International Journal of Computer Sciences and Engineering
Vol.13, Issue.11, pp.75-89, November 2025
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Survey Article

Knowledge Distillation: A Review of Methods and Applications in LiDAR Data

Razvan-Marian Avesalon[1], Radu Miron[2*]

[1]Siemens Industry Software, Brasov, Romania
[2]Alexandru Ioan Cuza" University, Faculty of Computer Science, Iasi, Romania

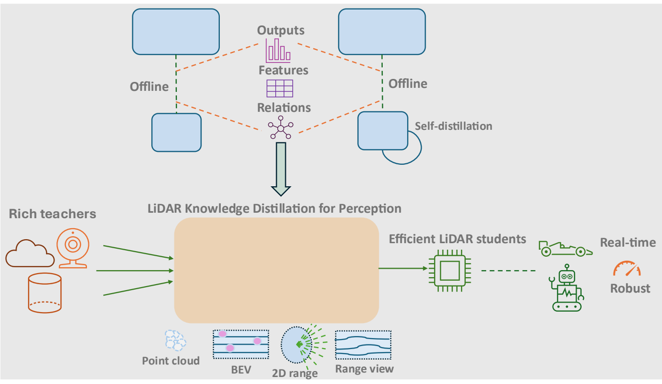
*Corresponding Author:

Received: 23/Sept/2025; **Accepted**: 25/Oct/2025; **Published**: 30/Nov/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i11.7589

Abstract: Knowledge distillation (KD) is a machine learning technique where a compact student model is trained by a larger teacher model to create efficient, high-performance models suitable for devices with limited computational resources. The student learns by mimicking the teacher's nuanced predictions, known as "soft targets", which provide a richer learning signal than traditional ground-truth labels. Methods are categorized by the source of knowledge—such as the teacher's final outputs (response-based), intermediate features (feature-based), or the relationships between data points (relation-based)—and by the training strategy, including offline, online, and self-distillation schemes. This review focuses on the application of KD to 2D and 3D LiDAR data for tasks like object detection and semantic segmentation in autonomous systems. In this domain, knowledge distillation is critical for developing lightweight models that can run in real-time, enabling cross-modal learning from expensive LiDAR to cheaper sensors, and addressing inherent challenges of point cloud data such as sparsity and sensor-specific domain gaps.

Keywords: Knowledge distillation, LiDAR, Autonomous Systems, Object Detection, Semantic Segmentation, Point Clouds

Graphical Abstract



1. Introduction

Knowledge Distillation (KD), also known as model distillation, is a sophisticated machine learning technique centered on transferring knowledge from a large, complex, and often cumbersome model, referred to as the teacher, to a smaller, more compact model, the student. The fundamental objective is to train this student model to not only perform a specific task but to mimic the behavior and replicate the high performance of the teacher, thereby creating a lightweight yet powerful model suitable for deployment in environments with constrained computational resources, such as mobile or edge devices. While frequently categorized as a form of model compression, knowledge distillation is fundamentally a knowledge transfer process. This distinction is critical, as traditional model compression techniques like pruning or quantization modify an existing model, whereas distillation involves training an entirely new student model from the ground up using the teacher as a guide. This paradigm allows for significant architectural flexibility, as the student and teacher do not need to be homogeneous. The core of the distillation process represents a paradigm shift from conventional supervised learning. Instead of learning only from ground-truth hard labels, the student model is trained to match the nuanced predictions of the teacher. This concept was formalized by [1], who introduced the idea of dark knowledge represented by the rich information encoded in the teacher's full probability distribution across all classes, not just the single correct answer. By learning from these soft targets, the student can learn a more generalized and robust function, as it captures how the teacher perceives similarities and relationships within the data space.

The initial concept, rooted in earlier work on model compression by [2], established response-based distillation, where the student mimics the teacher's final output layer. While simple and effective, this approach overlooks the wealth of information in a teacher's intermediate layers. This limitation led to the development of feature-based distillation, which transfers knowledge from the teacher's hidden layers. A pioneering method in this area was [3], which used hints from a teacher's intermediate feature maps to train thinner but deeper student networks. This demonstrated that guiding the student's feature extraction process directly, could significantly improve performance. Subsequent feature-based methods [3, 4] have explored distilling various aspects of feature maps, including attention maps, activation boundaries, and neuron selectivity patterns, to provide more comprehensive guidance to the student.

Further evolution led to relation-based distillation, which posits that the relationships between features or data samples constitute a higher form of knowledge. The work on Relational Knowledge Distillation [5] was instrumental, introducing distance-wise and angle-wise losses to compel the student to preserve the structural relationships between data points found in the teacher's embedding space. This approach proved highly effective, particularly for 2D metric learning tasks, where student models were sometimes able to outperform their teachers. Other methods have since explored transferring relational knowledge using graph-based representations or by modeling the similarity between feature maps.

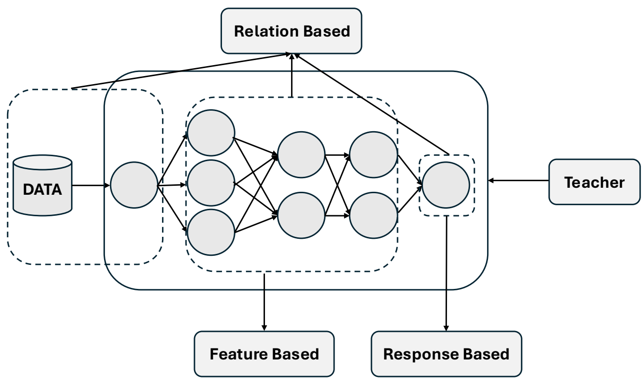


Figure 1. Knowledge sources in distillation. Response-based: the student mimics teacher soft targets at the output head. Feature-based: the student matches selected intermediate feature maps. Relation-based: the student preserves structural relations (e.g., affinities) computed from features across samples or regions.

