



# ParkGaussian: Surround-view 3D Gaussian Splatting for Autonomous Parking

Xiaobao Wei<sup>\*1</sup>, Zhangjie Ye<sup>\*2</sup>, Yuxiang Gu<sup>\*2</sup>, Junjie Zhu<sup>2</sup>, Yunfei Guo<sup>2</sup>, Yingying Shen<sup>1</sup>, Shan Zhao<sup>1</sup>, Ming Lu<sup>2</sup>, Haiyang Sun<sup>†1</sup>, Bing Wang<sup>1</sup>, Guang Chen<sup>1</sup>, Rongfeng Lu<sup>✉2</sup>, Hangjun Ye<sup>✉1</sup>

<sup>1</sup>Xiaomi EV <sup>2</sup>Hangzhou Dianzi University

\*Equal Contribution, <sup>†</sup>Project Leader, <sup>✉</sup> Corresponding Author

Parking is a critical task for autonomous driving systems (ADS), with unique challenges in crowded parking slots and GPS-denied environments. However, existing works focus on 2D parking slot perception, mapping, and localization, 3D reconstruction remains underexplored, which is crucial for capturing complex spatial geometry in parking scenarios. Naively improving the visual quality of reconstructed parking scenes does not directly benefit autonomous parking, as the key entry point for parking is the slots perception module. To address these limitations, we curate the first benchmark named ParkRecon3D, specifically designed for parking scene reconstruction. It includes sensor data from four surround-view fisheye cameras with calibrated extrinsics and dense parking slot annotations. We then propose ParkGaussian, the first framework that integrates 3D Gaussian Splatting (3DGS) for parking scene reconstruction. To further improve the alignment between reconstruction and downstream parking slot detection, we introduce a slot-aware reconstruction strategy that leverages existing parking perception methods to enhance the synthesis quality of slot regions. Experiments on ParkRecon3D demonstrate that ParkGaussian achieves state-of-the-art reconstruction quality and better preserves perception consistency for downstream tasks. The code and dataset will be released.

**Date:** January 4, 2026

**Project:** <https://wm-research.github.io/ParkGaussian/>

## 1 Introduction

Autonomous parking is a vital component of autonomous driving systems (ADS) (Olmos Medina et al., 2025). Unlike on-road driving, which typically occurs in structured and GPS-available environments, parking often takes place in narrow underground areas, crowded parking slots, and poor illumination (Millard-Ball, 2019). These factors pose challenges for accurate perception and localization, highlighting the necessity of specialized techniques for parking scenarios.

Early works (Huang et al., 2019; Min et al., 2021; Zinelli et al., 2019; Li et al., 2020; Wu et al., 2020; Lu et al., 2023; Yu et al., 2022) concentrate on parking perception, particularly parking slot detection, to identify and localize slots in surround-view images. These methods typically employ Inverse Perspective Mapping (IPM) to transform multi-view fisheye images into a Bird's-Eye View (BEV) representation for slot perception. Building upon these detection modules, subsequent studies (Tripathi and Yogamani, 2020; Li et al., 2024c; Qin et al., 2020) investigate Simultaneous Localization and Mapping (SLAM) in parking scenes, where parking slot landmarks serve as reliable references to enhance localization robustness in GPS-denied and visually repetitive environments. Recently, several approaches (Rathour et al., 2018; Hu et al., 2022; Yang et al., 2024; Chen et al., 2025a) explore end-to-end frameworks that jointly learn perception, planning, and control from sensor inputs, training on simulated parking datasets collected in CARLA (Dosovitskiy et al., 2017).

Although effective for simulated parking planning, there remains a large gap between CARLA and real-world parking environments. This gap highlights the need for a realistic simulator to assess slot perception and closed-loop planning performance in complex parking scenes.

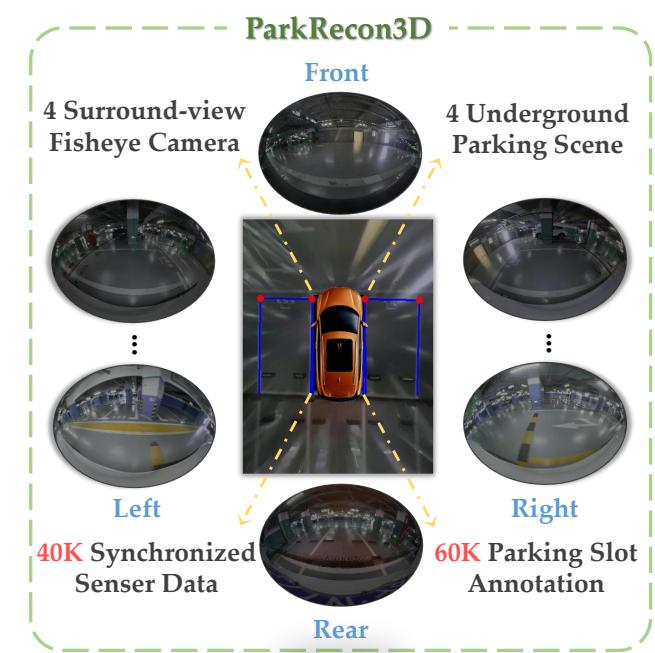
To bridge the gap between simulators and real-world environments, recent research (Dosovitskiy et al., 2017; Shah et al., 2018; Dauner et al., 2024) has increasingly focused on realistic driving simulation through 4D

scene modeling, encompassing both reconstruction and generation approaches for street-level scenarios. For reconstruction, early methods ([Turki et al., 2023](#); [Ost et al., 2021](#); [Tancik et al., 2022](#); [Tonderski et al., 2023](#); [Wu et al., 2023](#); [Yang et al., 2023](#)) build upon Neural Radiance Fields (NeRF) ([Mildenhall et al., 2020](#); [Wei et al., 2024](#)) to represent street scenes, marking the initial exploration for driving environments. But these approaches suffer from low efficiency. Recent advances ([Chen et al., 2025b](#); [Wang et al., 2025](#); [Wei et al., 2025a,b](#)) based on 3D Gaussian Splatting (3DGS) ([Kerbl et al., 2023](#); [Wu et al., 2025](#); [Liao et al., 2024](#)) have significantly improved rendering speed, which represents dynamic street scenes in a box supervised ([Yan et al., 2024](#); [Chen et al., 2024](#); [Zhou et al., 2024b,a](#); [Hess et al., 2025](#)) or a self-supervised manner ([Huang et al., 2024](#); [Li et al., 2024a](#); [Chen et al., 2023](#); [Peng et al., 2025](#); [Wei et al., 2025c](#)). For generation, recent models based on diffusion and controllable generation frameworks ([Rombach et al., 2022](#); [Gao et al., 2023](#); [Guo et al., 2025](#); [Zeng et al., 2025](#); [Li et al., 2025b,a](#)) synthesize street scenes from scene layouts or text.

These methods achieve photorealistic synthesis quality in typical on-road driving scenes. Building upon these, several works further develop reinforcement learning pipelines ([Gao et al., 2025](#)) and closed-loop simulations ([Yan et al., 2025](#); [Jiang et al., 2025](#); [Yang et al., 2025](#)) for training and evaluating autonomous driving systems ([Cao et al., 2025](#); [Liao et al., 2025](#)).

However, existing driving simulators primarily focus on modeling on-road driving, leaving parking scene simulation largely unexplored. Additionally, previous reconstruction methods mainly emphasize visual fidelity while overlooking the fundamental goal of simulation, which is to generate perception-aligned synthetic data for faithfully evaluating the performance and limitations of downstream models. For autonomous parking, where the parking slot detection module serves as the system’s entry, it is crucial to align visual fidelity in slot-related regions with the downstream perception models.

