ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XI-2-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission II, 4–11 July 2026, Toronto, Canada

Sensor Domain Adaptation for 3D Object Detection via LiDAR Super-Resolution

June Moh Goo*, Zichao Zeng, Jan Boehm

Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT UK - (june.goo.21, zichao.zeng.21, j.boehm)@ucl.ac.uk

Keywords: LiDAR Super-Resolution, Cross-Sensor Domain, 3D Object Detection, Autonomous Driving, Domain Adaptation

Abstract

LiDAR-based perception models' performance can degrade sharply when applied to data from sensors different to those they were trained on. LiDAR super-resolution aims to enhance sparse point clouds from low-cost sensors. This can help to bridge the sensor domain gap to higher resolution LiDAR. Prior work has primarily focused on reconstruction quality metrics for super-resolution with limited evaluation of downstream perception tasks. We address this gap by conducting a systematic analysis of how super-resolution quality impacts 3D object detection performance. We evaluate detection capability through zero-shot transfer experiments on the KITTI object dataset. Four representative detectors (SECOND, PointPillars, PV-RCNN, PointRCNN) trained on high-resolution data are directly applied to super-resolved low-resolution data without fine-tuning. Results reveal a critical insight: reconstruction improvements yield vastly different detection gains across architectures. PointPillars shows minimal improvement until reaching high reconstruction quality, then performance improves significantly. In contrast, PV-RCNN exhibits steady gains throughout. The highest-quality reconstruction closes up to 86% of the performance gap and enables detection in safety-critical scenarios, including distant vehicles and small pedestrians, where lower-quality methods fail entirely. This work establishes that LiDAR super-resolution effectiveness depends on both reconstruction quality and detector architecture.

1. Introduction

LiDAR has become one of the key sensors in autonomous driving (Abbasi et al., 2022, Karangwa et al., 2023), cultural heritage (Matrone et al., 2024, Pierdicca et al., 2020), forestry (Böhm et al., 2016, Dai et al., 2018), urban planning (Griffiths and Boehm, 2019, Sun et al., 2025), and robotics (Chen et al., 2025). It provides reliable 3D structural information of the environment. Tasks such as 3D object detection (Lang et al., 2019, Shi et al., 2020, Shi et al., 2019, Yan et al., 2018, Goo et al., 2025a) and scene understanding (Goo et al., 2024) rely heavily on these spatial cues. High-resolution LiDARs, for example 64-beam sensors, produce dense point clouds and lead to strong perception performance. However, they remain expensive and are difficult to deploy at scale. Low-resolution LiDARs, such as 16 or 32 beam models, are cheaper and easier to use (Ortiz Arteaga et al., 2019). Yet their sparse measurements inevitably reduce perception accuracy.

Deploying high-resolution LiDAR at scale is not always feasible. Autonomous vehicles, delivery robots, and mobile platforms face cost, weight, and power constraints (Ortiz Arteaga et al., 2019, Glennie and Hartzell, 2020). As a result, low-resolution sensors remain attractive in practice. The downside is that perception accuracy can drop in safety-critical scenarios. For example, a model trained on a dense 64-beam sensor may fail to detect small or distant objects when tested on sparse 16-beam data (Goo et al., 2026). This motivates the need for approaches that reduce the gap between sensors without requiring extensive retraining.

LiDAR-based perception models, such as 3D object detection, often achieve strong performance when evaluated on data collected from the same sensor used during training. However, their performance can degrade sharply when applied to data from different sensors (see Figure 1(a)). This performance drop, often called the cross-sensor domain gap, is a practical obstacle for robust LiDAR perception (Fang et al., 2024, Goo et al., 2026). This issue is mainly due to discrepancies in sensor resolution and vertical field of view. Such differences introduce fundamental structural mismatches, rather than simple distribution shifts. If low-resolution LiDAR data can be transformed into a representation that resembles high-resolution scans, the cross-sensor gap can be reduced. In that case, models trained on high-resolution data could be effectively reused across a wider range of sensor configurations.

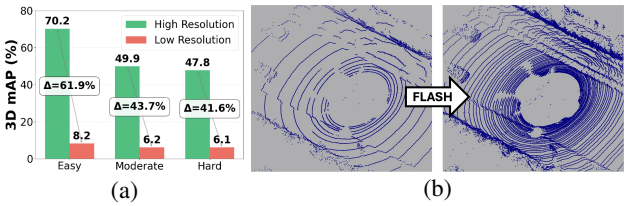

Figure 1. Cross-sensor performance degradation in 3D object detection and the effectiveness of LiDAR super-resolution. (a) Detection mAP drops drastically on the KITTI dataset across all difficulty levels with sparse point clouds. (b) Example of a super-resolved dense point cloud reconstructed by FLASH

Several approaches have been proposed to address this issue. Domain adaptation methods align feature distributions between source and target domains (Yi et al., 2021), while domain generalization builds models for unseen domains without target data (Kong et al., 2025, Kim et al., 2023). However, these methods do not directly address structural gaps such as differences in LiDAR resolution. It remains difficult to overcome performance loss caused by physical differences between sensors. Recently, LiDAR super-resolution has been introduced as another direction. It seeks to upsample sparse LiDAR scans into denser forms. Most of this work has focused on reconstruction quality or visual fidelity. The effect on downstream perception, espe-

* Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-93-2026

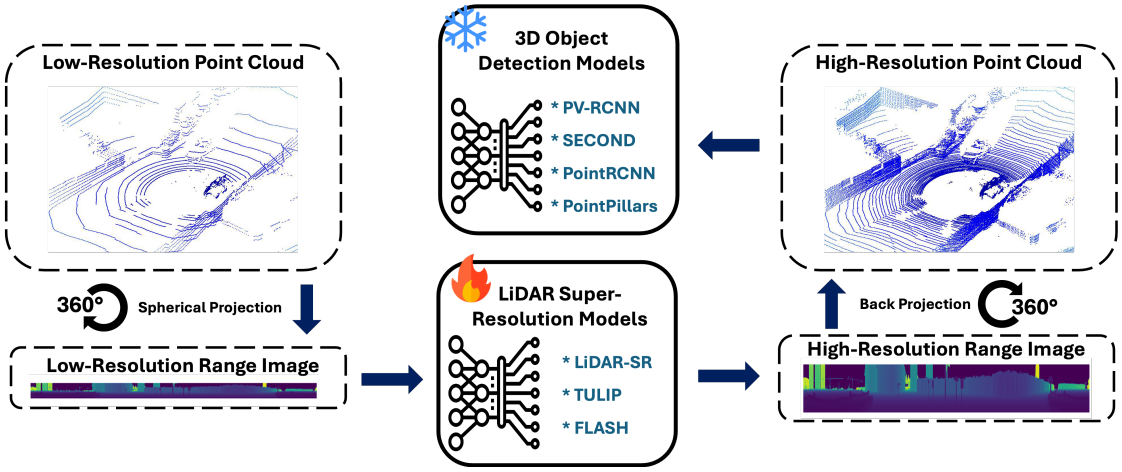

Figure 2. Cross-sensor evaluation pipeline. Low-resolution range images are upsampled through super-resolution models and back-projected into 3D point clouds for evaluation by pre-trained object detectors. The fire icon denotes training; the snowflake icon denotes no fine-tuning (pre-trained on high-resolution LiDAR data – KITTI object dataset).

cially for cross-sensor detection, has not been fully investigated.

In this paper we explore how 3D object detection performance is affected by existing LiDAR super-resolution approaches. LiDAR super-resolution can make low-cost sensors more compatible with models trained on high-resolution data (see Figure 1(b)). We focus on the 16-to-64 beam case and systematically evaluate its effect on 3D object detection. Notably, we uncover a non-linear relationship between reconstruction quality and detection performance. Different detector models have varying sensitivity to super-resolution improvements. The main contributions of this paper are:

- **Downstream-oriented evaluation of LiDAR super-resolution:** systematically compare how multiple super-resolution methods serve as sensor domain adaptation for zero-shot cross-sensor 3D object detection
- **Evaluation protocol:** establish a systematic benchmark by evaluating state-of-the-art detection models on original, artificially downsampled, and super-resolved point clouds for fair cross-sensor comparison
- **Experimental results:** all evaluated super-resolution methods improve 3D object detection, with gains scaling with reconstruction quality
- **Non-linear correlation analysis:** we reveal that reconstruction quality improvements do not translate linearly to detection performance, with detector models showing different sensitivities

2. Related Works

2.1 LiDAR Perception Overview

LiDAR has become a central sensor in 3D perception research. It provides dense point clouds and enables accurate estimation of geometry and distance. As a result, LiDAR is widely used for tasks such as 3D object detection, tracking, and scene understanding. Recent models, including PointPillars (Lang et al., 2019), SECOND (Yan et al., 2018), PV-RCNN (Shi et al., 2020) and PointRCNN (Shi et al., 2019), have shown strong performance. However, most studies assume access to high-resolution sensors, and performance degrades significantly in low-resolution settings.

2.2 Cross-Sensor Problem & Domain Gap

LiDAR perception models achieve strong results when tested on data from the same sensor used in training. However, accuracy drops sharply when applied to data collected by different sensors. This degradation is caused by differences in resolution and field of view. To mitigate this, many studies have explored domain adaptation (Yi et al., 2021, Xu et al., 2025). These methods reduce the distribution mismatch between source and target domains. Yet they usually assume access to at least some target data and do not resolve structural sensor differences. Domain generalization takes another approach. It aims to build models that generalize to unseen domains without target data (Kong et al., 2025, Kim et al., 2023). Techniques such as data augmentation, meta-learning, and regularization have been explored. Still, they struggle to overcome the performance gap created by variations in beam number and field of view.

2.3 Point Cloud Upsampling & LiDAR Super-Resolution

Point cloud interpolation and upsampling have been studied for many years. Early methods relied on simple interpolation or geometric constraints to fill sparse point sets. More recently, deep learning approaches such as PU-Net (Yu et al., 2018) and PU-GAN (Li et al., 2019) have enabled more sophisticated reconstruction. However, these methods target point cloud densification and do not account for the structured scan patterns of LiDAR sensors.

Building on this, LiDAR super-resolution has been proposed to transform low-resolution scans into denser representations. LiDAR-SR (Shan et al., 2020) introduced one of the first deep learning frameworks for LiDAR super-resolution, where sparse scans are projected into range images and upsampled using CNN-based networks. More recently, TULIP (Yang et al., 2024) adapted the Swin Transformer (Liu et al., 2022) architecture for range image super-resolution. It introduced row-based patching and non-square-window attention, specifically designed for the extreme aspect ratios typical of LiDAR range images, achieving state-of-the-art reconstruction performance across multiple datasets. FLASH (Goo et al., 2025b) further extended this line of work by incorporating frequency-domain analysis into the attention mechanism and replacing standard skip connections with multi-scale fusion, reporting improved reconstruction fidelity on the KITTI dataset.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-93-2026 | © Author(s) 2026. CC BY 4.0 License.