This review examines the evolution of knowledge distillation techniques as applied to LiDAR data, from foundational concepts to domain-specific and cross-modal applications.

1.1 Contributions

This survey makes the following contributions:

1. We introduce a unified taxonomy for KD in LiDAR that jointly covers knowledge sources (response-based, feature-based, relation-based), training schemes (offline, online, self), data representations (point cloud, BEV, RV/projections), and modality paths (intra- vs cross-modal), and we map existing methods onto this space;
2. We provide the first focused synthesis of KD for 2D range (single-plane LiDAR) perception, summarizing architectures and datasets and proposing practical KD recipes tailored to this setting;
3. We analyze LiDAR-specific design patterns, challenges, and failure modes across representations—sparsity, occlusion, beam/domain gaps, compression, and weather;
4. We outline forward-looking directions, including foundation-model teachers, temporal/world-model distillation, robustness-aware and planning-aware KD.

Compared to prior general surveys on knowledge distillation [6], which are modality-agnostic and largely centered on 2D vision and classification, our contribution is domain-specific: we extend existing taxonomies with two LiDAR-centric axes: the data representation (point cloud, BEV, range view, 2D range) and the modality path (intra- vs cross-modal), and we systematically map LiDAR KD methods into this multi-dimensional design space. We provide, to our knowledge, the first focused synthesis of knowledge distillation for 2D range (single-plane LiDAR) perception, including detectors, datasets, and practical recipes. We analyze LiDAR-specific design patterns and failure modes—sparsity, beam/domain gaps, compression, and weather—and connect them to concrete KD mechanisms and engineering guidelines. We curate recent LiDAR-focused KD algorithms for detection, segmentation, and mapping across BEV, point-cloud, and projected representations, which are only marginally covered or not covered at all in general KD surveys.

1.2 Search Methodology

To identify the articles for this review, the research methodology centered on a systematic search of academic databases using structured queries. The search strategy was formulated to specifically target the intersection of sensor technology, a core machine learning technique, and key application areas. We detail here the methodology used to collect and organize the literature surveyed in this work. Our aim was to cover research at the intersection of knowledge distillation (KD) and LiDAR-based perception, including both intra-modal distillation and cross-modal transfer where LiDAR is either the teacher or the student modality.

To this end, we queried several major digital libraries and search engines—IEEE Xplore, ACM Digital Library, SpringerLink, Elsevier ScienceDirect, arXiv, and Google Scholar—for works published between January 2015 and March 2025. This time window spans the period after the popularization of modern KD for deep networks and the emergence of large-scale 3D LiDAR benchmarks. We used combinations of three groups of keywords:

1. sensor/data terms such as "LiDAR", "point cloud", "bird's-eye view", "range view", "2D range";
2. distillation terms such as "knowledge distillation", "model distillation", "teacher student"; and
3. task terms such as "3D object detection", "semantic segmentation", "mapping", and "autonomous driving".

Typical query patterns included, for example, "LiDAR AND knowledge distillation", "point cloud AND distillation", and "bird's-eye view AND knowledge distillation", with syntax adapted to each database (title/abstract search where supported). From the resulting set of candidate papers, we applied the following inclusion criteria. A work was retained if it

1. explicitly employed a teacher–student training scheme that can be interpreted as KD (offline, online, or self-distillation),
2. involved LiDAR or LiDAR-derived representations (native point clouds, BEV grids, range/front views, or raw 2D scans) either as the primary input to the student or as the target modality of cross-modal KD, and
3. addressed a perception or mapping task relevant to autonomous systems (e.g., 3D/BEV object detection, LiDAR semantic or instance segmentation, traversable region detection, or semantic mapping).

We excluded works that used LiDAR but did not apply any form of KD, general KD methods with no 3D or LiDAR component (except for a small number of foundational KD papers used for background), and non-archival or incomplete materials. Screening was performed in two stages: initial filtering based on title and abstract, followed by full-text inspection to verify the actual use of KD, the role of LiDAR, and the specific task. When multiple versions of the same work existed (e.g., arXiv preprint and later conference paper), we kept the most complete peer-reviewed version. Each selected paper was then manually categorized along the five axes of our unified taxonomy (Fig. 2): knowledge source (response-, feature-, or relation-based, or combinations thereof), distillation scheme (offline, online, self-distillation), data representation (point cloud, BEV, range/front view, or raw 2D range), modality path (intra-modal vs cross-modal), and task type (object detection, semantic segmentation, mapping, etc.). This classification was based on the methodological description and loss formulations in each paper and is summarized in our overview tables (Tables 1–7).

The remainder of this paper is organized as follows. Section 1 introduces knowledge distillation (KD), positions it within autonomous perception, states our contributions, and outlines the search methodology (Fig. 2). Section 2 reviews 2D range data and LiDAR representations—raw polar scans, Cartesian point clouds, bird's-eye view (BEV), and range view (RV)—including practical processing implications and key datasets (DROW, JRDB, FROG). Section 3 develops the taxonomy of knowledge sources—response-, feature-, and relation-based—anchored by an overview (Fig. 1), detailed schematics (Figs. 3–5), and a summary table (Table 2). Section 4 presents distillation schemes—offline, online, and self-distillation—with conceptual diagrams (Figs. 6–8) and a consolidated comparison (Table 3). Section 5 surveys algorithms and applications across representations: BEV (cross-/intra-modal transfer and alignment), native point clouds (local geometry/topology and robustness), and projected 2D views (range images), with organization by representation (Table 1) and modality path (Table 4). Section 6 concludes with a synthesis of challenges and forward-looking directions, including foundation-model teachers, temporal/world-model distillation, and robustness- and planning-aware KD.