To address these limitations, we first curate a benchmark named **ParkRecon3D** (Fig. 1), specifically designed for parking-scene reconstruction. ParkRecon3D is built upon the open-sourced dataset from AVM-SLAM ([Li et al., 2024b](#)), containing sensor data captured by four calibrated fisheye cameras in underground parking lots. ParkRecon3D provides over 40K synchronized sensor frames and 60K annotated parking slots with accurate extrinsic calibration. Building on this benchmark, we propose



**Figure 1 ParkRecon3D:** a surround-view dataset for parking-aware reconstruction. It contains synchronized image streams from four fisheye cameras covering full 360° surroundings and four representative underground parking scenes. In total, the dataset includes 40K multi-camera sensor frames and 60K human-annotated parking-slot labels, supporting parking-slot perception and geometry-aware 3D reconstruction.

**ParkGaussian**, the first framework that integrates 3D Gaussian Splatting (3DGS) adapted to surround-view fisheye images to achieve high-quality 3D reconstruction of parking scenes. Furthermore, we introduce a slot-aware reconstruction strategy that leverages two widely used parking-slot detection methods (DMPR-PS ([Huang et al., 2019](#)) and GCN-Parking ([Min et al., 2021](#))) to identify slot regions in a learnable manner. This strategy not only improves synthesis quality in slot areas but also enhances alignment with downstream perception tasks, establishing a reliable simulator for autonomous parking systems.

Our main contributions are summarized as follows:

- We curate **ParkRecon3D**, the first benchmark tailored for 3D reconstruction in parking scenarios, featuring over 40K synchronized fisheye sensor frames and 60K precisely annotated parking slots with calibrated extrinsics, captured in underground parking lots.
- We present **ParkGaussian**, a novel framework that adapts 3DGS to surround-view fisheye sensors

and incorporates a slot-aware reconstruction strategy leveraging parking slot detectors to enhance reconstruction fidelity in task-critical slot regions.

- We conduct extensive experiments on ParkRecon3D, demonstrating that ParkGaussian achieves state-of-the-art reconstruction quality and stronger alignment with downstream perception.

## 2 Related Work

*Driving Scene Simulation.* Traditional autonomous driving simulators ([Shah et al., 2018](#); [Dosovitskiy et al., 2017](#)) demand substantial manual effort to construct environments and fall short of achieving photorealistic rendering. The emergence of neural field based approaches ([Ost et al., 2021](#); [Wu et al., 2023](#); [Tonderski et al., 2023](#); [Fischer et al., 2024](#)) has enabled realistic rendering and achieved a scalable neural scene representation for large-scale urban environments. Subsequently, 3DGS-based methods decompose street scenes in a box supervised ([Yan et al., 2024](#); [Chen et al., 2024](#); [Zhou et al., 2024b,a](#); [Hess et al., 2025](#)) or a self-supervised manner ([Huang et al., 2024](#); [Li et al., 2024a](#); [Chen et al., 2023](#); [Peng et al., 2025](#); [Wei et al., 2025c](#)). For example, OmniRe ([Chen et al., 2024](#)) constructs hierarchical scene representations that integrate dynamic object graphs, unifying the modeling of static backgrounds and dynamic entities such as vehicles and pedestrians to achieve high-fidelity reconstruction of complex urban driving environments.

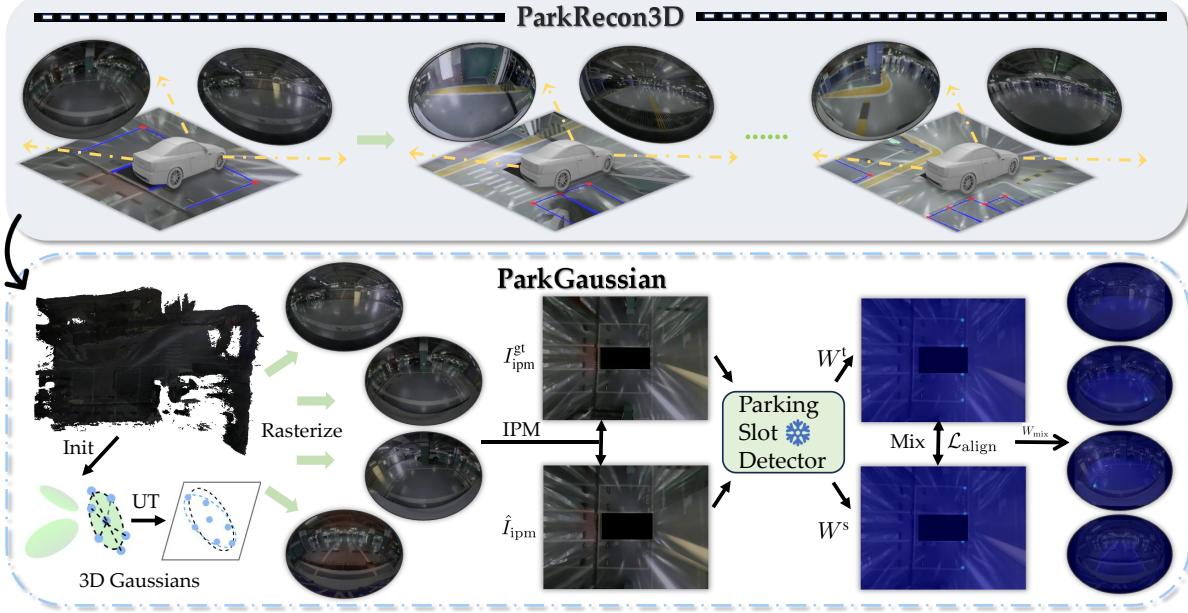
However, these methods primarily focus on driving scenes and are not suitable for underground parking environments. To achieve promising synthesis quality, they heavily rely on accurate multi-sensor setups, including dense LiDAR point clouds and calibrated GPS/IMU localization. In contrast, underground parking scenes present a much harsher setting, characterized by poor illumination, limited texture diversity and the absence of GPS signals required for accurate extrinsic calibration. These factors pose significant challenges for parking scene reconstruction.

*Autonomous Parking.* Autonomous parking serves as a foundation of ADS. Compared to driving, it presents more challenges in perception and localization, as parking lots are characterized as narrow and GPS-denied. Previous works in parking slot detection ([Huang et al., 2019](#); [Min et al., 2021](#)) simplify this task by translating surround-view images to BEV representation, which inevitably limits the ability to perceive the entire space. AVP-SLAM ([Qin et al., 2020](#)) and AVM-SLAM ([Li et al., 2024b](#)) provide a method for automatically navigating into parking lots and parking in the target spot by extracting semantic features from stable parking slots, highlighting the importance of parking slot detection. Beyond SLAM, there has been a growing trend toward end-to-end models ([Rathour et al., 2018](#); [Hu et al., 2022](#); [Yang et al., 2024](#); [Chen et al., 2025a](#)) that integrate perception, planning, and control within a unified learning framework for autonomous parking in the simulation environment ([Dosovitskiy et al., 2017](#)). Despite the appeal of fully end-to-end pipelines, the entry point of autonomous parking remains perception, and the quality of perception directly determines downstream planning accuracy. In particular, parking slot detection relies on recognizing fine-grained, geometry-driven patterns such as L-shaped and T-shaped corners, making local geometric consistency crucial. While previous studies have advanced parking specialized methods, research dedicated to 3D parking reconstruction remains scarce.

