Enhancing AVM-based parking-slot detection with synthetic data

Citation for the original published paper (version of record):
Zhang, S., Liu, Y., Gao, K. (2026). Enhancing AVM-based parking-slot detection with synthetic data.
JOURNAL OF INTELLIGENT AND CONNECTED VEHICLES, 9(1).
http://dx.doi.org/10.26599/JICV.2025.9210075

J Intel Connect Veh

Research Article

Enhancing AVM-based parking-slot detection with synthetic data

Song Zhang[1], Yang Liu[2,✉], Kun Gao[3]

Cite this article: Zhang S, Liu Y, Gao K. *J Intel Connect Veh* 2026, **9**(1): 9210075. https://doi.org/10.26599/JICV.2025.9210075

ABSTRACT: Parking-slot detection is pivotal for autonomous driving, facilitating automated parking and enhancing vehicle safety. However, the diversity of parking slot shapes and the complexity of surrounding environments incur significant costs in data collection and annotation. To address this challenge, we propose a novel data synthesis framework specifically tailored for around-view monitor (AVM) images. First, an inpainting-based generative algorithm eliminates foreground elements from real parking slot images to produce clean backgrounds. Subsequently, new foreground elements exhibiting diverse shapes, colors, and textures are superimposed onto these backgrounds. Furthermore, rather than relying on domain selection, we introduce a data selection strategy based on active learning that operates directly on the generated datasets. The controllable attributes of synthetic data facilitate the effective evaluation and optimization of the selection strategy across various scenarios. Experimental results on the panoramic surround view (PSV) dataset demonstrate that models trained exclusively with synthetic data achieve 1.32% higher precision than those trained only on real images. Moreover, integrating 40% real images with synthetic data increases precision by up to 1.74% and recall rates by up to 1.48%, highlighting the effectiveness and practical utility of our proposed approach.

KEYWORDS: autonomous driving; synthetic data; object detection; around-view monitor (AVM)

1 Introduction

Autonomous driving perception is a pivotal field, where parking-slot detection stands out as one of its most significant practical applications (Ma and Xue, 2024; Wang, 2017). Detection is typically performed on around-view monitor (AVM) images. Unlike imagery obtained from pinhole cameras, the acquisition and stitching of AVM data require specialized processing procedures (Liu et al., 2023). Specifically, the process involves collecting data from multiple fisheye cameras (generally 4 channels), followed by image projection and image stitching according to the intrinsic and extrinsic parameters of these cameras. However, the acquisition of AVM images incurs substantial costs due to the complex multicamera setup and rigorous calibration requirements, while manual annotation remains labor intensive (Yang et al., 2026). The diversity of parking scenarios—ranging from standard rectangular slots to irregular spaces and from different lighting conditions to various surface markings—demands extensive data coverage. Consequently, synthetic data generation emerges as a viable solution to address these challenges (Yang et al., 2026).

In the realm of autonomous driving systems, synthetic data have gained significant traction due to the difficulties associated with data collection and manual annotation. Extensive research has focused on data synthesis techniques as a potential solution to the challenge of AVM data collection (Lee et al., 2018). Generative approaches require historical data for training, and their generation effect often lacks controllability. Conversely, rendering-based methods involve high computational costs, and the fidelity of images produced by game engines frequently proves insufficient (Hou et al., 2022a). Approaches that superimpose objects onto real images achieve high realism because the background retains its authenticity; however, these methods require the identification of suitable empty spaces within the image for object placement. Building upon this strategy, we have enhanced this methodology to further improve image quality (Gupta et al., 2016).

We propose a data synthesis framework for parking-slot detection based on an inpainting-and-compositing mechanism. First, an inpainting network eliminates foreground elements from the AVM imagery to acquire a clean background. Subsequently, new parking slots defined by controllable parameters—such as shapes, colors, and textures—are superimposed onto the background via image composition. These parameters correspond to distinct data domains, allowing for the on-demand generation of datasets tailored to specific requirements. Fig. 1 presents examples of synthetic images across different domains. The subfigures highlight the diversity achieved by our synthesis engine: The top row depicts variations in lighting conditions and brightness; the middle row displays different geometric shapes and line curvatures; and the bottom row illustrates diverse ground textures and noise types. Such domain randomization is essential for training robust detection models. We conduct experiments on the panoramic surround view (PSV) dataset, and Fig. 2 illustrates the main pipeline of our parking-slot synthesis framework (Li et

[1] Z-one Technology Co., Ltd., Shanghai 201805, China. [2] COSCO SHIPPING Development Co., Ltd., Shanghai 200120, China. [3] Department of Architecture and Civil Engineering, Chalmers University of Technology, Goteburg SE-412 96, Sweden.

✉ Corresponding author. E-mail: liu.yang33@coscoshipping.com

Received: October 29, 2025; Revised: December 23, 2025; Accepted: December 30, 2025

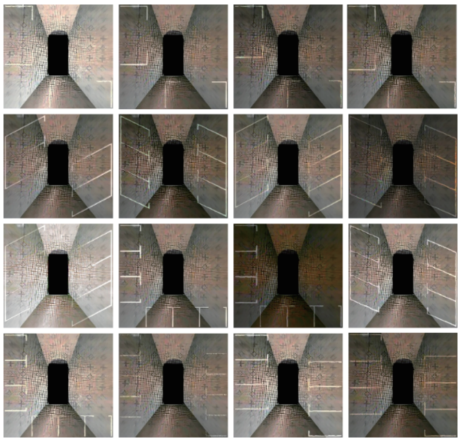

Fig. 1 Generated parking slots with various domains.

