# Feedback-Driven Learning for 3D Object Detection and Completion with Limited Supervision

Ananthu Ajith
Department of
Computer Science
Pondicherry University
Puducherry, India

## ABSTRACT
Object detection and completion in 3D have shown considerable potential in the field of autonomous vehicles, robots, and mixed reality, but it heavily depends on dense 3D annotations, which is difficult to scale. In this work, the authors have proposed a feedback-driven learning framework for 3D perception with limited supervision. It generates pseudo-labels using weak supervision and refines them using a closed-loop process, which includes the use of vision foundation models along with geometry-aware constraints. This has shown considerable potential in the field of 3D object detection and completion in a scalable and robust manner.

## General Terms
Weakly Supervised Learning, 3D Computer Vision

## Keywords
3D Object Detection, Geometry-Aware Learning, Weak Supervision, Vision Foundation Models, Multi-View Perception, Semantic–Geometric Alignment

## 1. INTRODUCTION
Three-dimensional object detection is a fundamental problem in autonomous driving, robotics, and mixed reality. However, existing methods, such as LiDAR-based SECOND and CenterPoint [1], heavily depend on dense 3D annotations, which are costly to obtain. Recently, weakly supervised and geometry-aware 3D object detection methods have been proposed to mitigate the annotation dependency. In particular, PETR [2], BEVFormer [3], PETRv2 [4] have demonstrated the effectiveness of 3D reasoning using transformers, while Back to Reality [5], GGA [6], WI3D [7] have shown the feasibility of 2D information in 3D object detection. Furthermore, WI3D [7] and FOMO-3D [8] have successfully incorporated semantics into 3D object detection. However, existing methods are limited by pseudo-label noise, a lack of semantic-geometry alignment, and the absence of an iterative refinement mechanism. To address the above limitations, this paper propose a novel feedback-driven learning approach to refine pseudo-labels in a closed-loop manner, which can improve both the accuracy of localization and semantics.

## 2. BACKGROUND AND RELATED CONCEPTS
Three-dimensional object detection methods can be broadly understood through the lens of input modality, supervision level, and guidance priors, each of which shapes the way a system in-terprets spatial information. Early approaches relied exclusively on LiDAR point clouds, exploiting their geometric precision but lacking sufficient semantic cues, as seen in classical LiDAR-only pipelines [1]. Later hybrid systems such as FOMO-3D [8] ad-dressed this limitation by fusing LiDAR geometry with 2D image semantics. In contrast, transformer-based detectors such as PETR

[2] and BEVFormer [3] demonstrated that even pure multi-camera image setups could achieve robust 3D reasoning when equipped with large receptive fields and temporal cue integration. In paral-lel, research diverged along supervision levels. Fully supervised methods delivered strong accuracy but demanded dense 3D anno-tations, motivating the rise of weakly supervised techniques, which instead rely on 2D bounding boxes [5], [6], weak depth priors, or pseudo-label generation from auxiliary modalities. For exam-ple, WI3D [7] leverages 2D vision foundation models to infer 3D structure through lifting and refinement, while pseudo-supervised architectures such as FOMO-3D [8] integrate pretrained 2D se-mantic embeddings to generalise beyond limited 3D labels. Un-supervised and self-supervised strategies—such as MAL-UPC [9] and USSPA [10]—shift focus toward reconstruction and comple-tion objectives to extract 3D structure without any paired labels. Orthogonal to modality and supervision is the notion of guidance priors, which encode domain knowledge into the detection process. GGA [6] employs explicit geometric ratios derived from category-level language priors, whereas BR [5] supplements weak super-vision with synthetic shape repositories. Meanwhile, symmetry-driven methods such as USSPA [10] and SymmCompletion [11] incorporate structural regularities to stabilise 3D predictions. Sys-tems like WI3D [7] and FOMO-3D [8] further extend these priors by leveraging pretrained 2D foundation models to inject high-level semantics into 3D reasoning. Together, this taxonomy establishes a conceptual foundation for understanding modern 3D perception pipelines and directly informs the need for hybrid frameworks that intelligently integrate multi-modal cues, structured priors, and se-mantic verification—thus setting the stage for the literature review that follows, in which each of the twelve surveyed works is exam-ined through this lens.