To bridge this gap, we introduce ParkRecon3D, the first benchmark specifically tailored for 3D reconstruction in parking scenarios, and ParkGaussian, a novel 3D Gaussian Splatting framework adapted to surround-view fisheye inputs with a slot-aware reconstruction strategy.

## 3 Method

In this section, we first describe the construction of the ParkRecon3D benchmark. We then review the preliminaries of 3D Gaussian Splatting (3DGS) and its extension with 3DGUT. Finally, we present the IPM projection and the slot-aware reconstruction modules that form the core of ParkGaussian.



**Figure 2** Overview of the ParkGaussian pipeline. Four surround-view fisheye images are first rendered from 3D Gaussian primitives using UT-based projection for stable splatting under strong distortion. The rendered views are then passed through a differentiable IPM module to obtain a unified BEV map. A pretrained parking-slot detector (DMPR-PS or GCN-Parking) processes both rendered and ground-truth BEV maps to produce teacher–student structural guidance, from which slot-aware weights are constructed. These weights supervise reconstruction in both the IPM space and the camera-view space, forming the alignment and slot-aware objectives. ParkGaussian jointly optimizes Gaussian attributes toward photometric fidelity and perception-aligned slot geometry.

### 3.1 ParkRecon3D Benchmark

To the best of our knowledge, there is currently no benchmark specifically designed for 3D reconstruction in parking environments. We therefore construct ParkRecon3D by reorganizing and extending the open-source dataset used for AVM-SLAM (Li et al., 2024b). The data are collected in an underground parking facility of approximately 220 m × 110 m with more than 430 parking spaces. The ego vehicle is equipped with a production around-view monitoring system consisting of four fisheye cameras mounted at the front, rear, left, and right sides of the vehicle. Each camera records images at 10 Hz with a resolution of 1280 × 960, and the system additionally produces synthesized Inverse Perspective Map (IPM) images at 1354 × 1632 from the four-view fisheye inputs. The sequences are organized into four representative scenes covering diverse parking layouts.

To provide an accurate geometric reference for reconstruction, we calibrate the extrinsic parameters of all four fisheye cameras using COLMAP (Schönberger and Frahm, 2016; Schönberger et al., 2016), instead of relying on wheel-encoder and IMU odometry that is noisy in underground environments. For parking-slot perception, we follow the annotation protocol of DMPR-PS (Huang et al., 2019) and manually label slot corner points in the BEV domain, yielding high-quality supervision for slot-level detection. Building on this data, we curate ParkRecon3D, the first benchmark tailored for 3D reconstruction in parking scenarios. It contains over 40K synchronized multi-fisheye frames and 60K manually verified parking-slot annotations with calibrated extrinsics. The benchmark will be released to the community as a comprehensive dataset for training and evaluating 3D reconstruction models in underground parking environments.

### 3.2 Preliminary

*3D Gaussian Splatting.* 3DGS (Kerbl et al., 2023) is an advanced method for representing 3D scenes, using a set of 3D Gaussian spheres to represent the scene. Its probability density function is formulated as:

$$G(x) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \quad (1)$$

where  $\mu \in \mathbb{R}^3$  is the mean value, which is the coordinates of the 3D Gaussian sphere in the world;  $\Sigma$  represents the covariance matrix, which can be decomposed into  $\Sigma = RSS^T R^T$ , where  $R \in SO(3)$  is the rotation matrix and  $T \in \mathbb{R}^{3 \times 3}$  is the scaling matrix, which is stored by the rotation quaternion  $q \in \mathbb{R}^4$ . For the colors in the pixels, the spherical harmonics (SH) and the opacity  $\alpha$  are assigned to the color of the Gaussian sphere, and the depth sorting is carried out according to the distance between the Gaussian sphere and the camera, and then the color of a pixel can be calculated by using  $\alpha$ -blended:

$$\mathbf{c} = \sum_{i=1}^N \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (2)$$

*3DGS with UT-based projection.* Vanilla 3DGS relies on an EWA splatting formulation that linearizes the camera projection using a first-order Jacobian. This becomes inaccurate for heavily distorted fisheye cameras and requires camera-specific Jacobian derivations, which is undesirable for multi-camera surround-view systems. To address this issue, we incorporate the unscented-transform (UT) projection used in 3DGUT (Wu et al., 2025) as a drop-in replacement within 3DGS.

Instead of linearizing the non-linear projection  $v = g(x)$ , the UT approximates the 3D Gaussian  $\mathcal{N}(\mu, \Sigma)$  using a small set of sigma points:

$$x_i = \begin{cases} \mu, & i = 0, \\ \mu \pm \sqrt{(n + \lambda)\Sigma}_{[i]}, & i = 1, \dots, 2n, \end{cases} \quad (3)$$

which are projected exactly through  $g(\cdot)$ . The mean and covariance of the resulting 2D Gaussian footprint are then computed as:

$$\begin{aligned} \mu_v &= \sum_i w_i^\mu g(x_i), \\ \Sigma_v &= \sum_i w_i^\Sigma (g(x_i) - \mu_v)(g(x_i) - \mu_v)^\top \end{aligned} \quad (4)$$

This UT-based projection removes the need for Jacobian derivation for each fisheye camera model and yields more stable 2D Gaussian footprints under strong distortion. In ParkGaussian, this module allows 3DGS to train directly on surround-view fisheye imagery and significantly improves geometric stability in underground parking scenarios.

### 3.3 Differentiable Surround-View IPM

Most parking slot detection models operate on bird's-eye-view (BEV) images rather than raw fisheye inputs. Therefore, directly rendering reconstructed fisheye views from 3DGS cannot be used by downstream detectors. Thus, we convert the rendered surround-view fisheye images into a unified BEV representation through a fully differentiable Inverse Perspective Mapping (IPM) module.

Following AVP-SLAM (Qin et al., 2020), each surround-view camera is calibrated offline with intrinsic matrix  $K_c$  and extrinsic parameters  $(R_c, t_c)$ . A fisheye pixel  $\mathbf{u} = [u, v, 1]^\top$  is first unprojected to a ray in the camera frame using the inverse fisheye model  $\pi_c^{-1}(\cdot)$ , and then intersected with the ground plane ( $z = 0$ ) in the vehicle frame:

$$\frac{1}{\lambda} \begin{bmatrix} x_v \\ y_v \\ 1 \end{bmatrix} = [R_c \ t_c]_{\text{col}:1,2,4}^{-1} \pi_c^{-1} \left( \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \right) \quad (5)$$

where  $[x_v, y_v]^\top$  is the ground-plane point in the vehicle coordinate system, and  $\lambda$  is the scale factor to enforce  $z = 0$ .

Then ground-plane points from all four fisheye cameras are fused onto a canonical BEV plane. Given the intrinsic matrix  $K_{\text{ipm}}$  of the synthesized IPM image, each point is reprojected to BEV pixel coordinates:

$$\begin{bmatrix} u_{\text{ipm}} \\ v_{\text{ipm}} \\ 1 \end{bmatrix} = K_{\text{ipm}} \begin{bmatrix} x_v \\ y_v \\ 1 \end{bmatrix} \quad (6)$$

producing a unified BEV map  $\hat{I}_{\text{ipm}}$  consistent with the geometric assumptions of existing parking-slot detectors. All steps in the IPM are expressed in closed-form and implemented as differentiable operations. This enables gradients from the downstream detection models to be propagated back to the 3D Gaussian representation, allowing the reconstruction to be directly optimized toward downstream parking slot perception objectives.