These methods generate high-resolution point clouds from sparse input, and typically evaluate quality using metrics such as Chamfer Distance or Intersection over Union. However, most existing work has focused on reconstruction fidelity rather than downstream tasks. In particular, the question of whether super-resolution can effectively reduce the cross-sensor performance gap in 3D perception remains largely unanswered.

2.4 Research Gap

In summary, prior studies fall into two main directions. The first focuses on domain adaptation and generalization, aiming to reduce distribution differences. The second focuses on point cloud upsampling and super-resolution, aiming to improve density and reconstruction quality. However, neither direction directly addresses the structural gap created by differences in LiDAR resolution, nor does it evaluate the downstream impact on perception. This paper fills this gap. We evaluate LiDAR super-resolution as a strategy to reduce the cross-sensor performance drop and systematically validate its effectiveness through 3D object detection experiments.

3. Methodology

3.1 Problem Formulation

Given a low-resolution LiDAR point cloud $P_l \in \mathbf{R}^{n_l \times 3}$ with H_l vertical channels, we generate a high-resolution point cloud $P_h \in \mathbf{R}^{n_h \times 3}$ with $H_h = 4 \cdot H_l$ channels. We transform this 3D upsampling into 2D super-resolution by projecting the point cloud onto a range image $I_l \in \mathbf{R}^{H_l \times W}$ using spherical projection with logarithmic transformation $\log(r + 1)$.

LiDAR super-resolution models predict a high-resolution range image $I_h \in \mathbf{R}^{H_h \times W}$ from I_l, where $H_h = 4H_l$ for $4\times$ upsampling. Training minimizes L1 loss:

$$L = \|I_h - \hat{I}_h\|_1 \tag{1}$$

where $\hat{I}_h$ denotes the network prediction.

3.2 Cross-Sensor Evaluation Framework

Our evaluation framework simulates a practical deployment scenario: a perception system trained on high-resolution LiDAR data must operate with a lower-cost sensor. Figure 2 illustrates the overall pipeline. We construct a three-stage evaluation: (1) simulate a cross-sensor gap by downsampling high-resolution point clouds, (2) apply super-resolution methods to recover spatial density, and (3) evaluate pre-trained 3D object detectors on the resulting point clouds without any fine-tuning.

This zero-shot transfer setting is critical for practical relevance. In real-world deployments, retraining detectors for every new sensor configuration is infeasible. Instead, super-resolution acts as a sensor-agnostic preprocessing step that transforms low-resolution inputs into representations compatible with models trained on high-resolution data. By keeping the detectors frozen, we isolate the effect of reconstruction quality on detection performance.

3.3 LiDAR Super-Resolution Methods

We evaluate three representative LiDAR super-resolution approaches that span different architectural paradigms:

LiDAR-SR (Shan et al., 2020) employs a CNN-based U-Net to upsample range images. It introduces Monte Carlo Dropout during inference to filter uncertain predictions at object boundaries.

TULIP (Yang et al., 2024) adapts the Swin Transformer architecture for range image super-resolution. It uses row-based patching and non-square window attention designed specifically for the extreme aspect ratios of LiDAR range images.

FLASH (Goo et al., 2025b) extends TULIP with two components: Frequency-Aware Window Attention, which combines spatial attention with FFT-based global frequency analysis, and Adaptive Multi-Scale Fusion, which replaces standard skip connections with learned position-specific feature aggregation.

3.4 3D Object Detection Models

To assess how different detector architectures respond to super-resolved inputs, we select four models that represent distinct design paradigms for processing 3D point clouds:

PointPillars (Lang et al., 2019) discretises the point cloud into vertical pillars and encodes each with a simplified PointNet (Qi et al., 2017a). This pillar-based representation prioritises computational efficiency but depends on sufficient point density within each pillar for reliable feature extraction.

SECOND (Yan et al., 2018) converts the point cloud into a sparse 3D voxel grid and applies sparse convolutions. Its voxel-based approach balances efficiency and spatial resolution, but voxel occupancy is directly affected by input point density.

PointRCNN (Shi et al., 2019) operates directly on raw points without voxelisation, using PointNet++ (Qi et al., 2017b) for feature extraction followed by a two-stage refinement. This point-based design is sensitive to the spatial distribution and completeness of the input point cloud.

PV-RCNN (Shi et al., 2020) combines voxel-based and point-based representations through a keypoint aggregation module. By fusing both representations, it can leverage complementary information at multiple spatial scales, making it potentially more robust to variations in point density.

These four detectors are trained exclusively on high-resolution 64-beam point clouds. At inference, they receive inputs from different sources (low-resolution, super-resolved, or high-resolution) without any adaptation or fine-tuning.

4. Experimental Results

4.1 Experimental Settings

Dataset and Preprocessing. We conduct experiments on the KITTI dataset (Geiger et al., 2012), which contains LiDAR scans from a Velodyne HDL-64E sensor. We use 3,712 point clouds for training and evaluate on 3,769 validation samples with ground truth annotations. To simulate low-resolution sensors, we downsample the original 64-channel scans by $4\times$ to create 16-channel inputs. Following common practice in range image processing, we apply logarithmic scaling $\log(r + 1)$ to the range values, which helps stabilize training by compressing the distribution of distant points.