## 3. LITERATURE REVIEW
Recent developments in 3D object detection have pursued multiple methodological directions to address the challenges of limited 3D annotations, sparse sensor data, and the complex nature of real-world scenarios. The existing contributions to this literature differ substantially along dimensions of input modalities, training super-vision, and the priors included within the learning process. The collected body of literature examined within this review investi-gates collectively four primary directions in research as shown in Figure 2. 1) weakly supervised 3D detection approaches that mit-igate the annotation cost associated with 3D object detection, ei-ther through the use of 2D labels or pseudo-labels ([5], [6], [7]);

2) unsupervised and self-supervised learning-based 3D completion frameworks which learn the geometric structure of 3D objects with-out any paired annotations ([9], [10], [11]); 3) transformer-based multi-view camera detectors that use attention mechanisms to rea-son across images ([2], [3], [4], [8]); 4) LiDAR or range-view-centric representations of the environment for advanced geomet-ric processing in efficient and

real-time multi-sensor scenarios ([1], [12]). Collectively, these four noteworthy contributions mark the field's transition away from traditional fully supervised pipelines toward hybrid, geometry-aware, and multimodal perception sys-tems. The subsequent sections examine the contributions of this literature in detail, addressing the innovative elements of each ap-proach, its limitations or potential issues, and its relevance to build-ing a more unified semantic–geometric refinement framework for weakly supervised 3D object detection.

Table 1 illustrates the most popular learning paradigms for 3D per-ception, which are grouped according to the level of supervision for learning 3D representations. The taxonomy identifies four ma-jor trends in 3D perception: weakly supervised 3D object detec-tion, unsupervised and self-supervised 3D completion, multi-view transformer-based 3D detection, and range-view learning with Li-DAR sensors. The differences among these paradigms lie in the level of supervision, input information, and underlying principles. The weakly supervised methods minimize the annotation burden by relying on 2D information or pseudo-annotations, whereas unsu-pervised methods take advantage of inherent geometric constraints without relying on any annotations. On the other hand, transformer-based and range-view-based methods increase detection accuracy at the cost of higher annotation burden and complexity.

This taxonomy will give the field a categorized overview, which will be the basis for the comparison analysis in the next section.

## 3.1 Weakly Supervised 3D Object Detection

Zhang et al. [6] introduced a general geometry-aware approach that learned 3D bounding boxes from only 2D annotations, leveraging high-level geometric priors. The model incorporates automatically acquired, category-specific width-to-height ratios into the detec-tion backbone (CenterPoint for outdoors and FCAF3D for indoors). Training optimises a joint loss combining 2D projection consis-tency with 3D spatial alignment. On SUN RGB-D and ScanNet, GGA achieved more than 90% of fully supervised accuracy with only 2D labels. The limitations observed in domain transfer are such that the estimates from one environment, say indoors, would not generalize well to outdoors.A representative geometry-aware weakly supervised detection pipeline is illustrated in Figure 1.

Li et al. [7] introduced WI3D, a weakly incremental learning paradigm that expands a pretrained 3D detector to new object cat-egories without additional 3D boxes. A dual-teacher refinement mechanism performs intra-modal self-distillation and cross-modal 2D–3D consistency alignment. According to the experiments on SUN RGB-D and ScanNet, the approach showed significant recall gains on novel categories with minimal forgetting on base classes. However, the accuracy of this framework would depend upon the quality of pseudo-labels and whether the camera-depth alignment is well-calibrated.

Xu et al. [5] approached weak supervision via synthetic scene gen-eration: BR synthesizes dense labels by assembling CAD models into "virtual" scenes given coarse center annotations and adapts back to real data. The method has achieved close-to-supervised ac-curacy on ScanNet by consuming less than 5 percent of the manual labelling cost. Its main limitation is its reliance on a comprehen-sive shape repository. Poor diversity in synthetic shapes degrades real-world performance significantly.

## 3.2 Unsupervised and Self-Supervised 3D Completion

Wu et al. [9] developed an adversarial multi-view learning frame-work for unsupervised 3D point-cloud completion. The system projects an incomplete input from multiple virtual camera view-points and trains a generator–discriminator pair to match rendered and observed depth maps. The resulting completion network re-constructs missing regions without paired ground-truth supervi-sion. MAL-UPC outperformed earlier self-reconstruction methods in terms of Chamfer Distance and F-score metrics on ShapeNet and KITTI-360 datasets. However, adversarial instability occasion-ally caused over-smoothed geometry in results, while the model re-quired substantial memory due to its multi-view rendering pipeline. Ma et al. [10] presented USSPA, an unsupervised shape-preserving autoencoder leveraging reflection symmetry as a self-supervisory signal. By enforcing mirror consistency on the latent features of partial point clouds, the model learns to reconstruct missing re-gions that are coherent with symmetric priors. Compared to the variational baselines, USSPA achieved increased symmetry fidelity on ShapeNet-Part and ScanNet. The method is constrained to ob-jects with explicit or near-explicit bilateral symmetries, while non-symmetric structures like plants or irregular furniture remain chal-lenging for reconstruction. Building on USSPA, Chen et al. [11] extended symmetry-aware reconstruction to unpaired real-scene data. SymmCompletion introduces an attention-weighted symme-try module to dynamically detect local reflection planes, instead of assuming global symmetry. The network jointly predicts plane pa-rameters and reconstructs geometry via a cross-entropy consistency loss. Experiments on PCN and KITTI datasets showed more realis-tic surface recovery, though at higher computational cost. This ap-proach shows that symmetry is still a powerful inductive bias even for unsupervised 3D learning.