### 3.4 Slot-aware Reconstruction

Based on the differentiable IPM mapping, we further introduce Slot-Aware Reconstruction, which injects task-driven supervisory signals from pretrained parking slot detectors directly into the 3DGS optimization process. This encourages Gaussian primitives to encode slot-critical geometry and results in reconstructions that are not only photometrically accurate but also structurally faithful for downstream perception.

After rendering the surround-view fisheye images  $\hat{I}_{\text{sur}}$  from 3DGS, the differentiable IPM operator  $\Phi_{\text{IPM}}$  produces a synthesized IPM image

$$\hat{I}_{\text{ipm}} = \Phi_{\text{IPM}}(\hat{I}_{\text{sur}}) \quad (7)$$

In parallel, we construct a ground-truth IPM image  $I_{\text{ipm}}^{\text{gt}}$  from the annotated ParkRecon3D data using the same IPM mapping. A pretrained parking-slot detector is then applied to both IPM images to extract task-relevant structural features. In this paper, we adopt two widely used methods, DMPR-PS (Huang et al., 2019) and GCN-Parking (Min et al., 2021), finetuned on ParkRecon3D. During reconstruction optimization, the perception network is kept frozen, providing stable, non-drifting guidance.

#### 3.4.1 Parking Corner Guidance

Taking DMPR-PS as an example, we feed the ground-truth and rendered IPM images through the same network:

$$H^t = f_{\text{DMPR-PS}}(I_{\text{ipm}}^{\text{gt}}), \quad H^s = f_{\text{DMPR-PS}}(\hat{I}_{\text{ipm}}) \quad (8)$$

where  $H^t$  and  $H^s$  denote the teacher and student feature maps, respectively, each containing corner confidence, direction fields, and offset predictions. Since our goal is to make the structural features extracted from the rendered IPM  $\hat{I}_{\text{ipm}}$  approach those extracted from the ground-truth IPM  $I_{\text{ipm}}^{\text{gt}}$ , we treat  $H^t$  as the teacher signal and  $H^s$  as the student representation. The corner-confidence channels of  $H^t$  and  $H^s$  are then used to construct slot-aware weight maps, guiding the reconstruction toward slot-critical geometric structures.

Let  $H_{\text{conf}}^t$  and  $H_{\text{conf}}^s$  denote the slot corner confidence channels of  $H^t$  and  $H^s$ , respectively. We convert these confidence maps into continuous slot-aware weights by applying a differentiable shaping function:

$$W = \sigma\left(\frac{H_{\text{conf}} - \tau}{T}\right)^{\gamma} \quad (9)$$

where  $T = 0.5$  is a temperature parameter controlling the softness of the mask,  $\tau = 0.25$  is a confidence threshold, and  $\gamma = 1$  produces a linear shaping. These choices follow our implementation and empirically balance sharp localization with stable gradients. This transformation yields soft masks  $W^t$  and  $W^s$  that concentrate supervision around high-confidence slot corners.

Based on the soft masks  $W^t$  and  $W^s$ , we construct a mixed slot-aware weight map that balances reliable geometric supervision from ground truth detections with adaptive cues from the rendered predictions:

$$W_{\text{mix}} = \alpha W^t + (1 - \alpha) \text{sg}(W^s), \quad \alpha = 0.8 \quad (10)$$

This formulation serves two purposes. First, the stop-gradient operator  $\text{sg}(\cdot)$  prevents the student weights from being directly updated, avoiding degenerate solutions such as uniformly low confidence and ensuring that gradients flow only through the reconstruction pathway. Second, the mixture combines the stability of teacher-derived supervision with the adaptiveness of student predictions, allowing the model to focus on slot-critical geometric regions that may vary across frames.

The resulting mixed weights  $W_{\text{mix}}$  are upsampled to the resolution of the rendered IPM image and subsequently back-projected into each surround-view fisheye camera. This ensures that slot-aware supervision is consistently applied both in the IPM space and the raw fisheye camera space.

### 3.4.2 Slot Edge-aware Extension

GCN-Parking (Min et al., 2021) extends DMPR-PS by additionally predicting slot edges between detected corners. Let  $A^t$  and  $A^s$  denote the teacher and student edge-score matrices obtained from the ground-truth and rendered IPM images, respectively. For each image, we select the top- $K_e$  edge candidates ( $K_e = 8$  in our experiments) and, for every edge  $(i, j)$ , define a line segment between the corresponding corner locations  $p_i$  and  $p_j$ :

$$\ell_{ij}(t) = (1 - t)p_i + t p_j, \quad t \in [0, 1] \quad (11)$$

We then rasterize each edge onto the IPM grid using a Gaussian tube with width  $\sigma = 1.5$  and  $N_s = 32$  samples:

$$e_{ij}(u, v) = \max_{t \in [0, 1]} \exp\left(-\frac{\|(u, v) - \ell_{ij}(t)\|_2^2}{2\sigma^2}\right) \quad (12)$$

Aggregating all selected edges  $e_{ij}(u, v)$  yields teacher and student edge weight maps  $W_{\text{edge}}^t$  and  $W_{\text{edge}}^s$ , which are mixed analogously to the corner masks,

$$W_{\text{edge}} = \beta W_{\text{edge}}^t + (1 - \beta) \text{sg}(W_{\text{edge}}^s), \quad \beta = 0.8 \quad (13)$$

and combined with the DMPR-PS-based mask to form the final slot-aware weights:

$$W'_{\text{mix}} = W_{\text{mix}} + \lambda_{\text{edge}} W_{\text{edge}} \quad (14)$$

which emphasizes both corner points and slot boundaries during reconstruction.

## 3.5 Training

During optimization, ParkGaussian is first supervised by photometric rendering loss in vanilla 3DGS for 20,000 iterations, then optimized with alignment loss and slot aware loss for 10,000 iterations.

*Photometric Rendering Loss.* Given rendered fisheye views  $\hat{I}_{\text{sur}}$  and ground-truth images  $I_{\text{sur}}$ , we use the standard 3DGS photometric objective:

$$\mathcal{L}_{\text{rgb}} = (1 - \lambda) \|\hat{I}_{\text{sur}} - I_{\text{sur}}\|_1 + \lambda \mathcal{L}_{\text{D-SSIM}} \quad (15)$$

where  $\lambda = 0.2$ .

*Alignment Loss.* Rendered IPM maps and ground-truth IPM maps are passed through the slot detection model to produce student and teacher confidence fields ( $W^s, W^t$ ). We further regularize the teacher and student weights to have similar spatial distributions on the teacher’s top- $K$  region. Let  $\Omega = \text{TopK}(W^t)$  and define the normalized distributions

$$\pi^s = \text{softmax}(W^s | \Omega), \quad \pi^t = \text{softmax}(W^t | \Omega)$$

The alignment loss is then calculated by:

$$\mathcal{L}_{\text{align}} = \text{KL}(\pi^s \| \pi^t) \quad (16)$$

*Slot-aware reconstruction loss.* The mixed structural weights  $W_{\text{mix}}$  supervise reconstruction both in the IPM domain and in the camera domain. First, weighted IPM consistency encourages the rendered IPM map  $\hat{I}_{\text{ipm}}$  to match the ground-truth  $I_{\text{ipm}}^{\text{gt}}$ :

$$\mathcal{L}_{\text{ipm}} = \frac{\|W_{\text{mix}} \odot |\hat{I}_{\text{ipm}} - I_{\text{ipm}}^{\text{gt}}|\|_1}{\|W_{\text{mix}}\|_1} \quad (17)$$

The same weights are then back-projected to each camera and modulate RGB reconstruction errors:

$$\mathcal{L}_{\text{cam}} = \frac{\|W_{\text{mix}} \odot |\hat{I}_{\text{sur}} - I_{\text{sur}}|\|_1}{\|W_{\text{mix}}\|_1} \quad (18)$$

Together, these two terms form the slot-aware reconstruction objective that emphasizes geometry aligned with parking-slot structures.