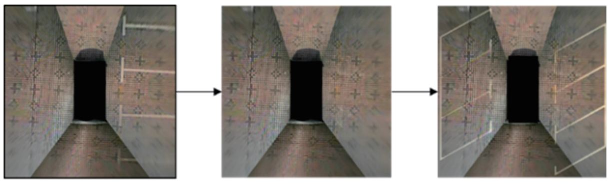

Fig. 2 Main pipeline of the proposed parking-slot synthesis framework.

al., 2020). Our method can also be applied to other scenarios beyond parking-slot synthesis, such as ground marking generation and lane line synthesis, where single lines can be converted into double lines. The data synthesized via this approach preserve high fidelity, as a significant portion of the background is derived from real-world imagery.

We observe that merely utilizing synthetic images does not necessarily improve algorithm performance. In fact, an excessive quantity of synthetic data can detrimentally affect the original models. To address this, we propose a data selection strategy based on active learning. This strategy employs the detection loss of marking point detection as an evaluation metric. Furthermore, rather than performing domain selection, we apply this selection process directly across the entire pool of generated images.

Experimental results on the PSV dataset, utilizing a graph convolutional network (GCN), demonstrate the efficacy of our synthetic data and selection strategy. Models trained exclusively with synthetic images achieve higher precision than those trained solely on real images across different parking slot detection models (Min et al., 2021). When we combine 40% real images with synthetic images, both precision and recall rates obtain significant improvements. These results suggest that synthetic data hold the potential to be sufficient for effective parking slot detection.

Our research addresses the challenge of data efficiency in parking-slot detection through a high-fidelity synthesis pipeline. The main contributions of this study are summarized as follows:

● To the best of our knowledge, we are the first to combine semantic image inpainting with active learning for AVM parking-slot synthesis. Unlike traditional approaches that overlay objects onto random backgrounds, our inpainting-based readding schema ensures background semantic consistency, significantly enhancing the fidelity of synthetic samples.

● We propose the close-to-far data selection (CFDS) strategy. Deviating from previous methods that rely on coarse-grained domain selection (e.g., selecting entire 'sunny' or 'rainy' datasets), our strategy utilizes a dynamic ranking mechanism to filter synthetic images at the instance level. This allows for a more precise optimization of the training distribution by discarding geometrically inconsistent samples.

● Through extensive experiments, we identify a critical 'saturation point' at a 40% real data mixture, where the model achieves optimal performance (1.74% precision gain). This finding offers a practical guideline for balancing synthetic and real data volumes in resource-constrained autonomous driving applications.

2 Relate work

2.1 Parking-slot detection

In recent years, deep learning has significantly advanced the field of parking-slot detection (Bui and Suhr, 2023; Zinelli et al., 2019). Notably, two-stage methods such as DeepPS and DMPR-PS adopt a detection-pairing strategy: These methods primarily identify parking slot markings via object detection techniques, subsequently utilizing learning-based algorithms to pair these markings into complete parking slots (Huang et al., 2019; Yang et al., 2025; Zhang et al., 2018). While this methodology benefits from the interpretability of intermediate results, it remains susceptible to error propagation between stages.

Concurrently, alternative approaches have explored distinct problem formulations. Semantic segmentation methods formulate parking-slot detection as a pixel-level classification task, directly predicting slot boundaries and orientations (Chen et al., 2025; Jiang et al., 2019; Wu et al., 2018). Similarly, graph neural networks (GNNs) have been employed to model the spatial relationships between detected markings. Furthermore, fully connected network architectures aim for end-to-end detection by directly regressing slot parameters from input images, thereby circumventing the need for explicit marking detection (Yu et al., 2020)

2.2 Data synthesis

In autonomous driving, data synthesis has emerged as a promising solution to mitigate the scarcity of annotated training datasets. Early approaches focused on simple augmentation techniques, such as superimposing foreground elements onto empty image regions, but these methods are limited in their ability to handle complex scene modifications (Gupta et al., 2016). Our proposed inpainting-based composition framework effectively addresses this limitation.

Recent breakthroughs in generative models have facilitated significant progress in image synthesis across diverse applications. Domain transfer methods such as CycleGAN have been applied to bridge the sim-to-real gap, enabling driving policies trained on synthetic data to generalize to real-world scenarios (Bewley et al., 2019; Zhu et al., 2017). Other studies have enhanced synthetic data realism by incorporating adversarial training (Hou et al., 2022b; Tripathi et al., 2019; Zhang et al., 2024), where discriminative models guide the synthesis process to produce more realistic outputs. For object detection, researchers have developed sophisticated data generation and augmentation strategies to generate diverse training samples (Jiang et al., 2023; Zoph et al., 2020). More recently, diffusion-based approaches such as FreeMask have demonstrated the ability to generate high-quality images derived from segmentation masks (Yang et al., 2023). Nonetheless, these methods generally lack precise control and quantification over specific domain attributes. This limitation impedes targeted modifications of particular image aspects, thereby restricting adaptability. It is also difficult to accurately analyze the impact of each factor in generating the entire image.