Training Details. All super-resolution models are trained on the same KITTI splits using L1 loss for fair comparison.

Method	MAE ↓	CD ↓	IoU ↑	F1 ↑
LiDAR-SR	0.7739	0.3226	0.2231	0.3521
TULIP	0.4357	0.1356	0.3862	0.5439
FLASH	0.3951	0.1121	0.4001	0.5682

Table 1. Overall reconstruction performance comparison with existing models on KITTI dataset

LiDAR-SR (Shan et al., 2020) uses the Adam optimiser with a learning rate of 10^{-4} and a decay factor of 10^{-5} per epoch. TULIP (Yang et al., 2024) and FLASH (Goo et al., 2025b) both adopt the AdamW optimiser with cosine annealing and warm restarts. For the 3D object detectors, we use the pre-trained weights provided by OpenPCDet (OpenPCDet Development Team, 2020), trained exclusively on high-resolution 64-beam KITTI data. No detector is fine-tuned on super-resolved outputs. All experiments are trained and tested on two NVIDIA A6000 GPUs (48GB).

Evaluation Metrics. We evaluate using complementary metrics that assess both reconstruction quality and downstream task performance. For super-resolution quality, we measure Mean Absolute Error (MAE) on range images, Chamfer Distance (CD) for 3D geometric fidelity, and Intersection over Union (IoU) computed on voxelized point clouds at 0.1 m resolution. We also report F1-score to quantify point cloud completeness. To evaluate practical utility, we measure 3D object detection performance using mean Average Precision (mAP_{3D}), which computes the average precision across object classes at an IoU threshold of 0.7 for bounding box overlap. To quantify how effectively super-resolution bridges the cross-sensor performance gap, we define the gap closure rate:

$$\text{Gap Closure} = \frac{\text{mAP}_{\text{SR}} - \text{mAP}_{\text{low}}}{\text{mAP}_{\text{high}} - \text{mAP}_{\text{low}}} \times 100\% \qquad (2)$$

where mAP_{SR}, mAP_{low}, and mAP_{high} denote the detection performance on the super-resolved, low-resolution, and high-resolution LiDAR point clouds, respectively.

4.2 LiDAR Super-Resolution Performance

Before analysing detection outcomes, we first compare the three super-resolution methods in terms of reconstruction quality, as this establishes the varying levels of point cloud fidelity that detectors will subsequently receive as input.

4.2.1 Quantitative Analysis Table 1 summarises reconstruction quality across the three super-resolution methods. FLASH achieves the lowest error across all metrics, followed by TULIP and LiDAR-SR. The gap between CNN-based LiDAR-SR and the transformer-based methods is substantial: FLASH reduces MAE by 51% and nearly doubles IoU relative to LiDAR-SR, while the margin between FLASH and TULIP is smaller but consistent (10% lower MAE, 13% lower Chamfer Distance). These reconstruction differences establish the basis for analysing how varying levels of point cloud fidelity translate to downstream detection performance in Section 4.3. Detailed architectural analysis and ablation studies are provided in (Goo et al., 2025b).

4.2.2 Qualitative Analysis Figure 3 illustrates reconstruction quality on three challenging KITTI scenarios, comparing super-resolution methods against high-resolution ground truth.

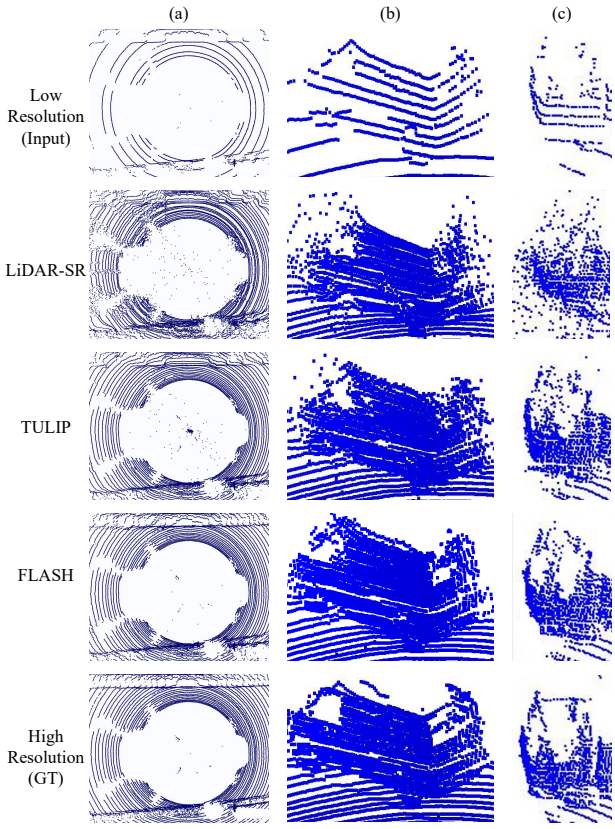

Figure 3. Comparison of reconstruction quality across super-resolution methods, showing differences in noise suppression (a), detail recovery (b), and edge preservation (c).

Figure 3 (a) examines noise handling near the sensor mounting area. While the original low-resolution input and ground truth show clean point clouds, super-resolution methods introduce different artifacts. LiDAR-SR generates substantial circular artifacts, TULIP reduces these but still produces scattered noise, while FLASH achieves the cleanest reconstruction closest to ground truth.

