 135–142.

David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, Mario Bijelic, and Felix Heide. 2024. Radar Fields: Frequency-Space Neural Scene Representations for FMCW Radar. In *ACM SIGGRAPH 2024 Conference Papers* (Denver, CO, USA) *(SIGGRAPH '24)*. Association for Computing Machinery, New York, NY, USA, Article 130, 10 pages. https://doi.org/10.1145/3641519.3657510

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. TensoRF: Tensorial Radiance Fields. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 333–350.

Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz, and Felix Heide. 2020. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Transactions on Graphics (ToG)* 39, 6 (2020), 1–18.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 5491–5500. https://doi.org/10.1109/CVPR52688.2022.00542

Yuki Fujimura, Takahiro Kushida, Takuya Funatomi, and Yasuhiro Mukaigawa. 2023. NLOS-NeuS: Non-line-of-sight Neural Implicit Surface . In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 10498–10507. https://doi.org/10.1109/ICCV51070.2023.00966

Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel Brostow. 2019. Digging Into Self-Supervised Monocular Depth Estimation . In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 3827–3837. https://doi.org/10.1109/ICCV.2019.00393

Felix Goudreault, Dominik Scheuble, Mario Bijelic, Nicolas Robidoux, and Felix Heide. 2023. LiDAR-in-the-Loop Hyperparameter Optimization . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 13404–13414. https://doi.org/10.1109/CVPR52729.2023.01288

Jianfei Guo, Nianchen Deng, Xinyang Li, Yeqi Bai, Botian Shi, Chiyu Wang, Chenjing Ding, Dongliang Wang, and Yikang Li. 2023. StreetSurf: Extending Multi-view Implicit Surface Reconstruction to Street Views. arXiv:2306.04988 [cs.CV] https://arxiv.org/abs/2306.04988

Anant Gupta, Atul Ingle, and Mohit Gupta. 2019. Asynchronous Single-Photon 3D Imaging . In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 7908–7917. https://doi.org/10.1109/ICCV.2019.00800

Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. 2018. Sub-Picosecond Photon-Efficient 3D Imaging using Single-Photon Sensors. *Scientific reports* 8, 1 (2018), 1–8.

Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. 2013. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)* 32, 4 (2013), 1–10.

Hanno Holzhüter, Jörn Bödewadt, Shima Bayesteh, Andreas Aschinger, and Holger Blume. 2023. Technical Concepts of Automotive LiDAR Sensors: A Review. *Optical Engineering* 62, 3 (2023), 031213–031213.

Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. 2023. Neural LiDAR Fields for Novel View Synthesis . In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 18190–18200. https://doi.org/10.1109/ICCV51070.2023.01672

Hilmar Ingensand and Timo Kahlmann. 2005. Single-Photon Detection for High Precision Ranging – A Trade-off Study. In *1st Range Imaging Research Day Proceedings*. Hilmar Ingensand and Timo Kahlmann, Zurich, Switzerland, 33–42.

Atul Ingle, Andreas Velten, and Mohit Gupta. 2019. High Flux Passive Imaging With Single-Photon Sensors . In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 6753–6762. https://doi.org/10.1109/CVPR.2019.00692

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

Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. 2024. Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 9492–9502. https://doi.org/10.1109/CVPR52733.2024.00907

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, and Rakesh Ranjan. 2024. PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 14565–14574. https://doi.org/10.1109/CVPR52733.2024.01380

Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. 2022. Panoptic Neural Fields: A Semantic Object-Aware Neural Scene Representation . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 12861–12871. https://doi.org/10.1109/CVPR52688.2022.01253

David B Lindell and Gordon Wetzstein. 2020. Three-dimensional imaging through scattering media based on confocal diffuse tomography. *Nature communications* 11, 1 (2020), 4517.

Jeffrey Yunfan Liu, Yun Chen, Ze Yang, Jingkang Wang, Sivabalan Manivasagam, and Raquel Urtasun. 2023. Real-Time Neural Rasterization for Large Scenes . In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 8382–8393. https://doi.org/10.1109/ICCV51070.2023.00773

Weihan Luo, Anagh Malik, and David B. Lindell. 2025. Transientangelo: Few-Viewpoint Surface Reconstruction Using Single-Photon Lidar . In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE Computer Society, Los Alamitos, CA, USA, 8723–8733. https://doi.org/10.1109/WACV61041.2025.00845

Anagh Malik, Benjamin Attal, Andrew Xie, Matthew O'Toole, and David B. Lindell. 2025. Neural Inverse Rendering from Propagating Light . In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 10534–10544. https://doi.org/10.1109/CVPR52734.2025.00985

Anagh Malik, Parsa Mirdehghan, Sotiris Nousias, Kiriakos N. Kutulakos, and David B. Lindell. 2023. Transient neural radiance fields for lidar view synthesis and 3D reconstruction. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 3133, 13 pages.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 405–421.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)* 41, 4 (2022), 1–15.

Desmond V. O'Connor and David Phillips. 1984. *Time-Correlated Single Photon Counting*. Vol. 33. Academic Press, London, UK. https://doi.org/10.1016/B978-0-12-524140-3.X5001-1

Julian Ost, Issam Laradji, Alejandro Newell, Yuval Bahat, and Felix Heide. 2022. Neural Point Light Fields . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 18398–18408. https://doi.org/10.1109/CVPR52688.2022.01787

Julian Ost, Andrea Ramazzina, Amogh Joshi, Maximilian Bömer, Mario Bijelic, and Felix Heide. 2025. LSD-3D: Large-Scale 3D Driving Scene Generation with Geometry Grounding. *arXiv preprint arXiv:2508.19204* (2025).