Recently, generative artificial intelligence (AI) has advanced significantly with diffusion models and neural radiance fields (NeRFs). For instance, PriorFusion demonstrates the integration of generative priors for robust perception, while S-NeRF++ enables autonomous driving simulation via neural reconstruction. While these methods achieve state-of-the-art fidelity, they pose specific challenges for parking-slot detection. Diffusion models often lack the precise geometric controllability required to generate strictly parallel parking lines without complex conditioning. Similarly, NeRF-based approaches typically require dense multiview sequences for 3D reconstruction, which are rarely available in standard single-frame AVM datasets. In contrast, our proposed 'Inpainting and Readding' schema offers a balanced solution, providing explicit geometric control and high efficiency suitable for 2D AVM imagery.

2.3 Data selection

Data selection techniques are designed to curate compact, task-specific subsets from extensive datasets. These methods can be broadly categorized into uncertainty-based, impact-based, and hybrid approaches. Uncertainty-based methods prioritize samples exhibiting high ambiguity, utilizing metrics such as maximum class probability (Settles et al., 2008), margin confidence (Scheffer et al., 2001), and prediction entropy (Holub et al., 2008). The underlying premise is that uncertain samples provide more learning value than confidently predicted samples. Conversely, impact-based approaches focus on the influence of individual samples on the training process, typically quantified through loss or gradient magnitudes (Paul et al., 2021), as exemplified by the gradient norm time distance (GraNd) metric. Additionally, certain studies track misclassification patterns across training epochs to pinpoint consistently challenging examples (Toneva et al., 2018). Hybrid methods combine multiple selection criteria to balance exploration and exploitation. For instance, density-based approaches consider both sample uncertainty and feature distribution to avoid selecting redundant examples from similar regions (Ebert et al., 2012).

In the context of domain adaptation, Roy et al. (2021) proposed easy-to-hard domain selection (EHDS) to establish a sequence for domain selection based on target domain metrics. However, this domain-centric approach relies on the assumption of distinct domain boundaries, which may prove inadequate for complex real-world scenarios where informative samples are often dispersed across multiple domains. In the specific case of parking-slot detection, the primary challenge involves selecting synthetic samples that effectively complement real data by encompassing the diverse variations inherent in AVM images. Consequently, this necessitates a transition from rigid domain selection to more flexible data selection strategies (Roy et al., 2021).

3 Method

3.1 Parking-slot synthesis framework

First, we extract the lane mask on the road through semantic segmentation and employ the large mask inpainting (LaMa) network to generate clean backgrounds. LaMa incorporates fast fourier convolutions (FFC), enabling image-wide receptive fields that are particularly effective for reconstructing global geometric structures and periodic textures common in road surfaces. We utilize the official model pretrained on the Places2 dataset to leverage its zero-shot generalization capabilities to effectively handle diverse road surfaces without fine-tuning our AVM data (Suvorov et al., 2022).

Subsequently, potential parking-slot regions are identified from the AVM images with the help of semantic segmentation (Tao et al., 2020). To convert these pixelwise masks into vectorized geometric shapes, the perimeter of each identified region is approximated using the Douglas-Peucker algorithm. Specifically, we set the approximation threshold ε to 1% of the contour's perimeter (i.e., $\varepsilon = 0.01 \times arcLength$). This adaptive thresholding ensures that jagged edges caused by segmentation noise are smoothed out, while the essential rectangular geometry and corner vertices of the parking slots are preserved.

Finally, these contour patterns are transformed into masks through homography (Douglas and Peucker, 1973; Gao et al., 2025). To accurately represent the planar mapping between the generated pattern and the AVM image, we employ homogeneous coordinates. The transformation is defined as

$$s\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{1}$$

where (x, y) and (x', y') represent coordinates of corresponding point pairs from the generated pattern and the AVM image, respectively. The transformation is modeled by a 3×3 homography matrix $\boldsymbol{H}$, with $\boldsymbol{s}$ representing the projective scaling factor. This formulation enables the perspective mapping required for synthesizing AVM images, where h_{11} to h_{33} are the elements of $\boldsymbol{H}$.

The complete workflow of our proposed framework is illustrated in Fig. 3. The synthetic image requires obtaining parking slots on the road. Since the AVM images already contain lane lines, directly adding parking slots would be unrealistic. Therefore, we first erase the lane markings from the road surface and then insert the parking slots into the cleared area.

To emulate the diversity of real-world scenarios and enhance the generalization capabilities of our model, we implement a domain randomization strategy to generate synthetic parking-slot data that closely mimic actual environments. Specifically, we generate images by randomly varying the following aspects:

- Quantity and location: Synthetic parking slots are positioned adjacent to roadsides, featuring adjustable lateral offsets and randomized slot quantities.
- Color: RGB values for parking-slot markings are sampled from real-world examples and subsequently perturbed a range of ± 30 per channel (Uint8 format).
- Shape: The geometry of the parking slots is randomly selected from perpendicular, parallel, or slanted configurations.
- Texture: The texture of parking lines is diversified utilizing patterns generated by ZenBG.
- Noise: Different noise types, including Gaussian noise, Laplacian noise, and salt-and-pepper noise, are added to simulate realistic conditions.
- Curvature: Elastic transformations introduce curvature to parking-slot lines, with the deformation factors set between 100 and 200.
- Brightness: Global illumination is modified by scaling the brightness channel with a factor uniformly sampled between 0.5 and 1.5.