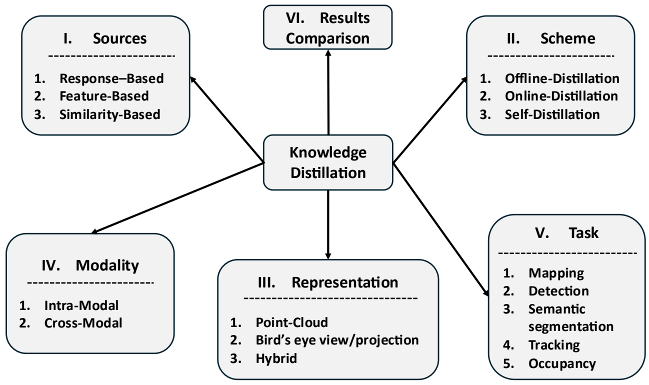


Figure 2. Unified taxonomy of knowledge distillation for LiDAR perception. Methods are organized along five axes: (1) knowledge source (response, feature, relation), (2) scheme (offline, online, self), (3) data representation (point cloud, BEV, range view), (4) modality path (intra-modal, cross-modal), and (5) task (object detection, semantic segmentation). Representative works are mapped to this space.

2. 2D Range Data

2D range data is a fundamental type of environmental perception data used extensively in robotics and autonomous systems. It consists of a set of distance measurements captured on a single, two-dimensional plane, effectively creating a 2D "slice" of the surrounding environment. This data is crucial for core robotic tasks such as obstacle detection, localization, and navigation.

The most common source of 2D range data is the 2D LiDAR (Light Detection and Ranging) sensor. This device operates by emitting a rotating laser beam. When the beam strikes an object, it reflects to the sensor. By measuring the time it takes for the light to travel to the object and return (time-of-flight), the sensor calculates the precise distance to that point. As the laser sweeps across a horizontal plane, it collects thousands of these distance measurements at discrete angular increments, typically generating a full 360-degree view [7].

The widespread adoption of 2D LiDAR sensors in robotics [8] is driven by a compelling blend of cost-effectiveness and robust performance. Their affordability compared to 3D counterparts has made them a standard, economical choice for a vast array of applications, from industrial automated guided vehicles (AGVs) to commercial and even hobbyist robots. In terms of performance, they deliver highly accurate distance measurements over long ranges, often proving superior to other low-cost sensors. A significant advantage over cameras is their lighting invariance; as active sensors providing their own illumination, they function reliably in conditions ranging from bright sunlight to complete darkness. Furthermore, a single 2D LiDAR can offer comprehensive 360-degree coverage, providing a wide field of view that would otherwise require a more complex and costly multi-camera setup to replicate. The way 2D range data is structured for processing by algorithms, especially deep learning models, is a critical consideration. The structure follows:

1. **Raw Data (Polar Coordinates):** In many robotics frameworks like the Robot Operating System (ROS), 2D LiDAR data is provided as a LaserScan message. This is essentially an array of range values (distances), where each

element's index corresponds to a specific angle. This native polar coordinate format (range, angle) is an ordered structure that some neural networks can process efficiently.

2. **Point Cloud (Cartesian Coordinates):** The raw polar data is often converted into a 2D point cloud using trigonometry to get a set of Cartesian coordinates. While geometrically intuitive, this representation is an unordered set of points. This irregularity and sparsity make it incompatible with standard Convolutional Neural Networks (CNNs) and require specialized architectures.
3. **Projected 2D Views:** To leverage the power and efficiency of 2D CNNs, a common strategy is to project the point cloud data into an image-like grid format. The two most prevalent projections are: Bird's-Eye-View (BEV), Range View (RV).

BEV is a top-down projection of the points onto a 2D grid. This view preserves the physical dimensions and spatial layout of objects on the ground, making it ideal for navigation and path planning. However, it can lose vertical information. RV is a projection onto a spherical or cylindrical plane, where the axes represent the laser's angles. This format is compact, dense, and preserves the sensor's perspective, including occlusion information. Its main drawback is the potential distortion of object shapes and sizes. Beyond just distance, many LiDAR sensors also return an intensity value for each point. This value measures the reflectivity of the surface the laser hit. For example, a reflective road sign will have a much higher intensity than dark asphalt. This intensity information provides powerful semantic cues that go beyond pure geometry. Including intensity as an additional data channel has been shown to be crucial for improving the performance of perception tasks like semantic segmentation, as it helps models distinguish between objects made of different materials. In (Table 1) we give an overview of existing work using specific data structures.

While 2D range data remains a cost-effective and widely used sensor modality for many mobile and industrial robots, it has been largely superseded by 3D LiDAR in the context of advanced autonomous driving research. This shift in focus is reflected in the available public datasets, with most large-scale benchmarks being built around 3D LiDAR sensors for complex driving scenarios [35]. Consequently, the body of research on knowledge distillation has predominantly concentrated on 3D point cloud data, whether for object detection or semantic segmentation. As a result, there is a comparative scarcity of academic papers that specifically address the application of knowledge distillation to native 2D range data. The recent introduction of what is described as the first public semantic dataset for 2D LiDAR underscores that this is an emerging area, suggesting that dedicated research into distilling knowledge for these more resource-constrained systems is still in its nascent stages.