## 3.3 Multi-View Transformer-Based Detection

Liu et al. [2] pioneered PETR, the first transformer architecture that performs 3D detection directly from synchronized multi-view im-ages. It encodes 2D pixel coordinates as 3D position embeddings that enable cross-view reasoning in one unified feature space. This framework eliminates any explicit depth prediction or cost vol-umes, thus reducing latency compared to stereo-based methods. When evaluated on nuScenes, it gives an mAP of 38.1 with six-camera inputs. Its main limitation is weak temporal modelling: each frame is processed independently, limiting motion under-standing.

Li et al. [3] extended multi-view reasoning to the temporal dimen-sion with BEVFormer, which builds a dynamic BEV representation from sequential images. The model applies deformable attention to aggregate both spatial and temporal cues, achieving effective tracking of objects across frames.

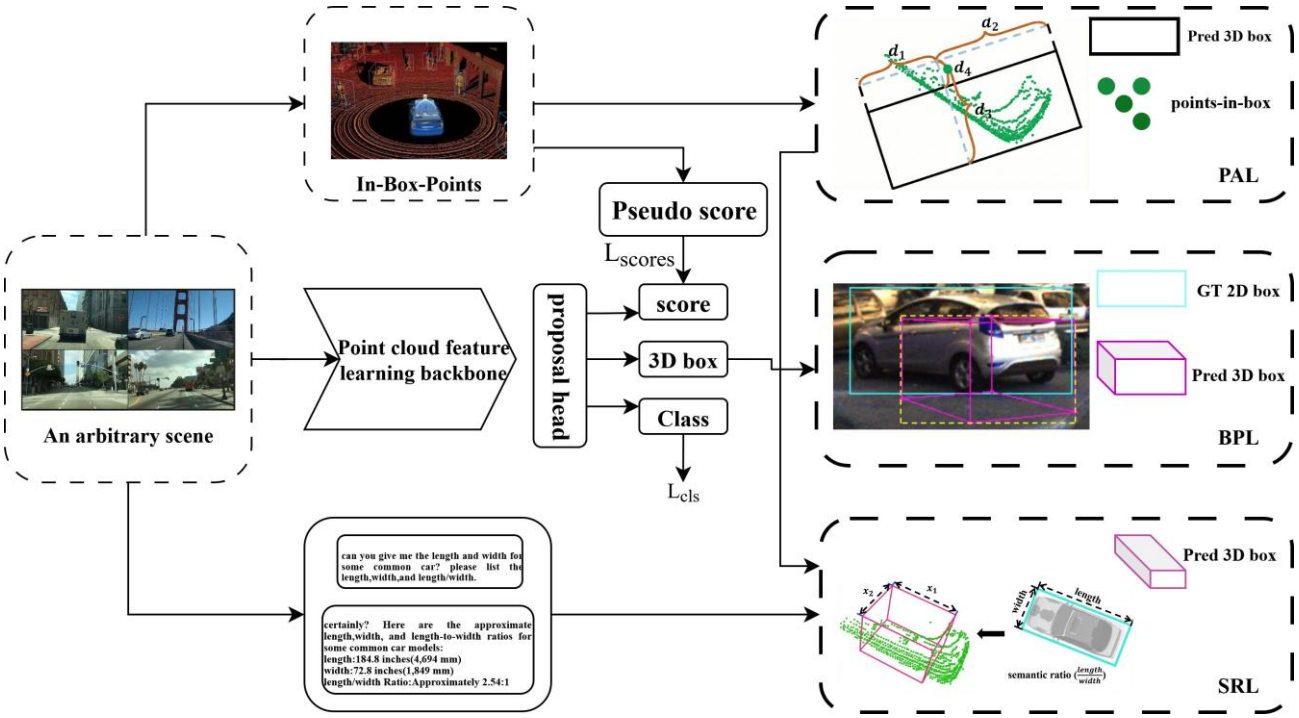


**Fig. 1. Conceptual illustration of a geometry-aware weakly supervised 3D object detection framework, showing how 2D supervision and geometric priors guide 3D prediction.**