After the geometric transformation defined in Eq. (1), simply overlaying the generated masks onto the background would result in unnatural artifacts. To ensure visual consistency, we employ a consistency-aware composition strategy. First, a Gaussian blur with a 5×5 kernel is applied to the mask edges to eliminate high-frequency aliasing. Subsequently, the foreground is composited using Poisson image editing, which solves the Poisson partial

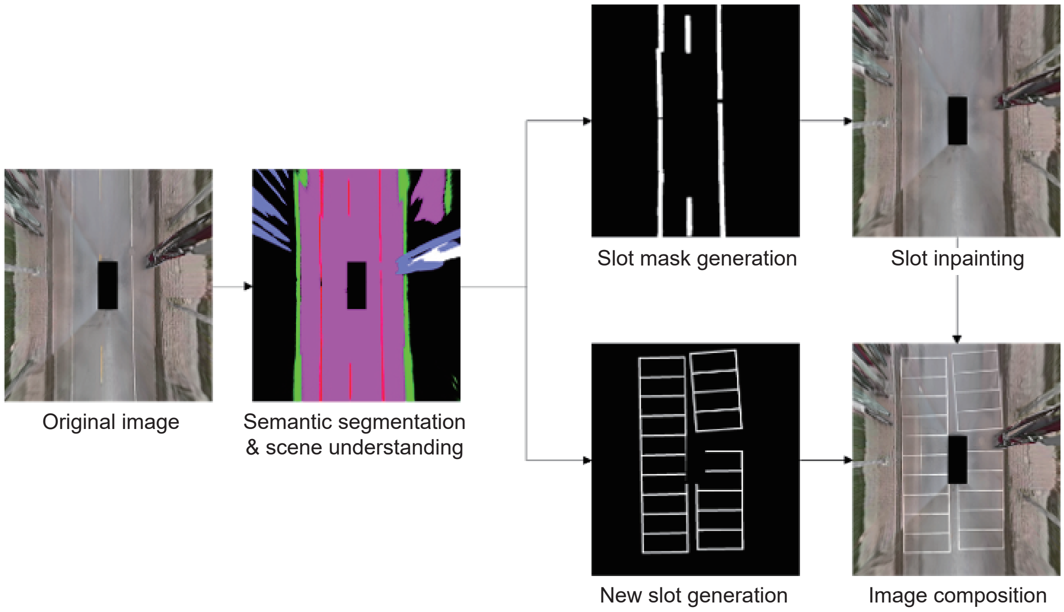

Fig. 3 Procedure of synthesizing parking slots in AVM images.

differential equation with Dirichlet boundary conditions. This seamlessly blends the slot's illumination and color gradients with the real background, ensuring high visual fidelity.

3.2 Data selection strategy

Our inpainting-based composition framework facilitates the generation of a substantial volume of images. However, quantitative evaluations indicate that an excessively large dataset can negatively impact model performance. Consequently, it is imperative to strategically select appropriate images from the synthetic pool to build a more efficient training dataset.

To address this, we introduce a strategy termed close-to-far data selection (CFDS). In this strategy, the ‘closeness’ of a synthetic target image to the original source data is evaluated based on the prediction accuracy of a model trained on real data. Specifically, the calculation scope of the point loss focuses on the discrepancy between the predicted marking points on the synthetic images (using a model pretrained on real PSV data) and the ground truth labels generated during synthesis. In parking-slot detection, the point loss is defined as

$$\mathrm{loss}_{\mathrm{point}} = \frac{1}{S^2}\sum_{n=1}^{S^2}\left\{(c_i-\widehat{c}_i)^2+1_i\left[(x_i-\widehat{x}_i)^2+(y_i-\widehat{y}_i)^2\right]\right\} \tag{2}$$

where S denotes the grid size of the output layer divided by the marking-point detector; (x_i, y_i) and c_i represent the projected position of the marking point and the confidence of cell i, respectively; the symbols with hats $\widehat{c}_i, \widehat{x}_i, \widehat{y}_i$ indicate the corresponding ground truth; and 1_i is a threshold function. Regarding the selection threshold, we adopt a dynamic ranking strategy instead of a fixed scalar value. We calculate the $\mathrm{loss}_{\mathrm{point}}$ for all generated synthetic images and sort them in ascending order. Images with lower loss are considered “Close” (high geometric consistency), while those with high loss are considered “Far” (unrealistic). We determine the cutoff quantity empirically; as demonstrated in our ablation study, we retain the top 24,000 images where the validation performance peaks. This formulation of point loss exhibits a strong positive correlation with detection accuracy, making it an ideal metric for data selection.

4 Experiments

We aim to comprehensively evaluate the effectiveness of our proposed method across various models on public datasets. Furthermore, we have investigated strategies utilizing synthetic data, including ensemble learning and a pretraining and fine-tuning pipeline integrated with data selections.

4.1 Implementation details

4.1.1 Benchmark dataset

Publicly accessible datasets suitable for parking-slot detection tasks are limited, particularly datasets that include masks for parking-slot lines essential for synthesizing distinct parking slots without duplication. Although several large-scale datasets exist in the parking domain, such as CNRPark and PKLot, they are primarily designed for parking lot occupancy detection via surveillance cameras. These datasets present two major limitations for our specific task: (1) viewpoint mismatch: They feature oblique perspective views with significant distortion, whereas our framework is tailored for the orthographic, top-down projection characteristic of AVM systems; (2) annotation granularity: They typically provide bounding boxes or occupancy status rather than the fine-grained parking-line segmentation masks required for our inpainting-based synthesis pipeline. Consequently, PSV stands as the only public dataset meeting our requirements; therefore, we choose it as our benchmark dataset (Li et al., 2020). The PSV dataset consists of 2550 training images and 1274 test images. Each subset contains images captured from distinct scenes at a resolution of 600 × 600 pixels, corresponding to a 10 m × 10 m physical plane area. Using domain randomization, we expand the initial training set to approximately 300,000 synthetic images for comprehensive experimentation.