Table 1. Classification of LiDAR data representation

BEV (Bird's-Eye-View)	Point Clouds	Raw 2D Range Data
[9] 3D-to-2D Distillation for Indoor Scene Parsing	[19] PointDistiller	[29] DR-SPAAM
[10] Lidar Distillation	[20] Topology Guided KD	[30] Li2Former
[11] DistillBEV	[21] Point-to-Voxel KD	[31, 32] DROW Dataset
[12] UniDistill	[22] Adversarial Learning on 3D Point Clouds	[33] JRDB Dataset
[13] BEV-LGKD	[23] Perturbed Self-Distillation	[34] FROG Dataset
[14] MMDistill	[24] Self-Distillation for Robust Lidar Segmentation	
[15] Distillation from 3D to BEV [16] SimDistill [17] TIGDistill-BEV [18] CRKD	[25] Image-to-Lidar Self-Supervised Distillation [26] Weak-to-Strong 3D Object Detection [27] Sunshine to Rainstorm: Cross Weather KD [28] PartDistill	

The field of 2D LiDAR data has seen a significant evolution, largely propelled by advancements in deep learning architectures and the availability of robust benchmarking datasets. The DR-SPAAM [29], which introduced an innovative spatial-attention and auto-regressive model to solve challenges in temporal data fusion, enables real-time performance on resource-constrained hardware. The latest advancement is represented by Li2Former [36], a Transformer-based architecture that achieves superior accuracy by modeling global context throughout the entire laser scan. The development and validation of these detectors have been made possible by a set of popular and increasingly challenging datasets. These include the original DROW dataset [31, 32], which catalyzed early research; the large-scale, multimodal JRDB (JackRabbot Dataset) [33], which offers immense complexity with its rich sensor data and annotations for actions and social groups and the FROG dataset [34], a modern benchmark specifically designed with dense annotations and crowded public scenarios to overcome the limitations of earlier datasets.

3. Source Theory

The paper [6] classifies these distillation sources into three well-established categories:

1. Response-based Distillation: This method, also known as logit-based knowledge distillation, uses the outputs of the model's last layer, known as logits, as the source of knowledge. The student model is trained to mimic these final predictions from the teacher.

2. Feature-based Distillation: This approach utilizes the representations from the intermediate layers of the teacher network as the source of knowledge, providing the student with step-by-step information that leads to the final prediction.
3. Relation-based Distillation: This form of distillation, newly named similarity-based knowledge distillation, transfers structural knowledge and relationships learned by the teacher, such as the pairwise similarities between features, instances, or classes, rather than the exact feature values themselves.

Knowledge distillation encompasses a variety of techniques for transferring knowledge from a large, complex teacher model to a smaller, more efficient student model. These methods can be broadly categorized based on the type of knowledge being transferred. The most straightforward approach is response-based distillation (logit-Based), where the student model is trained to mimic the final output predictions, or logits, of the teacher. This method leverages the teacher's softened probability distribution over the classes, which provides more information than just the ground-truth labels.

3.1 Response-Based

Modern response-based knowledge distillation techniques, as illustrated in (Figure 3) have evolved significantly beyond the foundational concept of mimicking a teacher model's final output distribution, or logits. Define teacher and student logits as z^T, z^S, temperature τ, soft target $p^S = \mathrm{softmax}(z^S/\tau)$. The student KD objective balances a hard and a soft term as:

$$L_{KD} = \lambda\tau^2 KL(p^T||p^S), \text{and}$$
$$L = \alpha L_{CE}(y, p^S) + (1-\alpha)L_{KD} \quad (1)$$

A prominent recent approach is the decoupling of the distillation loss, which most existing methods aim to do. This separates the knowledge from the target class and non-target classes, as proposed in methods like Decoupled Knowledge Distillation (DKD) [37], to improve student learning. DKD separates positive and negative examples, by adjusting the loss:

$$L^{DKD} = \lambda_{pos} \cdot KL(p_y^T||p_y^S) + \lambda_{neg}(\hat{p}_y^T||\hat{p}_y^S) \quad (2)$$

where p_y is the target-class probability and $\hat{p}_y$ sums over non-target classes. The T, S superscript denote probabilities coming from the teacher and the student network respectively.

Another key area of advancement is logit normalization, which refines the teacher's raw output before transfer. Instead of using a constant temperature to soften all predictions, methods like [38] propose customizing the temperature for each sample based on its logit distribution characteristics. Similarly, [39] posits that the magnitude of the teacher's confidence is not always necessary and works to reduce the teacher-student gap. Other techniques focus on logit softening or smoothing. For example, some methods perturb the teacher's logits with noise [40] or use an attention module to soften them before distillation [41] to act as a regularizer. While these logit-based methods are simple and effective, their application is limited to supervised learning and can be challenging when the teacher and student have different architectures. This often necessitates including feature-based or similarity-based knowledge to facilitate more effective knowledge transfer.

Response-based knowledge distillation serves as a foundational strategy for transferring information in LiDAR perception, focusing on training a student model to mimic the final output logits, or soft targets, of a more powerful teacher. This core concept is thoroughly explored in general frameworks like the one detailed in [42] where the student learns from the distillation of knowledge of the pivotal position logit. A powerful application of this technique is in cross-modal learning, where works such as [11] use the final output of a LiDAR teacher to train a camera-only student in a unified bird's-eye-view space. To better capture the teacher's uncertainties and refine the learning signal, more advanced methods have been developed; for instance, [43] models the teacher outputs imperfections using a cross-modal label guided distillation. This principle extends directly to other dense prediction tasks, with frameworks like [42] applying logit-mimicking to train efficient and accurate point cloud segmentation networks.

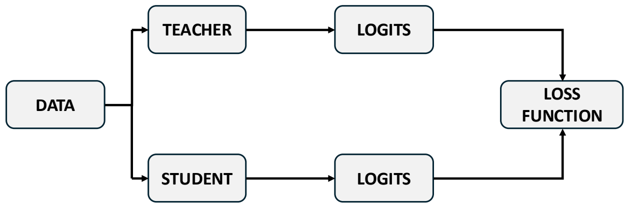


Figure 3. Process of Response-Based Knowledge Distillation. The Student model learns by mimicking the final output logits (soft targets) generated by a Teacher model. This transfer typically uses a Kullback-Leibler (KL) divergence loss between the Teacher's and Student's softened predictions, augmented with a standard cross-entropy loss against ground-truth labels.