**Table 1. Taxonomy of learning paradigms in recent 3D object detection literature**

| Category | Supervision | Primary Modality | Core Idea and Limitations |
|---|---|---|---|
| Weakly supervised 3D detection | Weak | RGB-D / RGB | Learns 3D bounding boxes using 2D annotations, geometric priors, or pseudo-labels; reduces annotation cost but is sensitive to domain shift and pseudo-label noise. |
| Unsupervised and self-supervised 3D completion | None | Point clouds | Reconstructs complete object geometry using structural priors such as symmetry or adversarial consistency; achieves strong geometric fidelity but lacks semantic discrimination. |
| Multi-view transformer-based detection | Full / Pseudo | Multi-view RGB | Uses attention mechanisms to aggregate information across synchronized camera views and time; provides high accuracy but requires heavy computation and accurate calibration. |
| LiDAR and range-view learning | Full | LiDAR | Projects point clouds into range-view representations for efficient processing; enables real-time inference but suffers from boundary distortion and sparsity effects. |

This has achieved state-of-the-art per-formance on nuScenes and Waymo datasets, with real-time infer-ence. However, the strong transformer backbone requires substan-tial GPU resources and is sensitive to camera calibration drift. Liu et al. [4] proposed PETRv2 as a generalization of PETR for uni-fying the three tasks of detection, depth estimation, and tracking. Combining BEV queries with camera intrinsic priors in PETRv2 embeds geometric constraints into transformer attention layers. Ex-periments demonstrate that the method outperforms BEVFormer

[3] in terms of depth accuracy and tracking continuity on nuScenes.

However, it still poses limitations in that its heavy reliance on dense multi-camera input and high computational cost render deploy-ment on embedded platforms challenging. Yang et al. [8] proposed FOMO-3D, a model that incorporates transformer-based detection with open-vocabulary recognition. This model can distill semantic knowledge from 2D vision foundation models like CLIP into a 3D detection backbone, allowing for recognition of rare and unseen categories. The experiments on nuScenes-LT and Waymo-Open-LT datasets have shown superior recall for tail classes while being competitive for frequent objects. Its main limitation is the depen dence on the quality of text-image alignment from the foundation model, which may not transfer precisely to 3D geometry.

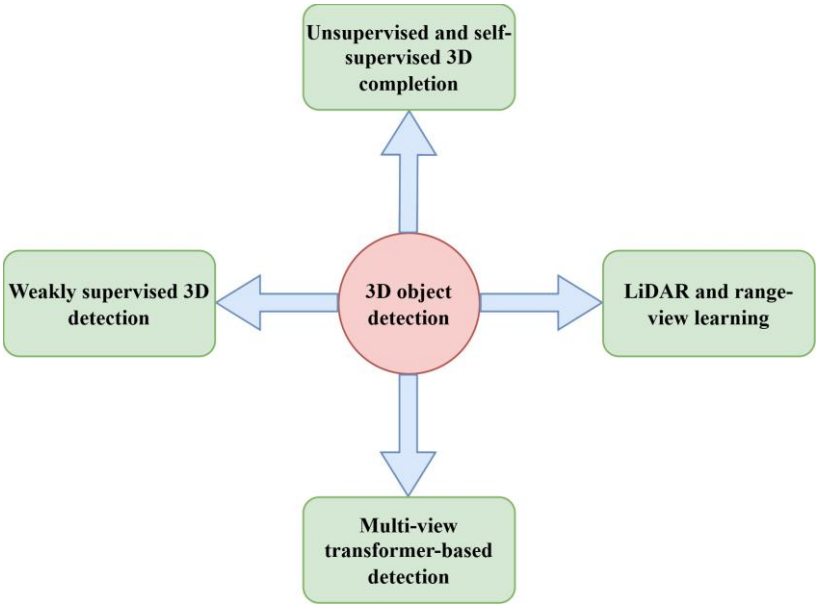


**Fig. 2. Taxonomy of recent 3D object detection and completion methods categorized primarily by the level of supervision.**