Figure 3 (b) focuses on a van’s rear window, where thin geometric structures challenge reconstruction methods. The window frame, barely visible in the low-resolution input, becomes clear in the ground truth. LiDAR-SR blurs this area entirely, losing frame definition. TULIP over-smooths the structure, merging the frame with surrounding surfaces. FLASH recovers the window geometry most accurately among the three methods, maintaining thin boundary definition that both LiDAR-SR and TULIP fail to preserve.

Figure 3 (c) evaluates edge preservation on large vehicles such as trucks, where sharp boundaries between vehicle and background are critical for detection. LiDAR-SR blurs vehicle edges substantially, creating ambiguous boundaries. TULIP tends to over-smooth vertical surfaces, blurring sharp boundaries. FLASH maintains sharp discontinuities.

Across all scenarios, FLASH produces reconstructions closest to high-resolution ground truth, avoiding noise generation while recovering fine details and preserving sharp boundaries. These qualitative observations align with the reconstruction metrics reported in Table 1.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-93-2026 | © Author(s) 2026. CC BY 4.0 License.

Models	Method	Easy		Moderate		Hard	
		mAP_{3D}	gap closure	mAP_{3D}	gap closure	mAP_{3D}	gap closure
PV-RCNN	High Resolution	68.21	-	49.83	-	45.03	-
	Low Resolution	14.33	-	12.69	-	11.60	-
	LiDAR-SR	21.60	+13.49%	15.91	+8.67%	13.17	+4.70%
	TULIP	55.08	+75.63%	39.00	+70.84%	34.93	+69.79%
	FLASH	**61.11**	**+ 86.82%**	**42.69**	**+ 80.78%**	**39.03**	**+ 82.05%**
SECOND	High Resolution	71.34	-	51.97	-	49.34	-
	Low Resolution	11.96	-	9.37	-	9.38	-
	LiDAR-SR	20.38	+14.18%	16.69	+17.18%	15.38	+15.01%
	TULIP	47.24	+59.41%	31.96	+53.03%	27.88	+46.30%
	FLASH	**56.98**	**+75.82%**	**40.26**	**+72.51%**	**30.22**	**+52.15%**
PointRCNN	High Resolution	70.16	-	49.88	-	47.76	-
	Low Resolution	8.23	-	6.17	-	6.14	-
	LiDAR-SR	14.81	+10.62%	12.71	+14.97%	12.49	+15.25%
	TULIP	41.26	+53.34%	30.51	+55.70%	25.98	+47.67%
	FLASH	**52.86**	**+72.06%**	**40.33**	**+78.15%**	**34.40**	**+67.90%**
PointPillars	High Resolution	65.97	-	49.50	-	44.59	-
	Low Resolution	12.33	-	10.53	-	9.35	-
	LiDAR-SR	20.11	+14.50%	15.86	+13.68%	14.42	+14.38%
	TULIP	42.29	+55.87%	30.15	+50.35%	25.53	+45.92%
	FLASH	**57.87**	**+84.90%**	**39.82**	**+75.17%**	**32.15**	**+64.71%**

Table 2. Comparison of 3D object detection performance (mAP) using different LiDAR super-resolution methods. All models are trained on high-resolution point clouds. At inference, predictions are obtained on low-resolution inputs with and without super-resolution methods (LiDAR-SR, TULIP, FLASH). Results are evaluated on KITTI validation set with four detectors (PV-RCNN, SECOND, PointRCNN, PointPillars) across three difficulty levels. The gap closure metric represents the recovery rate from low-resolution to high-resolution performance.

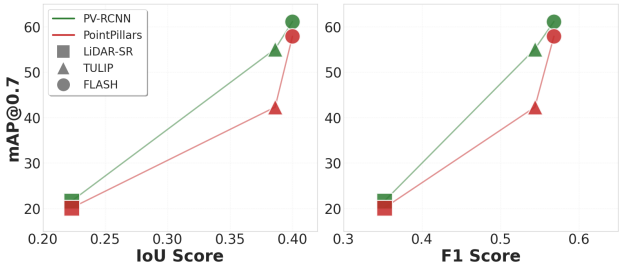

Figure 4. Reconstruction metrics versus detection performance for the two detectors showing the most distinct patterns: PV-RCNN (green) and PointPillars (red). Left: IoU vs mAP@0.7. Right: F1 vs mAP@0.7. PV-RCNN exhibits steady, near-linear improvement across super-resolution methods. PointPillars, however, shows accelerating performance gains as reconstruction quality improves.

4.3 3D Object Detection Performance

While reconstruction metrics measure point cloud quality, the ultimate value of LiDAR super-resolution lies in its impact on downstream perception tasks. To evaluate this practical utility, we assess detection performance in a zero-shot transfer setting that simulates real-world deployment scenarios. We train four representative 3D object detectors: SECOND, PointPillars, PV-RCNN, and PointRCNN, exclusively on high-resolution ground truth data from KITTI. These trained models are then directly applied to predict objects from various inputs: the original low-resolution scans, outputs from LiDAR-SR, TULIP, and FLASH, and high-resolution ground truth as an upper bound.