3.2 Feature-Based

Feature-based knowledge distillation, shown in (Figure 4) has emerged as a powerful paradigm, utilizing the rich, multi-level representations from a teacher model's intermediate layers as a valuable source of knowledge. The generic form of Feature-based knowledge distillation is the following:

$$L_{feat} = \sum_{l \in L} w_l \left\| \phi_T(F_l^T) - \phi_S(F_l^S) \right\|_p \quad (3)$$

where F_l^T, F_l^S are aligned features at layer l and ϕ_T, ϕ_S re learned of fixed transforms for alignment.

This approach provides the student with more detailed, step-by-step information than relying solely on the final output. The concept was pioneered in [50], which proposed guiding a student network by providing hints from the teacher's hidden layers.

This foundational work was followed by a variety of influential methods. For instance, [44] introduced Attention Transfer (AT), which compels the student to mimic the teacher's attention maps derived from feature channels. Other works took different approaches, such as [45], which focused on transferring the boundaries formed by the teacher's neurons rather than the feature values themselves.

Some early works explored different knowledge sources, such as [46], which matched the distributions of neuron selectivity patterns, and [47], which transferred the factors of features instead of the features themselves. The field saw a significant advancement with [48], which re-evaluated many assumptions and proposed distilling the pre-activation feature maps for more effective transfer. Techniques also became more granular, as seen in [49], where a softmax function is applied to the channels of feature maps to match the resulting distributions between the teacher and student.

More recently, research has focused heavily on the transformation function that aligns student and teacher characteristics. For example, [50] randomly masks pixels in the student's feature map and uses convolutional layers to reconstruct them to match the teacher's corresponding features. This idea was later simplified in [51], which demonstrated that a simple Multi-Layer Perceptron (MLP) can effectively transform the student's features without needing a masking strategy. State-of-the-art approaches have introduced novel transformation mechanisms; [52] uniquely treats the student's features as a noisy version of the teacher's and trains a diffusion model to "denoise" them into alignment. Other innovative approaches include [53], which decomposes feature loss into separate magnitude and angular difference terms, and [23], which transfers knowledge by distilling features in the frequency domain.

Finally, [4] uses an attention mechanism to refine the feature maps of both the teacher and the student prior to distillation, ensuring that the transfer focuses on the most discriminative regions. Cross-layer strategies have also been explored, such as in [54, 55], where features from a student layer are matched not only to the corresponding teacher layer but also to all preceding teacher layers.

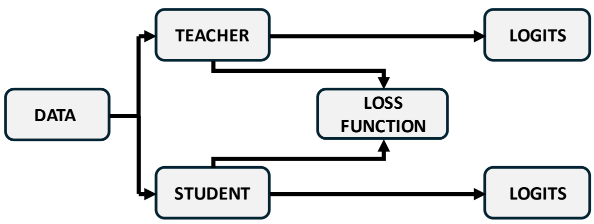


Figure 4. Feature-based knowledge distillation. The student is supervised to mimic the teacher's intermediate representations at selected layers, using lightweight alignment modules to match dimensions; only the student receives gradients.

Feature-based knowledge distillation for LiDAR data has evolved to transfer highly complex geometric and structural information from teacher to student models. This is accomplished through a variety of sophisticated techniques that go beyond simple feature mimicry. For instance, some frameworks focus on distilling fine-grained local geometric structures directly from the point cloud to ensure the student understands detailed shapes, as demonstrated in works like [56]. In parallel, other methods, such as those described in [20] focus on ensuring the student preserves the teacher's broader understanding of the overall shape and structure in the data. To provide more comprehensive guidance, knowledge can be transferred at multiple granularities, with approaches like [57] distilling information at both the fine-grained point level and the coarser voxel level simultaneously. The mechanism for enforcing this similarity can also vary; for example, [22] uses adversarial learning to align feature maps by training discriminators to make the student's and teacher's features indistinguishable.

3.3 Relation-Based

A third major type is relation-based distillation [55], depicted in (Figure 5), which focuses on relationships between different data points or layers. Rather than matching individual outputs or features, this approach encourages the student to learn the structural similarities and differences between samples, as perceived by the teacher. As a generic form relation-based knowledge distillation can be defined as:

$$\sum_{(x_i,\dots,x_n)\in X\_N} l((\psi(F_1^T,\dots F_n^T),\psi(F_1^S,\dots,F_N^S))) \tag{4}$$

where $(x_i,\dots,x_n)$ are tuples of data examples, $(F_1^T,\dots F_n^T)$ are teacher features, $(F_1^S,\dots,F_N^S)$ are student features, ψ is a relation function and l is a loss function that penalizes the differences between teacher and student.