**Table 2. Comparative summary of surveyed 3D detection and completion methods**

| Method | Year | Task Type | Supervision Level | Input Modality | Key Contribution | Main Limitation |
|---|---|---|---|---|---|---|
| GGA [6] | 2024 | 3D Detection | Weakly supervised | RGB-D | Incorporates category-level geo-metric priors to infer 3D bounding boxes from 2D annotations | Sensitive to domain bias and geo-metric prior mismatch |
| WI3D [7] | 2025 | 3D Detection | Weakly incremental | RGB-D | Enables category expansion using vision foundation models without additional 3D labels | Performance depends on pseudo-label quality |
| BR [5] | 2022 | 3D Detection | Synthetic supervision | RGB-D | Generates dense pseudo-labels us-ing synthetic scene reconstruction from CAD models | Limited by synthetic shape diver-sity |
| PETR [2] | 2022 | 3D Detection | Fully supervised | Multi-view RGB | Introduces position embedding transformation for camera-only 3D detection | Lacks explicit temporal modeling |
| BEVFormer [3] | 2022 | 3D Detection | Fully supervised | Multi-view RGB | Builds spatio-temporal BEV repre-sentations using deformable atten-tion | High computational cost |
| PETRv2 [4] | 2023 | Unified Perception | Fully supervised | Multi-view RGB | Unifies detection, tracking, and depth estimation with geometric priors | Requires dense multi-camera input |
| FOMO-3D [8] | 2025 | 3D Detection | Weakly supervised | RGB + LiDAR | Transfers open-vocabulary se-mantic knowledge from vision-language models to 3D detection | Semantic misalignment across modalities |
| MAL-UPC [9] | 2025 | 3D Completion | Unsupervised | Point clouds | Learns shape completion through multi-view adversarial reconstruc-tion | Training instability |
| USSPA [10] | 2023 | 3D Completion | Unsupervised | Point clouds | Exploits symmetry as a self-supervisory signal for geometric completion | Limited to symmetric objects |

### 3.4 LiDAR and Range-View Learning

Wilson et al. [12] emphasized the role of range-view representations in LiDAR-based 3D detection. They comprehensively examined the impact of resolution, field-of-view, and feature aggregation on detection efficacy. In addition, they proposed a simple range-view backbone with a hierarchical attention design which reduced the distortion induced by the spherical projection. The range-view backbone was tested on nuScenes and SemanticKITTI and demonstrated competitive accuracy with substantially lower latency when compared to voxel-based detectors. In conclusion, the study suggested that although range-view modelling is very viable for real-time systems, depth discontinuities along object boundaries led to sparsity. In both the transformer and range-view worlds, recent studies show a common trend towards sensor-agnostic representations that integrate both 2D and 3D reasoning. With transformer work (PETR [2], BEVFormer [3], PETRv2 [4]) leading advances in camera-only work, range-viewwork (Wilson et al. [12]) has the potential to lead the way with light-weight approaches for LiDAR data. On the periphery, weakly supervised detectors (GGA [6], WI3D [7]) will enhance the efficiency of labelled data. The field is iterating to hybrid, geometry-aware architecture designs that combine prior knowledge, self-supervision and foundation-model semantics.

## 4. COMPARATIVE ANALYSIS

A comparative summary of various 3D object detection and completion approaches is provided in Table 2. Weakly supervised approaches are found to reduce annotation costs by utilizing 2D labels, geometric information, or pseudo-labels. The performance of weakly supervised approaches is found to be vulnerable to domain shifts and pseudo-label noise.

Transformer-based approaches have been found to attain high detection accuracy by utilizing multiple camera views for spatial and temporal information aggregation. The approaches require dense supervision and are computationally expensive. Unsupervised and self-supervised object completion approaches have been found to attain high geometric fidelity by utilizing geometric prior information. The approaches are found to be less effective for object detection tasks as they are unable to attain semantic understanding.

LiDAR and range view-based approaches have been found to attain efficient point cloud processing and attain real-time inference. The approaches are found to have information loss at object boundaries and are less effective for small objects and distant objects.

A trade-off between annotation efficiency, computational complexity, and detection robustness is found to be a major issue for various object detection approaches.

## 5. RESEARCH GAP

While recent progress in 3D object detection and completion approaches has greatly diminished the reliance on dense 3D annotation, a substantial gap still exists between the efficiency of annotation, the reliability of geometry, and the accuracy of semantics. The weakly supervised methods, i.e., GGA, WI3D, and Back to Reality, have shown promising performance in greatly reducing the annotation burden with the help of 2D information and pseudo-labels, but they are highly vulnerable to noise, domain shifts, and cross-modal mismatches. On the other hand, unsupervised and self-supervised methods have shown promising performance in learning geometry, but they lack semantic understanding, which greatly limits their application to category-aware object detection tasks. The transformer-based multi-view approaches achieve state-of-the-art performance with the help of heavy computation, while range view approaches sacrifice geometry for efficiency. However, none of the current approaches guarantee reliable 3D localization, semantic validation, and scalability with limited supervision.This is where a significant research gap lies: there is no unified framework with feedback mechanisms for iteratively improving pseudo labels through geometric constraints and semantic verification. Closing this gap involves integrating geometric information with semantic feedback from foundation models in a closed-loop fashion for im-proved accuracy and generalization in weakly supervised 3D object detection.

## 6. PROPOSED METHOD

This work proposeS a framework for feedback-driven learning in the context of weakly supervised 3D object detection and comple-tion, which jointly considers geometric reasoning and semantic val-idation in a closed-loop refinement process.

This framework consists of a coarse pseudo-label generation, which is performed by leveraging information provided by weak supervision, where 2D object proposals and depth information are utilized for lifting regions of images into 3D space.