4.1.2 Experimental settings

We select three representative parking-slot detection models to validate the utility of synthetic data: DeepPS, DMPR-PS and GCN. Consistent with DeepPS, we employ precision-recall metrics for performance assessment. Specifically, $S_g = \{p_1^g, p_2^g\}$ is set as the ground truth, and $S_d = \{p_1^d, p_2^d\}$ is set as the detected marking point pair. If the following condition is satisfied, the detected pair S_d is considered a true positive, and the corresponding ground-truth pair S_g is deemed correctly detected:

$$\left\| \left(p_1^g - p_1^d, p_2^g - p_2^d \right) \right\|_2 < 10 \tag{3}$$

Otherwise, S_d is categorized as a false positive and S_g as a false negative parking-slot pair.

To visually verify the robustness of our model in industrial scenarios, we present the qualitative detection results on the Boden AVM dataset in Fig. 4. As observed, the model accurately localizes parking slots even under challenging conditions, confirming the effectiveness of our synthesis-based training strategy for real-world deployment.

4.2 Results and discussions

As presented in Table 1, we evaluate three models utilizing various data strategies: original data, full synthetic data, synthetic data after data selection, ensemble learning, join-training, and pretraining combined with fine-tuning. Among approaches relying exclusively on synthetic data, the implementation of our data selection strategy significantly enhances parking-slot detection performance, surpassing even the precision achieved with original training data.

When we combine synthetic data with original data, we find that including as little as 10% of real training images in the synthetic dataset results in better precision and recall rates compared to using only the original data. Model performance reaches its peak as the proportion of real images increases to 40%. However, adding real images beyond this threshold leads to diminishing returns. We attribute this phenomenon to 'information saturation' within the domain distribution. Our high-quality synthetic data, refined by the selection strategy, establish a robust foundation for the geometric and structural features required for detection. Consequently, the introduction of real data primarily serves to bridge the 'Sim-to-Real' domain gap (e.g., correcting texture and lighting discrepancies). Our empirical results suggest that a 40% real data ratio provides sufficient diversity to cover this domain shift. Beyond this point, additional real data introduce distributional redundancy with minimal marginal information gain. Furthermore, it may introduce label

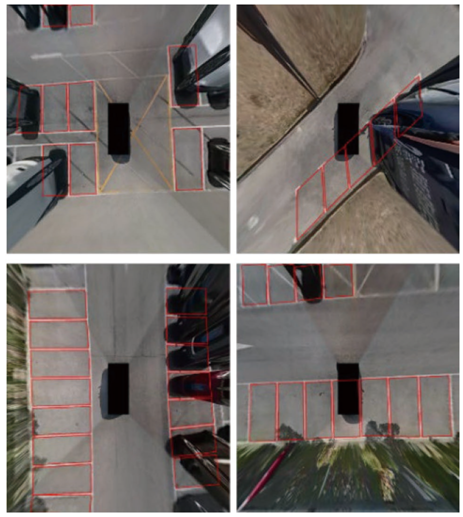

Fig. 4 Qualitative detection results on the self-collected Boden AVM dataset.

Table 1 Results of parking-slot detection of different models with different data processing methods on PSV

Training data	Paradigm	Model	Precision (%)	Recall (%)
Real	Real image	DeepPS	93.54	88.67
		DMPR-PS	91.35	82.97
		GCN	95.88	92.62
Synthetic	All synthetic image	DeepPS	42.61	52.63
		DMPR-PS	47.46	51.74
		GCN	49.26	56.37
	Data selection	DeepPS	94.86 (↑1.32)	66.63
		DMPR-PS	92.02 (↑0.67)	67.14
		GCN	96.42 (↑0.54)	66.27
	Ensemble learning	DeepPS	93.26	78.56
		DMPR-PS	90.49	77.65
		GCN	91.14	75.86
Real + Synthetic	Joint-training (10% Real)	DeepPS	93.72 (↑0.18)	88.93 (↑0.26)
		DMPR-PS	91.79 (↑0.44)	83.15 (↑0.18)
		GCN	96.24 (↑0.36)	93.11 (↑0.49)
	Joint-training (20% Real)	DeepPS	94.17 (↑0.63)	91.05 (↑2.38)
		DMPR-PS	92.01 (↑0.66)	83.36 (↑0.39)
		GCN	96.44 (↑0.56)	94.02 (↑1.40)
	Joint-training (40% Real)	DeepPS	94.45 (↑0.91)	90.15 (↑1.48)
		DMPR-PS	92.53 (↑1.18)	84.08 (↑1.11)
		GCN	97.62 (↑1.74)	93.95 (↑1.33)
	Joint-training (100% Real)	DeepPS	94.32 (↑0.78)	90.21 (↑1.54)
		DMPR-PS	92.17 (↑0.82)	83.24 (↑0.27)
		GCN	96.35 (↑0.47)	93.84 (↑1.22)
	Pretraining	DeepPS	95.05 (↑1.51)	90.34 (↑1.67)
		DMPR-PS	91.88 (↑0.53)	84.12 (↑0.04)
		GCN	95.92 (↑0.04)	93.11 (↑0.49)