We evaluate detection across three object categories: Car, Pedestrian, and Cyclist, following KITTI’s standard difficulty levels (Easy, Moderate, Hard). These levels reflect increasing detection challenges based on object occlusion, truncation, and distance: Easy includes fully visible, nearby objects; Moderate extends to partially occluded cases; Hard encompasses heavily occluded or distant objects with sparse point coverage. We report Average Precision at 3D IoU threshold of 0.7 ($AP_{3D}@0.7$) across all difficulty levels.

4.3.1 Quantitative Analysis Table 2 presents detection performance across four architectures and three difficulty levels. The results validate our core hypothesis: LiDAR super-resolution effectively bridges the performance gap between low-cost and high-end sensors. FLASH achieves 52-86% gap closure toward high-resolution ground truth, followed by TULIP (45-75%) and LiDAR-SR (4-17%).

On Hard difficulty, where the super-resolution impact is most practically relevant, all detectors show substantial gains with higher-quality reconstruction. PV-RCNN improves from 11.60 to 39.03 AP (82.05% gap closure), SECOND from 9.38 to 30.22 (52.15%), PointRCNN from 6.14 to 34.40 (67.9%), and PointPillars from 9.35 to 32.15 (64.71%) when using FLASH outputs. These gains significantly exceed those of TULIP and LiDAR-SR across all detectors.

The performance advantage of higher-quality super-resolution becomes more pronounced as difficulty increases. On Easy

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-93-2026

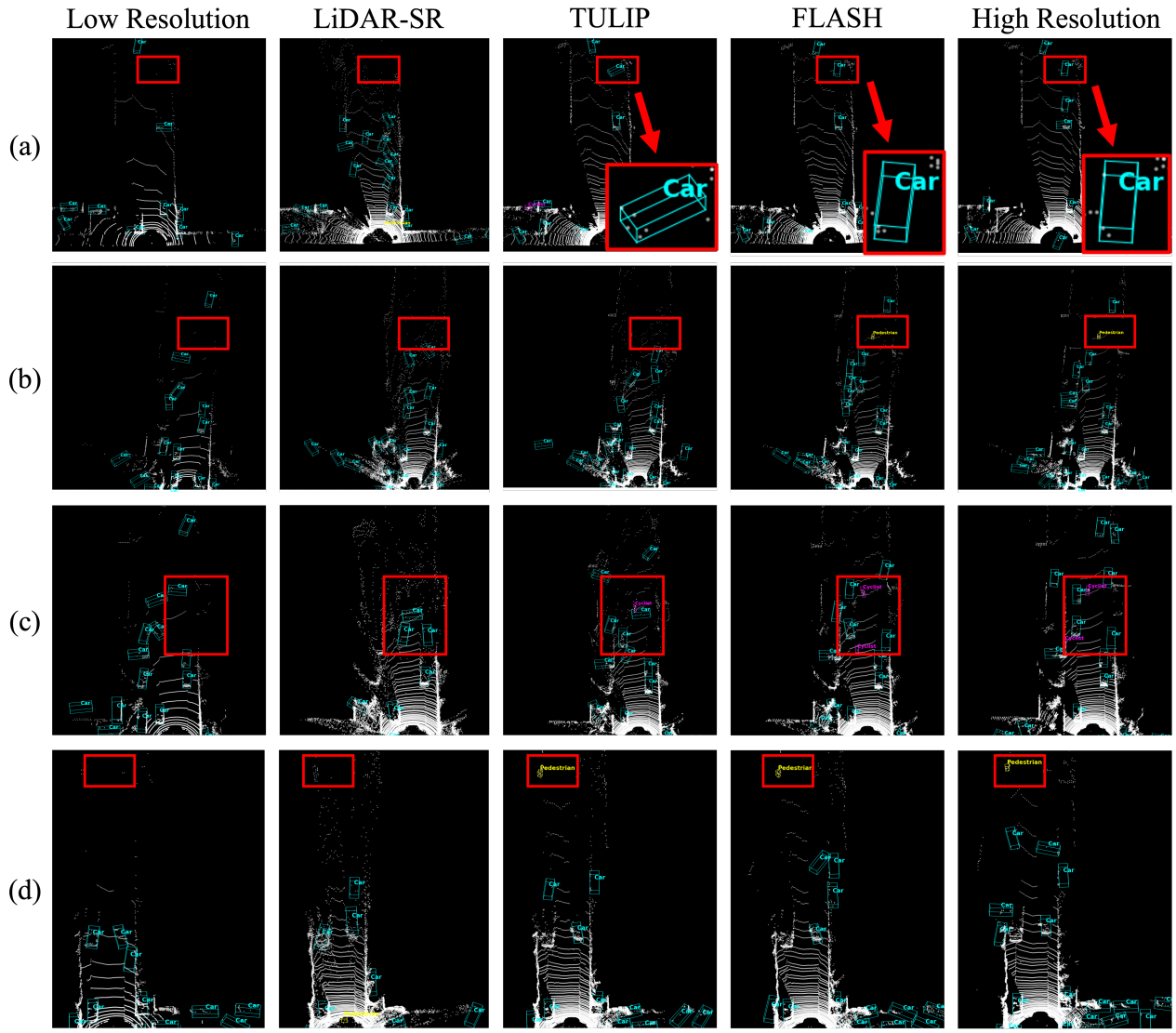

Figure 5. 3D object detection comparison across LiDAR super-resolution methods. Detection results on (a)-(d) four challenging scenes. Bounding boxes with class labels overlay the point clouds: Car, Pedestrian, Cyclist. Red boxes highlight regions where detection outcomes differ across methods.

tasks, both FLASH and TULIP achieve strong gap closure (86.82% vs 75.63% for PV-RCNN). However, on Hard difficulty, the margin widens: FLASH maintains 82.05% while TULIP drops to 69.79%. Notably, only the highest-quality reconstruction consistently achieves over 50% gap closure across all detectors on Hard difficulty. LiDAR-SR remains below 18% regardless of difficulty, suggesting that CNN-based reconstruction does not reach the fidelity threshold needed for meaningful detection improvement.