For the second part of the framework, a geometry refinement mod-ule is proposed, which imposes spatial consistency by consider-ing constraints such as projection alignment, point in box, and category-specific shape priors, thereby enhancing the structural plausibility of predicted 3D bounding boxes.

For semantic validation, a lightweight vision foundation model is proposed, which considers image regions corresponding to pre-dicted 3D boxes for evaluating semantic alignment between the predicted object categories and the visual appearance of the region in images.

The main contribution of the proposed framework is the iterative interaction between geometric refinement and semantic validation. The framework has a closed-loop architecture, which means that the refined predictions are re-evaluated and updated in multiple it-erations, thereby improving the quality of pseudo-labels and reduc-ing the error accumulation.

This feedback mechanism allows the framework to optimize geo-metric and semantic accuracy jointly, which is a scalable and robust solution for 3D perception in a weakly supervised setting.

## 7. EXPERIMENTAL SETUP

The proposed framework is tested on standard indoor and outdoor 3D perception benchmarks to evaluate its performance with limited supervision. Indoor benchmarking is done on the SUN RGB-D and ScanNet datasets. The proposed framework is tested on the KITTI dataset for outdoor benchmarking.

For training the proposed framework, weak supervision is used in the form of 2D bounding boxes and image data. Coarse pseudo-labels are used in the proposed framework with the help of 2D-to-3D lifting techniques. These pseudo-labels are refined using the proposed feedback mechanism.

For evaluating the proposed framework, standard detection metrics such as mean Average Precision (mAP) are used. It is used to evaluate the accuracy of the proposed framework in localizing the objects in the 3D environment. The performance of the proposed framework is also evaluated in terms of the refinement of pseudo-labels.

The proposed framework is implemented using a modular frame-work consisting of point cloud feature extraction, geometry-aware refinement, and semantic validation. The proposed framework is designed in such a manner that it is scalable and can be adapted to other datasets and modalities.

## 8. RESULTS AND EVALUATION

The performance of the suggested feedback-driven framework is assessed by comparing it with various weakly

supervised, transformer-based, and unsupervised 3D perception approaches on various benchmark tasks. The performance is assessed in terms of detection robustness, pseudo-label refinement, and semantic-geometric consistency.

A qualitative comparison of the suggested framework with various approaches is provided in Table 3, where various aspects are