Table 2 Results of parking-slot detection of different models with different data processing methods on Boden AVM

Training data	Paradigm	Model	Precision (%)	Recall (%)
Real	Real image	DeepPS	81.34	72.25
		DMPR-PS	83.15	74.62
		GCN	85.63	77.74
Synthetic	All synthetic image	DeepPS	33.74	42.85
		DMPR-PS	32.77	40.65
		GCN	34.27	43.82
	Data selection	DeepPS	82.45 (↑1.12)	60.37
		DMPR-PS	83.97 (↑0.82)	58.38
		GCN	85.87 (↑0.24)	63.26

noise inherent in manual annotations, which contrasts with the pixel-perfect precision of synthetic labels. These findings indicate that our synthetic data, supplemented by a moderate amount of real-world data, can effectively achieve superior training outcomes—a conclusion further supported by our pretraining experiments.

Additionally, we evaluated our approach on the nonpublic Boden AVM dataset, from which we selected 4000 images for training and 2000 images for testing. The results, presented in Table 2, reinforce our findings, demonstrating that our synthetic data strategy yields consistent performance improvements across different datasets.

4.3 Ablation study

4.3.1 Impact of different domains

To evaluate the effectiveness of different domains, we initially define a "base domain" comprising the shape, spatial arrangement and color of parking slots as one base domain, which determines the fundamental conditions of parking slot lines. Synthetic parking slots were generated utilizing the original parking-slot line masks, as illustrated in Fig. 5. Subsequently, we introduce individual additional domains characterized by randomized parameters, followed by the application of data selection. Table 3 summarizes the detection results derived from synthetic images selected across all domains. Our observations indicate that the precision and recall rates improved to varying degrees when a single domain was added. Moreover, incorporating all domains together significantly enhanced performance. This finding underscores that data selection is more suitable than domain selection for our synthetic data. Among the various domains, the base domain, which dictates the parking-slot line mask, exerts the most substantial impact on the detection performance. Leveraging attributes from original parking slots yields results significantly superior to those obtained using purely random values. Furthermore, the magnitude of improvement varied across different synthesis configurations. For random parking-slot mask data, curvature shows the greatest enhancement, while brightness shows the least. Conversely, when utilizing original parking-slot masks, brightness provided the greatest gain, while noise contributed minimally. This suggests that while different domains influence distinct data characteristics, the cumulative effect of multiple domains surpasses the contribution of any single domain.

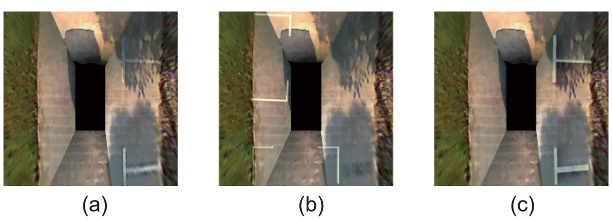

Fig. 5 (a) Original image; (b) synthetic image with random parking-slot mask; (c) synthetic image with original parking-slot mask.

Table 3 Parking-slot detection results of different domains with different parking-slot line masks

Domain	Line mask	Precision (%)	Recall (%)	Average increase (%)
Base	Random	70.32	43.47	0
	Original	97.88	84.84	0
Texture	Random	87.59	47.46	10.63
	Original	98.95	92.27	4.25
Noise	Random	89.76	45.95	10.96
	Original	99.34	88.59	2.61
Curvature	Random	83.98	51.93	11.06
	Original	99.15	91.75	4.09
Brightness	Random	81.38	43.56	5.57
	Original	99.21	92.14	4.31
All	Random	96.42	66.27	24.45
	Original	99.54	94.70	5.76

4.3.2 Data selection quantities

In this section, we evaluate distinct thresholds for point detection loss to filter varying quantities of synthetic images. As presented in Table 4, the results indicate that performance improves as the number of selected images increases. However, a performance decline is observed when the threshold is set excessively high, leading to the inclusion of low-quality images. Consequently, we adopt a subset of 24,000 images as the optimal configuration for subsequent experimental steps.

4.3.3 Ablation comparison of ensemble learning

To ensure a rigorous evaluation, we conduct a comparison analysis of ensemble learning methodologies. We partition the selected synthetic images into varying numbers of partitions and subsequently apply three different ensemble strategies. As presented in Table 5, the results indicate that a configuration of 10

Table 4 Results of different quantities for selected synthetic images

Quantity	Precision (%)	Recall (%)
10,000	96.72	50.48
17,000	95.64	62.16
24,000	96.42	66.27
32,000	93.43	63.84
40,000	93.07	57.89

Table 5 Parking-slot detection results of different partitions and different ensemble strategies

Partition	Option	Precision (%)	Recall (%)
1	—	96.42	66.27
	Affirmative	80.14	75.42
5	Unanimous	97.35	48.56
	Consensus	88.57	73.18
	Affirmative	80.33	76.65
7	Unanimous	98.02	49.36
	Consensus	90.54	74.44
	Affirmative	80.92	79.74
10	Unanimous	99.12	50.56
	Consensus	91.14	75.86
	Affirmative	77.68	74.64
15	Unanimous	92.16	44.59
	Consensus	87.34	70.41

partitions achieves the most uniform distribution, thereby yielding the optimal ensemble learning detection performance. Furthermore, utilizing the consensus ensemble strategy yields a more balanced precision and recall rate.