Different detector architectures exhibit distinct sensitivity patterns to reconstruction quality. Figure 4 illustrates this relationship for the two most contrasting cases. PV-RCNN scales near-linearly with reconstruction improvements, benefiting incrementally at every quality level due to its hybrid point-voxel feature extraction. In contrast, PointPillars shows minimal improvement with LiDAR-SR and TULIP outputs, then gains sharply with FLASH. This threshold behaviour reflects the pillar-based encoding’s dependence on minimum point density: below a critical density, pillars contain too few points for reliable feature extraction, but once this threshold is crossed, detection performance improves rapidly. These non-linear patterns confirm that the relationship between reconstruction quality and detection performance is architecture-dependent, and that evaluation on a single detector would miss this important characteristic.

4.3.2 Qualitative Analysis Figure 5 illustrates detection performance across four challenging scenarios, visualised using PV-RCNN to show how reconstruction quality translates to detection capability.

Figure 5 (a) examines distant car detection. The vehicle appears in the far background with minimal point coverage in the low-resolution input. Neither the low-resolution input nor the LiDAR-SR output provides sufficient geometric information for detection. TULIP successfully detects the vehicle but produces a misaligned bounding box with an incorrect heading angle. FLASH detects the car with pose accuracy matching the high-resolution ground truth, suggesting that higher reconstruction fidelity preserves the geometric structure needed for accurate localisation at range.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-93-2026 | © Author(s) 2026. CC BY 4.0 License.

Figure 5 (b) and (d) focus on pedestrian detection, the most safety-critical scenario in autonomous driving. Pedestrians present small targets with sparse point coverage, making them exceptionally challenging on super-resolved data. In both scenes, low-resolution input, LiDAR-SR, and TULIP all fail to detect the pedestrian, while FLASH and high-resolution ground truth succeed. The consistency across both scenes indicates that fine geometric detail preservation is essential for small object detection, and that a quality gap exists between TULIP-level and FLASH-level reconstruction for this task.

Figure 5 (c) presents two partially occluded cyclists, combining small size with complex non-convex geometry. Low-resolution input and LiDAR-SR detect neither cyclist. TULIP finds one but misses the second, suggesting that moderate reconstruction quality is insufficient to separate closely spaced instances. FLASH detects both, matching high-resolution performance. This result indicates that boundary-preserving reconstruction is critical for distinguishing adjacent objects.

Across all scenarios, detection capability correlates directly with reconstruction quality. The qualitative examples reinforce the quantitative findings in Table 2: the highest-fidelity reconstruction consistently enables detection in cases where lower-quality methods fail, particularly for distant vehicles, small pedestrians, and occluded cyclists that are most relevant to autonomous driving safety.

5. Limitations

This study has several limitations. First, all evaluated super-resolution methods operate on range images with fixed resolution ($H \times W$), which constrains the maximum point density that can be reconstructed. Methods that directly predict 3D points could potentially overcome this grid-based ceiling, but are not yet competitive for LiDAR super-resolution at this scale.

Second, the super-resolved point clouds contain only geometric information (xyz coordinates) without intensity values. Since several detectors leverage intensity for classification, particularly for distinguishing pedestrians from background clutter, the absence of this channel may underestimate the full potential of super-resolution as a domain adaptation strategy.

Third, our cross-sensor gap is simulated through uniform downsampling of 64-beam data to 16 beams. Real low-resolution sensors differ not only in beam count but also in vertical field of view, beam spacing, and noise characteristics. Validation on physically distinct sensor pairs would strengthen the generalisability of our findings.

6. Conclusion

We presented a systematic evaluation of how LiDAR super-resolution quality impacts 3D object detection in a cross-sensor transfer setting. By applying four representative detectors to super-resolved point clouds without fine-tuning, we isolated the relationship between reconstruction fidelity and downstream perception performance.

Our analysis reveals three key findings. First, super-resolution effectively bridges the cross-sensor domain gap, with the best-performing method closing up to 86% of the performance gap relative to high-resolution ground truth. Second, reconstruction improvements do not translate uniformly to detection gains: detector architectures exhibit distinct sensitivity patterns, ranging from near-linear scaling to threshold-like behaviour depending on their point cloud encoding strategy. Third, the advantage of higher-quality super-resolution becomes more pronounced at harder difficulty levels, where occlusion and distance compound the challenges of sparse input.

These findings suggest that strategic pairing of super-resolution methods and detector architectures is essential for deploying low-cost LiDAR sensors in safety-critical applications. Future work should extend this evaluation framework to additional perception tasks such as semantic segmentation and multi-object tracking, and validate the observed patterns on real cross-sensor data beyond simulated downsampling.

Acknowledgements

J. M. Goo and Z. Zeng are supported by the Engineering and Physical Sciences Research Council through an industrial CASE studentship with Ordnance Survey (Grant number EP/X524840/1 and EP/W522077/1).