*Final Loss.* The total loss used during ParkGaussian optimization is:

$$\mathcal{L} = \mathcal{L}_{\text{rgb}} + \lambda_{\text{align}} \mathcal{L}_{\text{align}} + \lambda_{\text{ipm}} \mathcal{L}_{\text{ipm}} + \lambda_{\text{cam}} \mathcal{L}_{\text{cam}} \quad (19)$$

where  $\lambda_{\text{align}} = 0.001$ ,  $\lambda_{\text{ipm}} = 0.1$ ,  $\lambda_{\text{cam}} = 0.1$ .

## 4 Experiments

### 4.1 Experimental Setup

*Baselines.* For reconstruction, we restrict comparison to a method explicitly validated for parking environments. Following previous protocols, we adopt OmniRe (Chen et al., 2024), 3DGUT (Wu et al., 2025), and Self-Cali-GS (Deng et al., 2025) that mainly focus on fisheye camera environments as a reconstruction baseline. For parking slot detection, we select the surround-view systems method (Huang et al., 2019; Min et al., 2021) as a baseline, which differs from monocular camera-based methods. Furthermore, we fine-tune them to better align with our benchmark. During the training, we use four scenes of our benchmark, from which 100 frames of four-channel surround fisheye inputs are sampled for each scene. Evaluation is performed every 10 frames to compute the relevant metrics. For the perception models, we use the corresponding frames' ground-truth IPM images to train the two detection networks. **Due to the space limitation, please refer to the appendix for more experiments.**

*Evaluation Metrics.* We test the underground garage scene on Self-Cali-GS (Deng et al., 2025), OmniRe (Chen et al., 2024), 3DGUT (Wu et al., 2025) and our method, and compute the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) metrics for the entire image to evaluate the impact of the underground garage's challenged conditions on 3D reconstruction. For the downstream task of parking slot detection, we use DMPR-PS and GCN Parking Slot methods to calculate Precision and Recall to evaluate the detection metrics of the reconstructed scene from a synthetic BEV perspective. The parking point Precision and Recall are defined as:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (20)$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (21)$$

Suppose the marked real parking space entry points are  $P_{r_1}(x_{r_1}, y_{r_1})$ ,  $P_{r_2}(x_{r_2}, y_{r_2})$ , and the identified parking space entry points are  $P_{i_1}(x_{i_1}, y_{i_1})$ ,  $P_{i_2}(x_{i_2}, y_{i_2})$ . If the relationship between two points satisfies:

$$\|(p_{r_1} - p_{i_1}, p_{r_2} - p_{i_2})\|_2 < \text{distance} \quad (22)$$

And the confidence meets its threshold, then it will be recognized as a true positive; otherwise, it is a false positive. If none of the above conditions are met, it will be recognized as a false negative. Where the *distance* can be adjusted. On this basis, the parking bitmap angle constructed from the parking entrance point is introduced, and the Precision and the Recall of the predicted parking space can be calculated by judging whether it meets:  $|\theta_r - \theta_i| < \text{angle}$  or  $360^\circ - |\theta_r - \theta_i| < \text{angle}$ , where the *angle* can be adjusted.

*Implementation Details.* Our ParkGaussian framework is implemented in PyTorch and trained for 30,000 iterations on a single NVIDIA RTX 4090 GPU using the Adam optimizer. For 3D Gaussian attributes, we adopt the learning-rate schedule used in 3DGS (Kerbl et al., 2023) and 3DGUT (Wu et al., 2025), where the position learning rate follows an exponential decay to 1% of its initial value. We also employ the Markov Chain Monte Carlo optimization strategy (Kheradmand et al., 2024) implemented in GSplat (Ye et al., 2025) to improve convergence and reconstruction stability. The Gaussian primitives are initialized from the COLMAP sparse point cloud in ParkRecon3D. All remaining hyperparameters follow those described in the method section.

**Table 1** Novel view synthesis performance on the proposed ParkRecon3D dataset. **Bold** denotes the best result and underline denotes the second best.

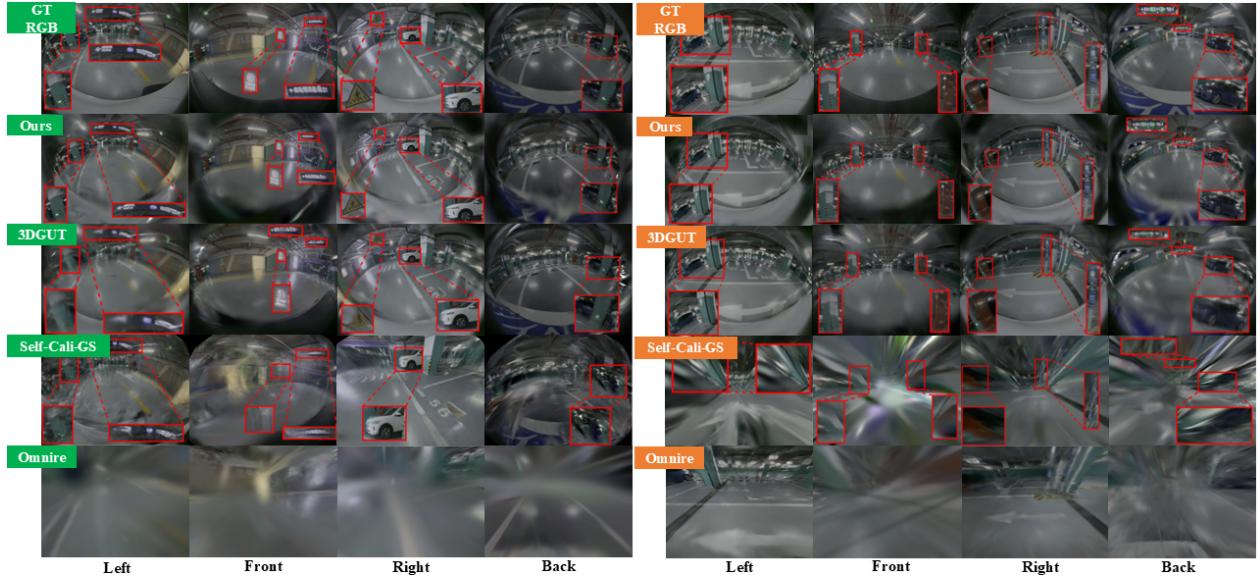
Scene	Modules	PSNR↑	SSIM↑	LPIPS↓
Scene1	Self-Cali-GS (Deng et al., 2025)	23.78	0.82	0.31
	3DGUT (Wu et al., 2025)	28.70	<u>0.92</u>	<u>0.21</u>
	OmniRe (Chen et al., 2024)	25.12	0.84	0.37
	Ours (w/ DMPR-PS)	<u>29.10</u>	<u>0.92</u>	<b>0.20</b>
	Ours (w/ GCN-Parking)	<b>30.09</b>	<b>0.93</b>	<b>0.20</b>
Scene3	Self-Cali-GS (Deng et al., 2025)	20.10	0.81	0.30
	3DGUT (Wu et al., 2025)	27.80	<u>0.92</u>	<b>0.20</b>
	OmniRe (Chen et al., 2024)	21.58	0.78	0.50
	Ours (w/ DMPR-PS)	28.89	<u>0.92</u>	0.21
	Ours (w/ GCN-Parking)	<b>30.27</b>	<b>0.93</b>	<b>0.20</b>