5 Additional topic

Our data synthesis framework extends to analogous scenarios, such as the conversion of single road lines into double lines and the modification of ground markings, as illustrated in Fig. 6. Given that the methodological and analytical frameworks for these applications parallel those previously discussed, detailed elaboration is omitted to avoid redundancy. However, it is important to note that these applications necessitate symbol identification prior to the execution of homography transformation—a process comprehensively detailed in our prior research (Bie et al., 2023).

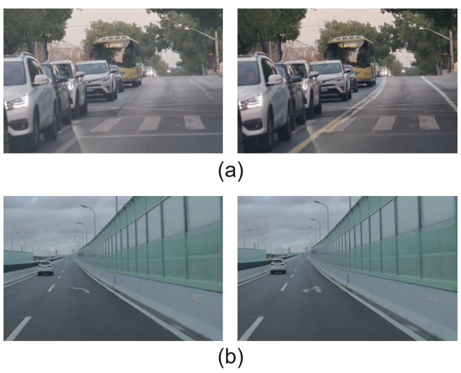

Fig. 6 Other applications of our data synthesis method: (a) single marking lines to double ones; (b) synthetic road markings.

6 Conclusions

In this study, we propose a data generation framework integrated with an active learning-based data selection strategy designed to enhance parking-slot detection for autonomous driving. Experimental evaluations demonstrate the superiority of our approach; specifically, by selectively integrating synthetic data with real-world samples, we achieved significant improvements in precision and recall compared to baselines trained exclusively on real data. Our findings highlight that prioritizing geometric fidelity through dynamic ranking, rather than indiscriminate data expansion, is crucial for optimizing the training distribution.

Furthermore, this work holds substantial practical value for the autonomous driving industry. By enabling high-quality data generation without expensive manual annotation, our method effectively addresses the data acquisition bottleneck. The successful validation on industrial-grade datasets (e.g., from the SAIC ES33 platform) further confirms its readiness for real-world deployment, offering a scalable solution for Sim-to-Real adaptation in commercial parking perception systems.

While the current validation focuses on parking-slot detection, the proposed “inpainting-and-selection” paradigm is generic. Future research will focus on extending this methodology to broader vehicle perception tasks (e.g., lane markings and road signs), thereby enriching the comprehensive data landscape for autonomous driving.

Replication and data sharing

Our code is available at: https://github.com/zhangsongdmk/avm_parkinglot_syn.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Bewley, A., Rigley, J., Liu, Y., Hawke, J., Shen, R., Lam, V. D., et al., 2019. Learning to drive from simulation without real world labels. In: 2019 International Conference on Robotics and Automation (ICRA), 4818–4824.

Bie, X., Zhang, S., Meng, C., Mei, J., Li, J., He, X., 2023. Synthetic data for 2D road marking detection in Autonomous Driving. https://doi.org/10.4271/2023-01-7046

Bui, Q. H., Suhr, J. K., 2023. One-stage parking slot detection using component linkage and progressive assembly. *IEEE Intell Transp Syst Mag*, **15**, 33–48.

Chen, L., Dong, T., Li, X., Xu, X., 2025. Logistics engineering management in the platform supply chain: An overview from logistics service strategy selection perspective. *Engineering*, **47**, 236–249.

Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, **10**, 112–122.

Ebert, S., Fritz, M., Schiele, B., 2012. RALF: A reinforced active learning formulation for object class recognition. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 3626–3633.

Gao, Z., Jia, B., Xie, D., Wang, W., Wu, J., 2025. A discussion on the complexity and transit mechanisms of urban traffic systems. *Engineering*, **44**, 24–29.

Gupta, A., Vedaldi, A., Zisserman, A., 2016. Synthetic data for text localisation in natural images. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2315–2324.

Holub, A., Perona, P., Burl, M. C., 2008. Entropy-based active learning for object recognition. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 1–8.

Hou, J., Chen, Q., Cheng, Y., Chen, G., Xue, X., Zeng, T., et al., 2022a. SUPS: A simulated underground parking scenario dataset for autonomous driving. In: 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2265–2271.

Hou, Y., Li, C., Lu, Y., Zhu, L., Li, Y., Jia, H., et al., 2022b. Enhancing and dissecting crowd counting by synthetic data. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2539–2543.

Huang, J., Zhang, L., Shen, Y., Zhang, H., Zhao, S., Yang, Y., 2019. DMPR-PS: A novel approach for parking-slot detection using directional marking-point regression. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), 212–217.

Jiang, W., Wu, Y., Guan, L., Zhao, J., 2019. DFNet: Semantic segmentation on panoramic images with dynamic loss weights and residual fusion block. In: 2019 International Conference on Robotics and Automation (ICRA), 5887–5892.

Jiang, Y., Gu, C., Xue, Z., Zhang, X., Liu, Y., 2023. Mask-guided image person removal with data synthesis. *IET Image Process*, **17**, 2214–2224.

Lee, M., Kim, S., Lim, W., Sunwoo, M., 2018. Probabilistic occupancy filter for parking slot marker detection in an autonomous parking system using AVM. *IEEE Trans Intell Transp Syst*, **20**, 2389–2394.

Li, W., Cao, L., Yan, L., Li, C., Feng, X., Zhao, P., 2020. Vacant parking slot detection in the around view image based on deep learning. *Sensors*, **20**, 2138.

Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F., Qu, X., 2023. Can language models be used for real-world urban-delivery route optimization? *Innovation*, **4**, 100520.