The student might be trained to preserve the pairwise distances or the relational graph structure of the teacher's feature embeddings. Relation based knowledge distillation, also called similarity-based distillation, transfers higher-order structural knowledge by focusing on relationships within the teacher's representations rather than the raw values themselves. These techniques operate at various levels of abstraction. Some methods focus on internal feature relationships; pioneering work like [58] computed a "Flow of Solution Procedure" by taking the inner product of features between layers, while more specific approaches like [59] create and distill a channel correlation matrix. This idea was advanced in [60], which minimizes the distance between the teacher and student channel correlation maps. Beyond internal features, other methods distill instance-level similarities between different data samples. For example, [5] transfers knowledge of the distance and angle correlations between sample characteristics, and [61] distills the correlation of samples within a batch. This concept was extended in [62], which utilizes a memory bank to consider global relations between samples beyond just the current batch. At a higher level of abstraction, class-level similarity methods distill the relationships between class representations. DSD [63] distills the pairwise similarity of classes from the logits, while the popular [64] significantly enhances student performance by distilling both inter- and intra-class similarities. Finally, other techniques capture spatial similarity; for instance, [65] was one of the first to propose pooling features into different regions and distilling the similarity relationships between them. This principle has been adapted to LiDAR perception through various mechanisms. For example, in [66] the student is trained to mimic the teacher's attention maps, thereby learning which parts of a scene the teacher considers most important in relation to others. More explicit structural representations are also used; frameworks like [19] model the scene as a graph to transfer the complex relationships between different objects or regions.

The scope of these relationships can vary, with methods in papers such as [21] focusing on preserving the similarity structure between different data using affinity distillation. Ultimately, these techniques can capture very high-level information, as demonstrated in [67] which transfers both the geometric structure and the semantic relationships. In (Table 2) we provide a comprehensive classification of state-of-the-art knowledge distillation methods relevant to LiDAR data application.

4. Scheme Theory

The strategy governing the interaction between the teacher and student during the training process is known as the distillation scheme. The choice of scheme is a critical architectural decision that depends on factors like the availability of a pre-trained teacher and the desired flexibility of the training process.

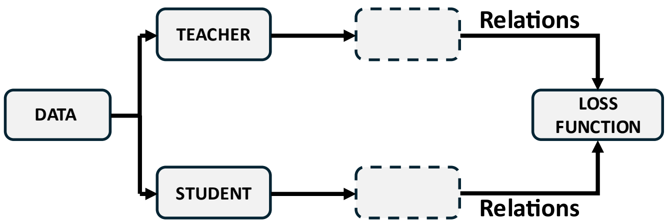


Figure 5. Relation-based knowledge distillation. The student is trained to preserve the structural relationships captured by the teacher—such as pairwise affinities, distances, or graph relations—across features or instances, rather than matching raw feature values.

4.1 Offline-distillation

Offline distillation (Figure 6) is the most traditional and widely used scheme. It is a two-stage process:

1. A large, powerful teacher model is first fully pre-trained on a large dataset until it reaches convergence.
2. The weights of this pre-trained teacher are then frozen. In a separate, subsequent stage, the smaller student model is trained to mimic the outputs of this static teacher.

This approach is straightforward and effective, but it has a key limitation: it requires a powerful, pre-trained teacher to be available from the outset.

The performance of the students is fundamentally capped by the quality of the pre-trained teacher. This scheme is most common when the student architecture is a simplified version of the teacher, for example, a network with fewer layers, fewer channels per layer, or a quantized version where the numerical precision of the parameters is reduced.

The transfer between models in offline knowledge distillation can be cross-modal, leveraging powerful models from other

Table 2. Classification by Knowledge Source

Response-based	Feature-Based	Relation-based
[37] Decoupled Knowledge Distillation (DKD)	[3] FitNets: Hints for Thin Deep Nets	[58] A Gift from Knowledge Distillation
[38] NormKD: Normalized Logits for KD	[4] Attention Transfer: Distillation of Activation Boundaries	[59] Exploring Inter-Channel Correlation for KD
[39] SphericalKD	[48] A Comprehensive Overhaul of Feature Distillation	[60] Channel Correlation Distillation
[40] Distilling Knowledge from Noisy Teachers	[49] Channel-wise Knowledge Distillation	[5] Relational Knowledge Distillation
[41] Student-Friendly Knowledge Distillation	[50] Masked Generative Distillation	[61] Similarity-Preserving Knowledge Distillation
[42] Efficient 3D Object Detection with KD	[51] Simple Framework via Channel-wise Transformation	[68] Cross-Image Relational Knowledge Distillation
[11] DistillBEV [43] LabelDistill	[52] Knowledge Diffusion for Distillation [4] Attention-Guided Feature Distillation [20] Topology-Guided KD [57] Multi-to-Single Knowledge Distillation [22] Adversarial Learning on 3D Point Clouds	[63] Double Similarity Distillation (DSD) [64] Distillation from a Stronger Teacher [65] Structured Knowledge Distillation [19] PointDistiller [21] Point-to-Voxel Knowledge Distillation [67] High-Order Structural Relation Distillation

domains like vision, or intra-modal, using a larger LiDAR model to teach a smaller one. An example of cross-modal distillation is presented in [11], where a LiDAR-based teacher model guides a multi-camera student by aligning features in the Bird's-Eye-View (BEV) space. Similarly, the [69] framework uses a pre-trained image network to supervise a point cloud network by establishing correspondences in both image-plane and bird's-eye views. The paper [25] proposes using a self-supervised image model to teach a 3D LiDAR model, using super-pixels to pool and match features from both modalities without requiring manual annotations. Intra-modal approaches are also common; for instance, [70] transfers knowledge from an accurate two-stage 3D detector to a faster one-stage detector. For segmentation tasks, [21] distills knowledge from an over-parameterized teacher to a slim student network by measuring semantic similarity at both point and voxel levels. Another approach, detailed in [15], effectively transfers rich geometric information from a 3D voxel-based model to a more efficient BEV-based model.