## 4.2 Main Results

### 4.2.1 Novel View Synthesis

Tab. 1 shows the results on the proposed ParkRecon3D dataset. The results show that ParkingGaussian achieves state-of-the-art performance in the Benchmark. The previous method has a poor performance because they were merely emphasizes visual fidelity while overlooking the fundamental goal of simulation. For a more intuitive understanding of the model’s reconstruction performance, we present the side-by-side visual comparison between the previous work and ours in Fig. 3, which evident that under the inherent constraints of underground parking garages, although 3DGUT (Wu et al., 2025) and Self-Cali-GS (Deng et al., 2025) can construct the spatial topology of the overall scene, they have obvious limitations in the robustness of detail representation. OmniRe (Chen et al., 2024) exhibits significantly inferior reconstruction quality, with severe blurring and structural loss across all perspectives.

Compared with all these methods, our approach achieves high-level visual alignment with the ground-truth scenes from all perspectives and enables more accurate scene reconstruction by incorporating a slot-aware reconstruction strategy that leverages parking-slot detectors to enhance reconstruction fidelity in task-critical slot regions.



**Figure 3** Novel view synthesis visualization in 4 surround view fisheye images.

**Table 2** DMPR-PS model comparison: results on improved AVM-SLAM dataset for parking slot point detection.

Variant	Scene1		Scene3	
	Precision↑	Recall↑	Precision↑	Recall↑
GT-DMPR-PS (Huang et al., 2019)	0.86	0.22	0.49	0.21
Ours w/o Per DMPR-PS	0.71	0.08	0.48	0.18
Ours w/ Per DMPR-PS	0.74	0.10	0.47	0.19

#### 4.2.2 Parking Slot Detection

Tab. 2, 3 demonstrate that our slot-aware reconstruction strategy substantially enhances the performance of parking slot detection. Both perception networks achieve performance close to the ground truth when operating on our reconstruction model, and both show noticeable improvements after incorporating the perception module during reconstruction. In Fig 4, we provide a visual comparison between the detection results on real images and those obtained from the renderings of our reconstruction model. As can be visually observed, our approach maintains high-fidelity reconstruction of the global scene and, at the same time, accurately reproduces the fine-grained structures of the parking slots.

#### 4.3 Ablation on Slot-aware Strategy

We conduct comprehensive ablation studies to disentangle the contribution of each component in our slot-aware reconstruction framework. Five variants are evaluated: (1) Direct IPM  $L_1$  supervision without slot-aware cues; (2) Feature-level supervision using perception outputs only; (3) Teacher-only weighting using ground-truth IPM features; (4) Student-only weighting using rendered IPM features; and (5) The full slot-aware design that mixes teacher–student weights with distribution alignment.

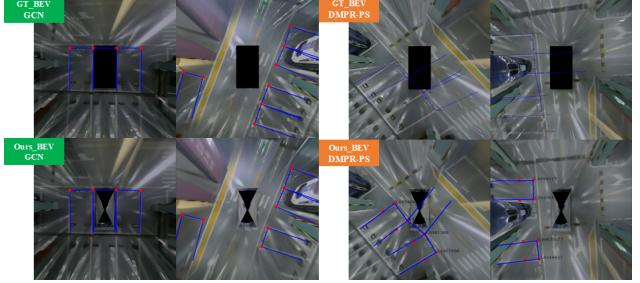
**Table 4** Ablation study of the proposed slot-aware reconstruction strategy on ParkRecon3D. Metrics include rendering quality and downstream parking-slot detection.

Variant	PSNR↑	SSIM↑	LPIPS↓	Prec.↑	Rec.↑
<i>Scene1</i>					
Direct IPM $L_1$ supervision	24.94	0.88	0.33	0.62	0.06
Feature-level supervision only	27.43	0.91	0.28	0.64	0.05
Teacher-only weighting	29.56	0.92	0.21	0.90	0.41
Student-only weighting	28.62	0.90	0.22	0.81	0.40
<b>Full slot-aware (ours)</b>	<b>30.09</b>	<b>0.93</b>	<b>0.20</b>	<b>0.97</b>	<b>0.43</b>
<i>Scene3</i>					
Direct IPM $L_1$ supervision	24.99	0.89	0.35	0.58	0.04
Feature-level supervision only	27.36	0.91	0.27	0.43	0.11
Teacher-only weighting	28.77	0.93	0.21	0.88	0.45
Student-only weighting	26.72	0.90	0.29	0.71	0.41
<b>Full slot-aware (ours)</b>	<b>30.27</b>	<b>0.93</b>	<b>0.20</b>	<b>0.95</b>	<b>0.48</b>

The ablation results in Tab. 4 reveal several important insights into the effectiveness of the proposed slot-aware strategy. First, naively applying an IPM-space  $L_1$  loss yields poor reconstruction quality and extremely weak parking-slot detection performance. This is largely because multi-view projections introduce conflicts at view boundaries, injecting additional noise into the IPM domain. Feature-level supervision improves rendering scores but still fails to recover reliable slot geometry, indicating that the optimization objectives of the perception model and the reconstruction model are not aligned. Their feature distributions differ substantially, causing the reconstructed geometry to deviate from the structures required by the perception network. Both teacher-only and student-only weighting substantially boost performance by focusing supervision on slot-relevant regions. However, they exhibit complementary behaviours: teacher-only weighting yields stable but less adaptive

**Table 3** GCN model comparison: results on improved AVM-SLAM dataset for parking slot detection.

Variant	Scene1		Scene3	
	Precision↑	Recall↑	Precision↑	Recall↑
GT-GCN (Min et al., 2021)	0.99	0.49	0.98	0.50
Ours w/o Per GCN	0.95	0.40	0.94	0.48
Ours w/ Per GCN	0.97	0.43	0.95	0.48

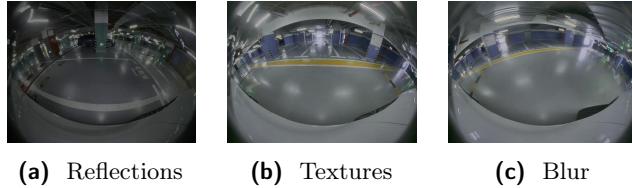


**Figure 4** Parking slot detection visualization in IPM images.

supervision, while student-only weighting adapts to rendered predictions but is more susceptible to noise. Our full slot-aware strategy, which blends teacher–student weighting and imposes distribution alignment, consistently achieves the best results across both scenes. It achieves the highest rendering quality and, by aligning reconstruction with the perception model, significantly improves downstream precision and recall, showing that integrating structural priors with prediction consistency is essential for robust parking-aware reconstruction.

## 5 Limitation

ParkRecon3D still contains several inherent challenges of underground parking environments, including indoor specular reflections, highly repetitive textures, and motion blur caused by long exposure under weak lighting (Fig. 5). These characteristics are difficult to model accurately and will be further addressed in future work.