Ma, C., Xue, F., 2024. A review of vehicle detection methods based on computer vision. *J Intel Connect Veh*, **7**, 1–18.

Min, C., Xu, J., Xiao, L., Zhao, D., Nie, Y., Dai, B., 2021. Attentional graph neural network for parking-slot detection. *IEEE Robot Autom Lett*, **6**, 3445–3450.

Paul, M., Ganguli, S., Dziugaite, G. K., 2021. Deep learning on a data diet: Finding important examples early in training. *Adv Neural Inf Process Syst*, **34**, 20596–20607.

Roy, S., Krivosheev, E., Zhong, Z., Sebe, N., Ricci, E., 2021. Curriculum graph co-teaching for multi-target domain adaptation. In: 2021 IEEE /CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5351–5360.

Scheffer, T., Decomain, C., Wrobel, S., 2001. Active hidden markov models for information extraction. In: International Symposium on Intelligent Data Analysis, 309–318.

Settles, B., Craven, M., Friedland, L., 2008. Active learning with real annotation costs. In: Proceedings of the NIPS Workshop on Cost-Sensitive Learning, 1–10.

Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., et al., 2022. Resolution-robust large mask inpainting with fourier convolutions. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2149–2159.

Tao, A., Sapra, K., Catanzaro, B., 2020. Hierarchical multi-scale attention for semantic segmentation. https://doi.org/10.48550/arXiv.2005.10821

Toneva, M., Sordoni, A., Combes, R. T. D., Trischler, A., Bengio, Y., Gordon, G. J., 2018. An empirical study of example forgetting during deep neural network learning. https://doi.org/10.48550/arXiv.1812.05159

Tripathi, S., Chandra, S., Agrawal, A., Tyagi, A., Rehg, J. M., Chari, V., 2019. Learning to generate synthetic data via compositing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 461–470.

Wang, F. Y., 2017. Artificial intelligence and intelligent transportation: Driving into the 3rd axial age with ITS. *IEEE Intell Transp Syst Mag*, **9**, 6–9.

Wu, Y., Yang, T., Zhao, J., Guan, L., Jiang, W., 2018. VH-HFCN based parking slot and lane markings segmentation on panoramic surround view. In: 2018 IEEE Intelligent Vehicles Symposium (IV), 1767–1772.

Yang, L., Xu, X., Kang, B., Shi, Y., Zhao, H., 2023. Freemask: Synthetic images with dense annotations make stronger segmentation models. *Advances in Neural Information Processing Systems*, **36**, 18659–18675.

Yang, L., Yuan, M., Liu, Y., Qu, X., Hu, Z., Zhang, Z., et al., 2025. Optimization of task scheduling and resource allocation for autonomous vehicle testing in vehicle-road-cloud collaborative systems. *Expert Syst Appl*, **299**, 129943.

Yang, Y., Zhan, J., Xu, M., Liu, Y., Qu, X., 2026. Toward climate-neutral urban mobility: Understanding shared e-scooter carbon emission patterns through multi-city evidence in Europe. *Transp Res Part A Policy Pract*, **203**, 104736.

Yu, Z., Gao, Z., Chen, H., Huang, Y., 2020. SPFCN: Select and prune the fully convolutional networks for real-time parking slot detection. In: 2020 IEEE Intelligent Vehicles Symposium (IV), 445–450.

Zhang, L., Huang, J., Li, X., Xiong, L., 2018. Vision-based parking-slot detection: A DCNN-based approach and a large-scale benchmark dataset. *IEEE Trans Image Process*, **27**, 5350–5364.

Zhang, S., Zhang, J., Yang, L., Chen, F., Li, S., Gao, Z., 2024. Physics guided deep learning-based model for short-term origin–destination demand prediction in urban rail transit systems under pandemic. *Engineering*, **41**, 276–296.

Zhu, J. Y., Park, T., Isola, P., Efros, A. A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2223–2232.

Zinelli, A., Musto, L., Pizzati, F., 2019. A deep-learning approach for parking slot detection on surround-view images. In: 2019 IEEE Intelligent Vehicles Symposium (IV), 683–688.

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J., Le, Q. V., 2020. Learning data augmentation strategies for object detection. In: European Conference on Computer Vision, 566–583.

Song Zhang received the M.S. degree from the Department of Precision Instruments and Mechanical in Tsinghua University in 2007. Currently, he works at Z-one Technology Co., Ltd., SAIC Motor Corporation. His research interests focus on deep learning perception algorithms for autonomous driving applications, generative models, data synthesis, reinforcement learning, and 3D reconstruction algorithms.

Yang Liu holds the M.S. degree in traffic safety engineering from Beijing Jiaotong University and obtained the M.S. degree in management from the Renmin Business School in 2016 while working. He currently serves as the Deputy General Manager of the Industrial Investment Department at COSCO Shipping Development Company. His research focuses include transportation industry investment, smart logistics, intelligent scheduling, and traffic information engineering. He has participated in 863 national projects, as well as multiple key initiatives of the Ministry of Transport and the Ministry of Railways, and has published numerous papers in related fields.

Kun Gao received the B.S. and Ph.D. degrees from Tongji University. He is currently a Docent (Associate Professor) at Chalmers University of Technology. His research focuses on smart and sustainable transport systems, including electrification, shared mobility, connected automation, and their analysis using machine learning and big data. He has led research projects in these areas and has published in top-tier transportation and interdisciplinary journals.