References

Abbasi, R., Bashir, A. K., Alyamani, H. J., Amin, F., Doh, J., Chen, J., 2022. Lidar point cloud compression, processing and learning for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 24(1), 962–979.

Böhm, J., Bredif, M., Gierlinger, T., Krämer, M., Lindenberg, R., Liu, K., Michel, F., Sirmacek, B., 2016. The IQmulus urban showcase: Automatic tree classification and identification in huge mobile mapping point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 301–307.

Chen, C., Xu, Y., Zhao, L., Li, Y., Sun, S., Li, L., Wang, Z., Yang, B., 2025. LUOJIA Explorer: A Multi-robot Autonomous Exploration and Mapping System for Unknown Spaces. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 269–274.

Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. *ISPRS journal of photogrammetry and remote sensing*, 144, 400–411.

Fang, J., Zhou, D., Zhao, J., Wu, C., Tang, C., Xu, C.-Z., Zhang, L., 2024. Lidar-cs dataset: Lidar point cloud dataset with cross-sensors for 3d object detection. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 14822–14829.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 3354–3361.

Glennie, C., Hartzell, P., 2020. Accuracy assessment and calibration of low-cost autonomous lidar sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 371–376.

Goo, J. M., Li, J., Wicaksono, D., Boehm, J., 2025a. Checkerboard Target Measurement in Unordered Point Clouds with Coloured ICP. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-G-2025, 529–534. http://dx.doi.org/10.5194/isprs-archives-XLVIII-G-2025-529-2025.

Goo, J. M., Zeng, Z., Boehm, J., 2024. Zero-Shot Detection of Buildings in Mobile LiDAR using Language Vision Model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 107–113.

Goo, J. M., Zeng, Z., Boehm, J., 2025b. Real-Time LiDAR Super-Resolution via Frequency-Aware Multi-Scale Fusion. *arXiv preprint arXiv:2511.07377*.

Goo, J. M., Zeng, Z., Boehm, J., 2026. A Comprehensive Survey on Deep Learning-Based LiDAR Super-Resolution for Autonomous Driving. *arXiv preprint arXiv:2602.15904*.

Griffiths, D., Boehm, J., 2019. Improving public data for building segmentation from Convolutional Neural Networks (CNNs) for fused airborne lidar and image data using active contours. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 70–83.

Karangwa, J., Liu, J., Zeng, Z., 2023. Vehicle detection for autonomous driving: A review of algorithms and datasets. *IEEE Transactions on Intelligent Transportation Systems*, 24(11), 11568–11594.

Kim, H., Kang, Y., Oh, C., Yoon, K.-J., 2023. Single domain generalization for lidar semantic segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17587–17598.

Kong, W., Zeng, Z., Wen, D., Wei, J., Peng, K., Goo, J. M., Boehm, J., Stiefelhagen, R., 2025. Exploring Single Domain Generalization of LiDAR-based Semantic Segmentation under Imperfect Labels. *arXiv preprint arXiv:2510.09035*.

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O., 2019. Pointpillars: Fast encoders for object detection from point clouds. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12697–12705.

Li, R., Li, X., Fu, C.-W., Cohen-Or, D., Heng, P.-A., 2019. Pu-gan: a point cloud upsampling adversarial network. *IEEE International Conference on Computer Vision (ICCV)*.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L. et al., 2022. Swin transformer v2: Scaling up capacity and resolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12009–12019.

Matrone, F., Gallitto, F., Lingua, A. M., Maschio, P. F., 2024. Exploitation of the Number of Return Echoes for DTM Extraction from Point Clouds Acquired by LiDAR UAS DJI Zenmuse L1. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 233–240.

OpenPCDet Development Team, 2020. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet.

Ortiz Arteaga, A., Scott, D., Boehm, J., 2019. Initial investigation of a low-cost automotive lidar system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 233–240.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., Frontoni, E., Lingua, A. M., 2020. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing*, 12(6), 1005.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Shan, T., Wang, J., Chen, F., Szenher, P., Englot, B., 2020. Simulation-based lidar super-resolution for ground vehicles. *Robotics and Autonomous Systems*, 134, 103647.

Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., Li, H., 2020. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10529–10538.

Shi, S., Wang, X., Li, H., 2019. Pointrcnn: 3d object proposal generation and detection from point cloud. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 770–779.

Sun, S., Chen, C., Yang, B., Xu, Y., Zhao, L., He, Y., Jin, A., Li, L., 2025. ALM-LED: autonomous LiDAR mapping in underground space with Luojia explorer anti-collision drone. *ISPRS Journal of Photogrammetry and Remote Sensing*, 230, 346–373.

Xu, J., Yang, W., Kong, L., Liu, Y., Zhou, Q., Zhang, R., Li, Z., Chen, W.-M., Fei, B., 2025. Visual foundation models boost cross-modal unsupervised domain adaptation for 3d semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*.

Yan, Y., Mao, Y., Li, B., 2018. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10), 3337.

Yang, B., Pfreundschuh, P., Siegwart, R., Hutter, M., Moghadam, P., Patil, V., 2024. Tulip: Transformer for upsampling of lidar point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15354–15364.

Yi, L., Gong, B., Funkhouser, T., 2021. Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15363–15373.

Yu, L., Li, X., Fu, C.-W., Cohen-Or, D., Heng, P.-A., 2018. Pu-net: Point cloud upsampling network. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.