**Figure 5** Challenges in underground parking lots.

## 6 Conclusion

In this work, we present ParkRecon3D, a novel framework that integrates 3D Gaussian Splatting with a slot-aware reconstruction strategy for high-quality parking scene modeling. Built upon the proposed ParkRecon3D benchmark, our method achieves accurate and efficient 3D reconstruction in underground parking environments while enhancing synthesis quality in parking slot regions. ParkRecon3D provides a reliable foundation for the development and evaluation of future autonomous parking systems.

## References

- Jiajun Cao, Qizhe Zhang, Peidong Jia, Xuhui Zhao, Bo Lan, Xiaoan Zhang, Zhuo Li, Xiaobao Wei, Sixiang Chen, Liyun Li, et al. Fastdrivevla: Efficient end-to-end driving via plug-and-play reconstruction-based token pruning. *arXiv preprint arXiv:2507.23318*, 2025.
- Chao Chen, Shunyu Yao, Yuanwu He, Tao Feng, Ruojing Song, Yuliang Guo, Xinyu Huang, Chenxu Wu, Ren Liu, and Chen Feng. End-to-end visual autonomous parking via control-aided attention. *arXiv preprint arXiv:2509.11090*, 2025a.
- Peng Chen, Xiaobao Wei, Qingpo Wuwu, Xinyi Wang, Xingyu Xiao, and Ming Lu. Mixedgaussianavatar: Realistically and geometrically accurate head avatar via mixed 2d-3d gaussians. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 945–954, 2025b.
- Yurui Chen, Chun Gu, Junzhe Jiang, Xiatian Zhu, and Li Zhang. Periodic vibration gaussian: Dynamic urban scene reconstruction and real-time rendering. *arXiv:2311.18561*, 2023.
- Ziyu Chen, Jiawei Yang, Jiahui Huang, Riccardo de Lutio, Janick Martinez Esturo, Boris Ivanovic, Or Litany, Zan Gojcic, Sanja Fidler, Marco Pavone, et al. Omnidire: Omni urban scene reconstruction. *arXiv preprint arXiv:2408.16760*, 2024.
- Daniel Dauner, Marcel Hallgarten, Tianyu Li, Xinshuo Weng, Zhiyu Huang, Zetong Yang, Hongyang Li, Igor Gilitschenski, Boris Ivanovic, Marco Pavone, et al. Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking. *Advances in Neural Information Processing Systems*, 37:28706–28719, 2024.
- Youming Deng, Wenqi Xian, Guandao Yang, Leonidas Guibas, Gordon Wetzstein, Steve Marschner, and Paul Debevec. Self-calibrating gaussian splatting for large field of view reconstruction. *arXiv preprint arXiv:2502.09563*, 2025.
- Alexey Dosovitskiy, German Ros, Felipe Codella, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017.

- Tobias Fischer, Lorenzo Porzi, Samuel Rota Bulo, Marc Pollefeys, and Peter Kontschieder. Multi-level neural scene graphs for dynamic urban environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21125–21135, 2024.
- Hao Gao, Shaoyu Chen, Bo Jiang, Bencheng Liao, Yiang Shi, Xiaoyang Guo, Yuechuan Pu, Haoran Yin, Xiangyu Li, Xinbang Zhang, et al. Rad: Training an end-to-end driving policy via large-scale 3dgs-based reinforcement learning. *arXiv preprint arXiv:2502.13144*, 2025.
- Ruiyuan Gao, Kai Chen, Enze Xie, Lanqing Hong, Zhenguo Li, Dit-Yan Yeung, and Qiang Xu. Magicdrive: Street view generation with diverse 3d geometry control. *arXiv preprint arXiv:2310.02601*, 2023.
- Xiangyu Guo, Zhanqian Wu, Kaixin Xiong, Ziyang Xu, Lijun Zhou, Gangwei Xu, Shaoqing Xu, Haiyang Sun, Bing Wang, Guang Chen, et al. Genesis: Multimodal driving scene generation with spatio-temporal and cross-modal consistency. *arXiv preprint arXiv:2506.07497*, 2025.
- Georg Hess, Carl Lindström, Maryam Fatemi, Christoffer Petersson, and Lennart Svensson. Splatad: Real-time lidar and camera rendering with 3d gaussian splatting for autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 11982–11992, 2025.
- Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision*, pages 533–549. Springer, 2022.
- Junhao Huang, Lin Zhang, Ying Shen, Huijuan Zhang, Shengjie Zhao, and Yukai Yang. Dmpr-ps: A novel approach for parking-slot detection using directional marking-point regression. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 212–217. IEEE, 2019.
- Nan Huang, Xiaobao Wei, Wenzhao Zheng, Pengju An, Ming Lu, Wei Zhan, Masayoshi Tomizuka, Kurt Keutzer, and Shanghang Zhang. S3gaussian: Self-supervised street gaussians for autonomous driving. *arXiv preprint arXiv:2405.20323*, 2024.
- Junzhe Jiang, Nan Song, Jingyu Li, Xiatian Zhu, and Li Zhang. Realengine: Simulating autonomous driving in realistic context. *arXiv preprint arXiv:2505.16902*, 2025.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering, 2023.
- Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems*, 37:80965–80986, 2024.
- Hao Li, Jingfeng Li, Dingwen Zhang, Chenming Wu, Jieqi Shi, Chen Zhao, Haocheng Feng, Errui Ding, Jingdong Wang, and Junwei Han. Vdg: Vision-only dynamic gaussian for driving simulation. *arXiv preprint arXiv:2406.18198*, 2024a.
- Wei Li, Libo Cao, Lingbo Yan, Chaohui Li, Xiebing Feng, and Peijie Zhao. Vacant parking slot detection in the around view image based on deep learning. *Sensors*, 20(7):2138, 2020.
- Ye Li, Wenchao Yang, Dekun Lin, Qianlei Wang, Zhe Cui, and Xiaolin Qin. Avm-slam: Semantic visual slam with multi-sensor fusion in a bird’s eye view for automated valet parking. *arXiv preprint arXiv:2309.08180*, 2024b.
- Ye Li, Wenchao Yang, Dekun Lin, Qianlei Wang, Zhe Cui, and Xiaolin Qin. Avm-slam: Semantic visual slam with multi-sensor fusion in a bird’s eye view for automated valet parking. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7937–7943. IEEE, 2024c.
- Ying Li, Xiaobao Wei, Xiaowei Chi, Yuming Li, Zhongyu Zhao, Hao Wang, Ningning Ma, Ming Lu, and Shanghang Zhang. Manipdreamer: Boosting robotic manipulation world model with action tree and visual guidance. *arXiv preprint arXiv:2504.16464*, 2025a.
- Ying Li, Xiaobao Wei, Xiaowei Chi, Yuming Li, Zhongyu Zhao, Hao Wang, Ningning Ma, Ming Lu, and Shanghang Zhang. Manipdreamer3d: Synthesizing plausible robotic manipulation video with occupancy-aware 3d trajectory. *arXiv preprint arXiv:2509.05314*, 2025b.
- Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, et al. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12037–12047, 2025.