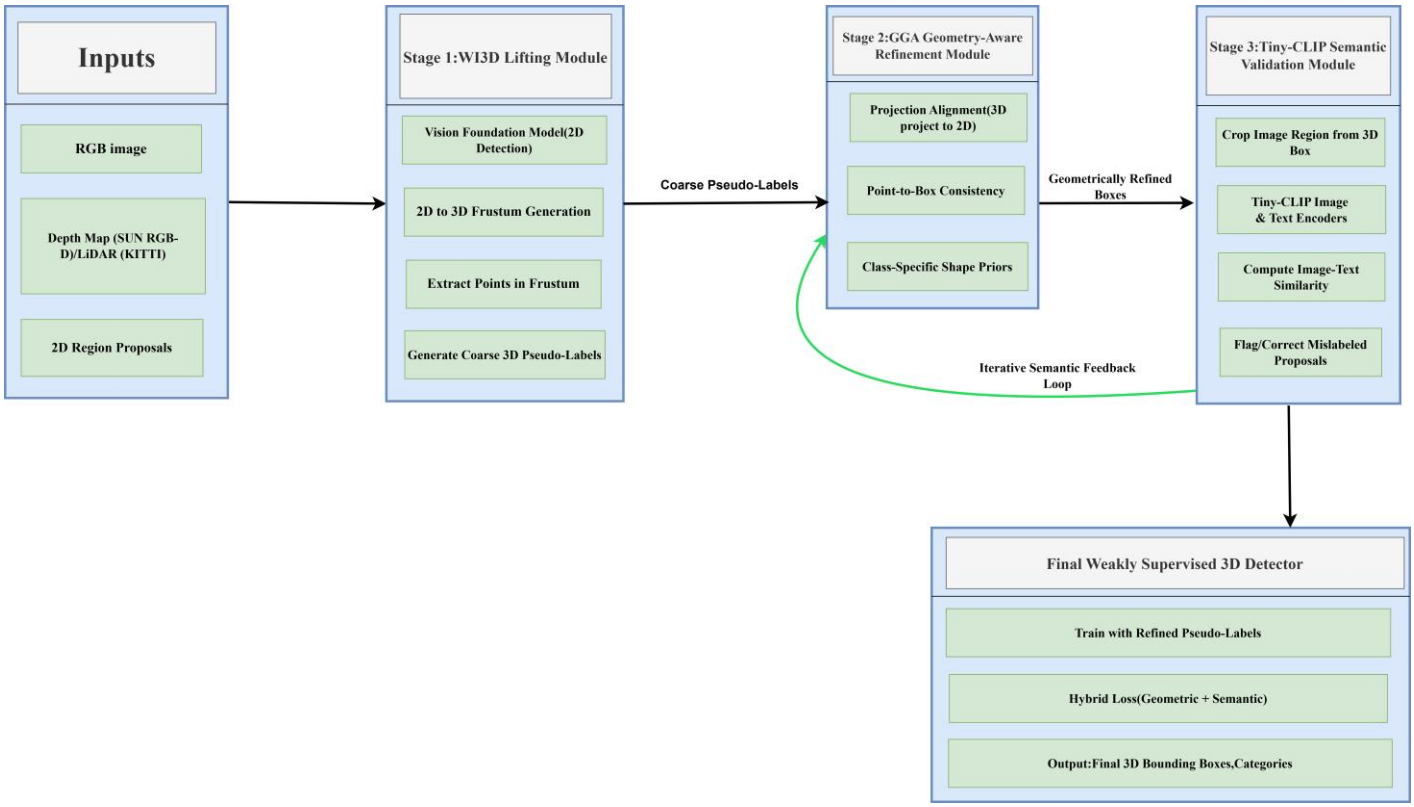


**Fig. 3. Conceptual illustration of a geometry-aware weakly supervised framework highlighting the interaction between coarse 3D hypotheses derived from limited 2D supervision and iterative refinement through geometric consistency and semantic feedback mechanisms.**

compared. The suggested framework is compared with various approaches in terms of supervision, pseudo-label, semantic, and robustness. The suggested framework is different from various approaches in that it considers geometry refinement and semantic validation. The suggested framework is compared with weakly supervised approaches such as GGA [6], WI3D [7], and Back to Reality [5]. The suggested framework is better than various approaches in terms of robustness with noisy supervision. The suggested framework is different from various approaches in that it considers pseudo-label refinement. The suggested framework is compared with transformer-based approaches such as PETR [2], BEVFormer [3], and PETRv2 [4]. The suggested framework is better than various approaches in terms of performance with reduced supervision dependency.

This method IS compared with other unsupervised methods, such as MAL-UPC [9], USSPA [10], and SymmCompletion [11], this method performs better in terms of semantic consistency and geometry preservation. Unsupervised methods rely on reconstruction, and they do not have a category-level understanding. However, this framework has incorporated semantic validation.

Further qualitative analysis of the results shows that the ini-tial pseudo labels have localization and classification errors, but through the feedback refinement process, the results improve.

Ablation analysis of the results shows that geometry refinement improves the accuracy of the structure, and semantic validation improves the accuracy of classification. However, when both geometry refinement and semantic validation are combined, the results improve.

Overall, it is noted that the suggested framework offers a good trade-off between efficiency, detection accuracy, and robustness, making it appropriate for large-scale 3D perception with limited supervision.

## 9. CONCLUSION AND FUTURE WORK

In this work, a feedback-based learning paradigm is proposed for 3D object detection and completion in a weakly supervised setting. It is demonstrated that the proposed approach can overcome the limitations of existing weakly supervised 3D object detection meth-ods, such as pseudo-label noise and semantic-geometric misalign-ment, by incorporating geometry-aware refinement and semantic validation in a closed-loop paradigm.

It is envisioned that the proposed 3D weakly supervised object detection and completion paradigm will pave the way towards a promising research direction in which 3D structure understanding can benefit from the effectiveness of combining structural con-straints with foundation model-based feedback to enhance both ac-curacy and category-level consistency. Future work will concen-trate on generalizing the proposed paradigm to dynamic scenes, exploring more advanced foundation models, and improving the robustness of the proposed paradigm to domain shift. Furthermore, unifying 3D object detection and completion in a single paradigm is a promising research direction.

**Table 3. Qualitative comparison of the proposed framework with existing methods**

| Method | Supervision | Pseudo-Label Handling | Semantic Integration | Robustness |
|---|---|---|---|---|
| GGA [6] | Weak | Static refinement | Limited | Moderate |
| WI3D [7] | Weak | Teacher-based refinement | Strong | Moderate |
| BR [5] | Weak | Synthetic labels | Limited | Moderate |
| PETR [2] | Full | Not applicable | Moderate | High |
| BEVFormer [3] | Full | Not applicable | Moderate | High |
| PETRv2 [4] | Full | Not applicable | Moderate | High |
| MAL-UPC [9] | None | Not applicable | None | Low |
| USSPA [10] | None | Not applicable | None | Low |
| SymmCompletion [11] | None | Not applicable | None | Moderate |
| **Proposed** | Weak | Iterative feedback refinement | Strong | High |

# 10. REFERENCES

[1] Y. Yan *et al.*, "SECOND: Sparsely Embedded Convolutional Detection," *Sensors*, 2018.