4.2 Online-distillation

In contrast to the sequential nature of offline distillation, online distillation (Figure 7) trains the teacher and student models simultaneously in an end-to-end process. In this paradigm, there is no need for a pre-trained teacher. Instead, a cohort of models (which can be of the same or different architectures) are trained together, and they learn from each other throughout the training process. Each model in the ensemble acts as both a teacher for the others and a student learning from collective knowledge. This approach is more flexible, avoids the two-stage training bottleneck, and is particularly useful when a powerful pre-trained teacher is not available or is too costly to create. Online knowledge distillation methods for LiDAR data involve the simultaneous, collaborative training of teacher and student models, where knowledge is transferred dynamically rather than from a fixed, pre-trained expert. A prominent example is found in [71], which proposes an online Camera-to-LiDAR distillation scheme. In this framework, an auxiliary camera-based network provides rich semantic cues to the primary LiDAR model during training through both feature-level and logit-level distillation, but is completely discarded during inference to maintain efficiency. Similarly, the [72] framework utilizes an ”auxiliary modal fusion and multi-scale fusion-to-single knowledge distillation (MSFSKD)” approach where richer semantic and structural information from multi-modal data is distilled online to a pure 3D network.

Figure 6. Offline knowledge distillation scheme. A pre-trained teacher is frozen, and a compact student is trained in a second stage to mimic the teacher via response-, feature-, or relation-based losses alongside the task loss. The setup supports intra-modal (LiDAR to LiDAR) and cross-modal transfer (e.g., LiDAR to camera in BEV or image to LiDAR via correspondence); only the student receives gradients.

4.3 Self-distillation

Self-distillation (Figure 8) is a fascinating special case of online distillation where a single network architecture acts as both the teacher and the student. This seemingly paradoxical setup is instantiated in several ways. A common approach is to use the deeper, more semantically rich layers of a network as a teacher to provide supervision for the shallower layers of the same network. This acts as a powerful form of intra-model regularization, encouraging consistency across the network’s depth and often leading to improved generalization and performance without the need for any external teacher model. It is a computationally efficient way to boost a model’s performance by leveraging its own internal knowledge.

Figure 7. Online knowledge distillation. Multiple models are co-trained without a pre-trained teacher; each act as both teacher and student by exchanging soft targets and/or intermediate features. Auxiliary branches (e.g., from another modality) provide guidance during training and are discarded at inference.

Table 3. Classification by Distillation Scheme

Offline-Distillation	Online-Distillation	Self-Distillation
[11] DistillBEV	[71] Lidar2Map	[24] Self-Distillation for Robust Lidar Segmentation
[69] HVDistill	[72] 2DPass	[23] Perturbed Self-Distillation
[49] Image-to-Lidar Self-Supervised Distillation		
[70] Diversity KD for 3D Object		
[21] Point-to-Voxel KD		
[15] Knowledge Distillation from 3D to BEV		

The paper [24] presents such a framework where a student model is guided by a teacher of the same architecture. The teacher’s weights are an exponential moving average of the student’s, and it is strengthened with Test-Time Augmentation (TTA) to generate more reliable soft labels for the student to learn from. Another approach, detailed in [23], creates an auxiliary ”perturbed branch” by applying transformations to the input point cloud and trains the model to enforce predictive consistency between the original and perturbed branches, thereby generating its own supervision signal. Furthermore, process-based online methods like the Mean Teacher approach have been applied to tasks like LiDAR-Radar segmentation, where the teacher model’s stability is enhanced by being an exponential moving average of the student model’s parameters, improving robustness against missing modalities. In (Table 3) we summarize important KD methods for LiDAR data, categorized by their distillation scheme.

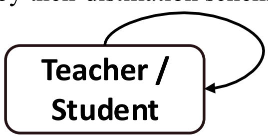


Figure 8. Self-distillation. A single architecture serves as both teacher and student: the teacher is an exponential moving average (EMA) of the student and produces soft targets (optionally with test-time augmentation), while the student is trained with a mix of task loss, distillation loss, and prediction-consistency across perturbed inputs; only the student is used at inference.

5. Algorithms

5.1 Bird’s eye view

In the realm of 2D range data, knowledge distillation has been applied to various perception tasks to enhance model efficiency and performance. For instance, in [9] a framework was introduced to transfer 3D features from point clouds to improve 2D networks for indoor scene parsing using a feature-based, semantic-aware adversarial loss, enabling the 2D network to infer simulated 3D features from 2D images without requiring 3D data during inference. Similarly, [10] utilizes knowledge distillation to bridge the domain gap between different LiDAR beams for 3D object detection by performing mimicking operations on dense Bird's Eye View (BEV) features.

Knowledge distillation in Bird's-Eye-View (BEV) has become a pivotal technique for enhancing 3D perception in autonomous driving, primarily by transferring knowledge from geometrically precise LiDAR sensors to cost-effective camera-based systems. The BEV space acts as a unified representation where features from different modalities can be aligned, though this process is challenged by feature distribution discrepancies. Seminal works like [11] established a foundational approach where a camera-based student model is trained to imitate the BEV features of a powerful LiDAR-based teacher, using balancing strategies to focus on crucial foreground objects. This concept was generalized in [12] which supports multiple distillation paths (e.g., LiDAR-to-camera, fusion-to-camera) and employs sparse distillation on key foreground points to mitigate the negative impact of misaligned background information. Other methods have introduced specialized techniques to tackle specific challenges; for instance, [13] uses raw LiDAR points to generate foreground masks that guide the distillation between camera-based models, effectively filtering out irrelevant background noise. Similarly, the [14] framework enhances performance by incorporating a geometric compensation module alongside BEV and response distillation to reduce the modality gap. The versatility of BEV as a distillation space is further highlighted in [15] which successfully transfers rich structural knowledge from complex 3D voxel-based models to more efficient BEV-based students for segmentation tasks. To further minimize the modality gap, [16] proposes a student with a simulated multi-modal architecture to better mimic a fusion-based teacher. The [14] framework introduces a two-stage process that includes a geometric compensation module to reduce the modality gap, followed by both BEV feature and response distillation. Further refining the process, [17] focuses on transferring high-level semantics by learning the inner geometry of objects rather than performing strict feature mimicry. The BEV space also facilitates knowledge transfer between different sensor suites, as demonstrated in [18] which improves a camera-radar student by distilling knowledge from a LiDAR-camera teacher. These diverse methods underscore the role of BEV as a canonical space for creating robust, efficient, and accurate perception models for autonomous driving.