- Zimu Liao, Siyan Chen, Rong Fu, Yi Wang, Zhongling Su, Hao Luo, Linning Xu, Bo Dai, Hengjie Li, and Zhilin Pei. Fisheye-gs: Lightweight and extensible gaussian splatting module for fisheye cameras. *arXiv preprint arXiv:2409.04751*, 2024.
- Yuheng Lu, Chenfeng Xu, Xiaobao Wei, Xiaodong Xie, Masayoshi Tomizuka, Kurt Keutzer, and Shanghang Zhang. Open-vocabulary point-cloud object detection without 3d annotation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1190–1199, 2023.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis, 2020.
- Adam Millard-Ball. The autonomous vehicle parking problem. *Transport Policy*, 75:99–108, 2019.
- Chen Min, Jiaolong Xu, Liang Xiao, Dawei Zhao, Yiming Nie, and Bin Dai. Attentional graph neural network for parking-slot detection. *IEEE Robotics and Automation Letters*, 6(2):3445–3450, 2021.
- Javier Santiago Olmos Medina, Jessica Gissella Maradey Lázaro, Anton Rassölklin, and Hernán González Acuña. An overview of autonomous parking systems: Strategies, challenges, and future directions. *Sensors*, 25(14):4328, 2025.
- Julian Ost, Fahim Mannan, Nils Thuerey, Julian Knodt, and Felix Heide. Neural scene graphs for dynamic scenes, 2021.
- Chensheng Peng, Chengwei Zhang, Yixiao Wang, Chenfeng Xu, Yichen Xie, Wenzhao Zheng, Kurt Keutzer, Masayoshi Tomizuka, and Wei Zhan. Desire-gs: 4d street gaussians for static-dynamic decomposition and surface reconstruction for urban driving scenes. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6782–6791, 2025.
- Tong Qin, Tongqing Chen, Yilun Chen, and Qing Su. Avp-slam: Semantic visual mapping and localization for autonomous vehicles in the parking lot. *arXiv preprint arXiv:2007.01813*, 2020.
- S Rathour, Vijay John, MK Nithilan, and Seiichi Mita. Vision and dead reckoning-based end-to-end parking for autonomous vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 2182–2187. IEEE, 2018.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.
- Shital Shah, Debadeepa Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, pages 621–635. Springer, 2018.
- Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P. Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8238–8248, 2022.
- Adam Tonderski, Carl Lindström, Georg Hess, William Ljungbergh, Lennart Svensson, and Christoffer Petersson. Neurad: Neural rendering for autonomous driving. *arXiv preprint arXiv:2311.15260*, 2023.
- Nivedita Tripathi and Senthil Yogamani. Trained trajectory based automated parking system using visual slam on surround view cameras. *arXiv preprint arXiv:2001.02161*, 2020.
- Haithem Turki, Jason Y Zhang, Francesco Ferroni, and Deva Ramanan. Suds: Scalable urban dynamic scenes. In *Computer Vision and Pattern Recognition (CVPR)*, 2023.
- Yu Wang, Xiaobao Wei, Ming Lu, and Guoliang Kang. Plgs: Robust panoptic lifting with 3d gaussian splatting. *IEEE Transactions on Image Processing*, 2025.
- Xiaobao Wei, Renrui Zhang, Jiarui Wu, Jiaming Liu, Ming Lu, Yandong Guo, and Shanghang Zhang. Nto3d: Neural target object 3d reconstruction with segment anything. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20352–20362, 2024.

- Xiaobao Wei, Peng Chen, Guangyu Li, Ming Lu, Hui Chen, and Feng Tian. Gazegaussian: High-fidelity gaze redirection with 3d gaussian splatting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13293–13303, 2025a.
- Xiaobao Wei, Peng Chen, Ming Lu, Hui Chen, and Feng Tian. Graphavatar: Compact head avatars with gnn-generated 3d gaussians. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8295–8303, 2025b.
- Xiaobao Wei, Qingpo Wuwu, Zhongyu Zhao, Zhuangzhe Wu, Nan Huang, Ming Lu, Ningning Ma, and Shanghang Zhang. Emd: Explicit motion modeling for high-quality street gaussian splatting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 28462–28472, 2025c.
- Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moenne-Loccoz, and Zan Gojcic. 3dgut: Enabling distorted cameras and secondary rays in gaussian splatting. *arXiv preprint arXiv:2412.12507*, 2025.
- Zizhang Wu, Weiwei Sun, Man Wang, Xiaoquan Wang, Lizhu Ding, and Fan Wang. Psdet: Efficient and universal parking slot detection. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 290–297. IEEE, 2020.
- Zirui Wu, Tianyu Liu, Liyi Luo, Zhide Zhong, Jianteng Chen, Hongmin Xiao, Chao Hou, Haozhe Lou, Yuantao Chen, Runyi Yang, Yuxin Huang, Xiaoyu Ye, Zike Yan, Yongliang Shi, Yiyi Liao, and Hao Zhao. Mars: An instance-aware, modular and realistic simulator for autonomous driving. *CICAI*, 2023.
- Tianyi Yan, Dongming Wu, Wencheng Han, Junpeng Jiang, Xia Zhou, Kun Zhan, Cheng-zhong Xu, and Jianbing Shen. Drivingsphere: Building a high-fidelity 4d world for closed-loop simulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27531–27541, 2025.
- Yunzhi Yan, Haotong Lin, Chenxu Zhou, Weijie Wang, Haiyang Sun, Kun Zhan, Xianpeng Lang, Xiaowei Zhou, and Sida Peng. Street gaussians: Modeling dynamic urban scenes with gaussian splatting. In *ECCV*, 2024.
- Jiawei Yang, Boris Ivanovic, Or Litany, Xinshuo Weng, Seung Wook Kim, Boyi Li, Tong Che, Danfei Xu, Sanja Fidler, Marco Pavone, and Yue Wang. Emernerf: Emergent spatial-temporal scene decomposition via self-supervision. *arXiv preprint arXiv:2311.02077*, 2023.
- Xuemeng Yang, Licheng Wen, Tiantian Wei, Yukai Ma, Jianbiao Mei, Xin Li, Wenjie Lei, Daocheng Fu, Pinlong Cai, Min Dou, et al. Drivearena: A closed-loop generative simulation platform for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 26933–26943, 2025.
- Yunfan Yang, Denglong Chen, Tong Qin, Xiangru Mu, Chunjing Xu, and Ming Yang. E2e parking: Autonomous parking by the end-to-end neural network on the carla simulator. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 2375–2382. IEEE, 2024.
- Vickie Ye, RUILONG Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025.
- Jinze Yu, Jiaming Liu, Xiaobao Wei, Haoyi Zhou, Yohei Nakata, Denis Gudovskiy, Tomoyuki Okuno, Jianxin Li, Kurt Keutzer, and Shanghang Zhang. Mttrans: Cross-domain object detection with mean teacher transformer. In *European Conference on Computer Vision*, pages 629–645. Springer, 2022.
- Kai Zeng, Zhanqian Wu, Kaixin Xiong, Xiaobao Wei, Xiangyu Guo, Zhenxin Zhu, Kalok Ho, Lijun Zhou, Bohan Zeng, Ming Lu, et al. Rethinking driving world model as synthetic data generator for perception tasks. *arXiv preprint arXiv:2510.19195*, 2025.
- Hongyu Zhou, Jiahao Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, and Yiyi Liao. Hugs: Holistic urban 3d scene understanding via gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21336–21345, 2024a.
- Xiaoyu Zhou, Zhiwei Lin, Xiaojun Shan, Yongtao Wang, Deqing Sun, and Ming-Hsuan Yang. Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21634–21643, 2024b.
- Andrea Zinelli, Luigi Musto, and Fabio Pizzati. A deep-learning approach for parking slot detection on surround-view images. In *2019 IEEE intelligent vehicles symposium (IV)*, pages 683–688. IEEE, 2019.