[2] Y. Liu *et al.*, "PETR: Position Embedding Transformation for Multi-View 3D Object Detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[3] Z. Li *et al.*, "BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatio-Temporal Transformers," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[4] Y. Liu *et al.*, "PETRv2: A Unified Framework for 3D Perception from Multi-Camera Images," *arXiv preprint arXiv:2301.02604*, 2023.

[5] X. Xu *et al.*, "Back to Reality: Weakly-Supervised 3D Object Detection with Shape-Guided Label Enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022

[6] G. Zhang *et al.*, "General Geometry-Aware Weakly Supervised 3D Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[7] M. Li *et al.*, "WI3D: Weakly Incremental 3D Detection via Vision Foundation Models," *IEEE Transactions on Multimedia*, 2025.

[8] A. Yang *et al.*, "FOMO-3D: Using Vision Foundation Models for Long-Tailed 3D Object Detection," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

[9] L. Wu *et al.*, "Unsupervised 3D Point Cloud Completion via Multi-View Adversarial Learning," *IEEE Transactions on Visualization and Computer Graphics*, 2025.

[10] H. Ma *et al.*, "USSPA: Unsupervised Symmetric Shape-Preserving Autoencoder for 3D Point-Cloud Completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. .

[11] S. Chen *et al.*, "SymmCompletion: Symmetry-Guided Point-Cloud Completion," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

[12] B. Wilson *et al.*, "What Matters in Range-View 3D Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[13] A. Lang et al., "PointPillars: Fast Encoders for Object Detection from Point Clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[14] S. Shi et al., "PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-tion (CVPR)*, 2019.

[15] S. Shi et al., "PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[16] Y. Yin et al., "CenterPoint: Center-Based 3D Object Detection and Tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[17] C. R. Qi et al., "VoteNet: A Deep Learning Framework for 3D Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[18] P. Philion and S. Fidler, "Lift, Splat, Shoot: Encoding Images from Arbitrary Camera Rigs by Implicitly Unprojecting to 3D," in *Proceedings of the European Conference on Com-puter Vision (ECCV)*, 2020.

[19] Y. Huang et al., "BEVDet: High-Performance Multi-Camera 3D Object Detection in Bird's-Eye View," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[20] H. Li et al., "BEVDepth: Acquisition of Reliable Bird's-Eye-View Representations via Depth Estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[21] Z. Li et al., "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified BEV Representation," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[22] Z. Hou et al., "Multi-View 3D Object Detection with Sparse Supervision," in *Proceedings of the IEEE International Con-ference on Computer Vision (ICCV)*, 2021.

[23] W. Yuan et al., "PCN: Point Completion Network," in *Proceedings of the International Conference on 3D Vision (3DV)*, 2018.

[24] Y. Yang et al., "FoldingNet: Point Cloud Auto-Encoder via Deep Grid Deformation," in *Proceedings of the IEEE Confer-ence on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[25] H. Xie et al., "GRNet: Gridding Residual Network for Dense Point Cloud Completion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[26] X. Yu et al., "PoinTr: Diverse Point Cloud Completion

with Geometry-Aware Transformers," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[27] X. Yu et al., "AdaPoinTr: Adaptive Geometry-Aware Point Cloud Completion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[28] L. Tchapmi et al., "TopNet: Structural Point Cloud Decoder," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[29] Y. Huang et al., "PF-Net: Point Fractal Network for 3D Point Cloud Completion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[30] S. Xiang et al., "SnowflakeNet: Point Cloud Completion by Snowflake Point Deconvolution," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[31] Z. Zhou et al., "SeedFormer: Patch Seeds Based Point Cloud Completion with Upsample Transformer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recog-nition (CVPR)*, 2023.

[32] Z. Chen et al., "PointLDM: Latent Diffusion Models for Point Cloud Completion," *IEEE Transactions on Multimedia*, 2024.

[33] A. Milioto et al., "RangeNet++: Fast and Accurate LiDAR Semantic Segmentation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[34] X. Zhu et al., "Cylindrical and Asymmetrical Convolution for LiDAR Segmentation," in *Proceedings of the IEEE Confer-ence on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[35] J. Kim et al., "Range-View Learning for LiDAR-Based 3D Perception," *IEEE Transactions on Intelligent Vehicles*, 2023.

[36] S. Patel et al., "Efficient Range-View Representation for Real-Time LiDAR-Based 3D Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-tion (CVPR)*, 2024.