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

Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. 2018. Signal Processing Based Pile-up Compensation for Gated Single-Photon Avalanche Diodes. arXiv:1806.07437 [physics.ins-det] https://arxiv.org/abs/1806.07437

Andrea Ramazzina, Mario Bijelic, Stefanie Walz, Alessandro Sanvito, Dominik Scheuble, and Felix Heide. 2023. ScatterNeRF: Seeing Through Fog with Physically-Based Inverse Neural Rendering. *arXiv preprint arXiv:2305.02103* (2023).

Andrea Ramazzina, Stefanie Walz, Pragyan Dahal, Mario Bijelic, and Felix Heide. 2024. Gated fields: Learning scene reconstruction from gated videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10530–10541.

Joshua Rapp and Vivek K Goyal. 2017. A Few Photons Among Many: Unmixing Signal and Noise for Photon-Efficient Active Imaging. *IEEE Transactions on Computational Imaging* 3, 3 (2017), 445–459. https://doi.org/10.1109/TCI.2017.2706028

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. 2020a. Seeing around corners with edge-resolved transient imaging. *Nature communications* 11, 1 (2020), 5929.

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

Ralph H Rasshofer, Martin Spies, and Hans Spies. 2011. Influences of weather phenomena on automotive laser radar systems. *Advances in radio science* 9 (2011), 49–60.

Dominik Scheuble, Hanno Holzhueter, Steven Peters, Mario Bijelic, and Felix Heide. 2025. Lidar Waveforms are Worth 40x128x33 Words. *2025 IEEE/CVF International Conference on Computer Vision (ICCV*.

Siyuan Shen, Zi Wang, Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiying Li, and Jingyi Yu. 2021. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 7 (2021), 2257–2268.

Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben P. Mildenhall, Pratul Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. 2022. Block-NeRF: Scalable Large Scene Neural View Synthesis . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 8238–8248. https://doi.org/10.1109/CVPR52688.2022.00807

Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa. 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 72, 12 pages. https://doi.org/10.1145/3588432.3591516

Tang Tao, Longfei Gao, Guangrun Wang, Yixing Lao, Peng Chen, Hengshuang Zhao, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, and Kaicheng Yu. 2024. LiDAR-NeRF: Novel LiDAR View Synthesis via Neural Radiance Fields. In *Proceedings of the 32nd ACM International Conference on Multimedia* (Melbourne VIC, Australia) *(MM '24)*. Association for Computing Machinery, New York, NY, USA, 390–398. https://doi.org/10.1145/3664647.3681482

Haithem Turki, Jason Y. Zhang, Francesco Ferroni, and Deva Ramanan. 2023. SUDS: Scalable Urban Dynamic Scenes . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 12375–12385. https://doi.org/10.1109/CVPR52729.2023.01191

Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. 2013. Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* 32, 4, Article 44 (July 2013), 8 pages. https://doi.org/10.1145/2461912.2461928

Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. 2023. Neural Fields Meet Explicit Geometric Representations for Inverse Rendering of Urban Scenes . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 8370–8380. https://doi.org/10.1109/CVPR52729.2023.00809

Hanfeng Wu, Xingxing Zuo, Stefan Leutenegger, Or Litany, Konrad Schindler, and Shengyu Huang. 2024. Dynamic LiDAR Re-Simulation Using Compositional Neural Fields. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 19988–19998. https://doi.org/10.1109/CVPR52733.2024.01889

Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. 2019. A Theory of Fermat Paths for Non-Line-Of-Sight Shape Reconstruction . In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 6793–6802. https://doi.org/10.1109/CVPR.2019.00696

Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. 2023. UniSim: A Neural Closed-Loop Sensor Simulator . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 1389–1399. https://doi.org/10.1109/CVPR52729.2023.00140

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. PlenOctrees for Real-time Rendering of Neural Radiance Fields . In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 5732–5741. https://doi.org/10.1109/ICCV48922.2021.00570

Junge Zhang, Feihu Zhang, Shaochen Kuang, and Li Zhang. 2024. NeRF-LiDAR: Generating Realistic LiDAR Point Clouds with Neural Radiance Fields. arXiv:2304.14811 [cs.CV] https://arxiv.org/abs/2304.14811

Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. NeRF++: Analyzing and Improving Neural Radiance Fields. arXiv:2010.07492 [cs.CV] https://arxiv.org/

abs/2010.07492

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)* 40, 6 (2021), 1–18.

Yongyi Zhao, Ankit Raghuram, Hyun K Kim, Andreas H Hielscher, Jacob T Robinson, and Ashok Veeraraghavan. 2021. High resolution, deep imaging using confocal time-of-flight diffuse optical tomography. *IEEE transactions on pattern analysis and machine intelligence* 43, 7 (2021), 2206–2219.

Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, and Changjun Jiang. 2024. LiDAR4D: Dynamic Neural Fields for Novel Space-Time View LiDAR Synthesis . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 5145–5154. https://doi.org/10.1109/CVPR52733.2024.00492