5.2 PointCloud

Knowledge distillation for 3D point clouds requires specialized frameworks that address the data's inherent sparsity, irregularity, and unstructured nature, which render direct application of 2D image-based methods ineffective. To overcome these challenges, research has shifted from simple feature mimicry to transferring higher-order structural and geometric knowledge. For instance, [19] introduces a method to distill local geometric structures using dynamic graph convolution, complemented by a reweighted learning strategy to focus on information-rich points and voxels. Similarly, [20] leverages topological data analysis to ensure the student model preserves the essential geometric structures of the teacher. Another key approach, detailed in [21] transfers knowledge at multiple granularity by distilling both fine-grained point wise and coarse-grained voxel wise outputs, and uniquely captures relational knowledge through inter-point and inter-voxel affinity distillation. Beyond offline teacher-student setups, other paradigms have been explored. The paper [22] proposes PointKAD, which uses discriminators to align feature maps and logits between models. Self-distillation has also proven effective. [23] enforces predictive consistency between original and augmented point clouds to create supervisory signals in data-scarce scenarios, while [24] improves model robustness by using a time-averaged version of the student model as its own teacher, enhanced with test-time augmentation. Furthermore, cross-modal techniques such as [25] leverage powerful, pre-trained 2D vision models like CLIP to enrich 3D representations without requiring 3D labels. Other notable methods include [73] which uses a channel-wise autoencoder. Beyond cross-modal transfer, KD is instrumental in improving the robustness of LiDAR-only detectors. The inherent sparsity of point clouds and frequent occlusions can cause models to miss objects or inaccurately estimate their properties. To mitigate this, [26] introduces an innovative approach where a teacher model is trained on object-complete point clouds, which are generated by aggregating scans over time to fill in missing data. The knowledge from this strong teacher is then distilled into a student model that operates on sparse, single-frame LiDAR data, significantly improving its ability to handle occlusions.

Lidar data, generated by laser pulses measuring time-of-flight, forms discrete 3D point clouds that present unique challenges for deep learning models. These challenges stem from the data's inherent properties, including non-uniform point distribution with "scanning holes" and "scanning lines," as detailed in [75]. The data also exhibits 1D curve-like structures, as explored in [75], and contains scanning anomalies like "reflection noise" and "ghosts". Furthermore, Lidar point clouds are intrinsically sparse, particularly at longer ranges, and possess an irregular structure, making them difficult for traditional convolutional networks to process efficiently. The computational cost of 3D operations, the complexity of data annotation, and a significant "domain gap" across different LiDAR sensors further add to the difficulties, as highlighted in [62] and [68]. The presence of crucial reflectance information, which can be lost during compression, is another challenge, leading to solutions like [76]. Domain adaptation is another critical challenge, as performance can degrade when a model is deployed with a different LiDAR sensor or in different weather conditions than it was trained on. [27] directly addresses this by using a teacher trained in clear weather to guide a student operating in adverse conditions, aligning instance features through density and shape similarity metrics. To address the multifaceted challenges of point clouds, some methods combine different forms of knowledge. The framework presented in [77] leverages both fine-grained point-level knowledge and more abstract, structural voxel-level knowledge, providing a more holistic supervisory signal for the student model. The work [57] proposes a framework where a teacher model, having access to multiple, aggregated scans of a scene, distills its comprehensive understanding to a student that processes only a single scan. This is achieved through a combination of feature-level, logit-level, and instance-aware similarity distillation, allowing the student to perform well even with less input information. Work [78] leverages powerful, pre-trained vision models by distilling their rich, general-purpose knowledge into a point cloud segmentation network, circumventing the need for dense 3D labels. Taking this a step further, [71] distills knowledge from a vision-language model, enabling a 3D network to understand and segment fine-grained object parts based on textual descriptions.

5.3 Projected 2D views

Knowledge distillation applied to projected 2D views of LiDAR data represents a crucial strategy for balancing computational efficiency with high performance, particularly for deployment on resource-constrained hardware. This approach first converts sparse, unstructured 3D point clouds into dense, regular 2D representations like spherical range images, which can be processed by highly optimized 2D CNNs. However, this projection is a lossy process that can cause distortion and information loss, especially for small or distant objects. To mitigate this, knowledge distillation is employed to transfer rich, high-fidelity knowledge from a powerful teacher model—often one that processes full 3D data—to a lightweight student model operating on the 2D projection.

A prime example is in semantic segmentation, where foundational models like [79] and [80] established the viability of using range images. Building on this, the work [81]

Table 4. Classification by Model

Intra-Model	Cross-Model
[70] Diversity KD for 3D Object Detection	[11] DistillBEV
[21] Point-to-Voxel KD	[69] HVDistill
[15] Knowledge Distillation from 3D to BEV	[25] Image-to Lidar Self-Supervised Distillation
[10] Lidar Distillation	[12] UniDistill
[26] Weak-to-Strong 3D Object Detection	[13] BEV-LGKD
[42] Efficient 3D Object Detection with KD [24] Self-Distillation for Robust Lidar Segmentation [23] Perturbed Self-Distillation	[18] CRKD [61 PartDistill [71] LiDAR2Map [72] 2DPass [78] Segment Any Point Cloud Sequences

introduced a soft-label distillation method (SLKD) to effectively train an efficient student network for off-road segmentation on range images. This paradigm also extends to cross-modal and cross-representation scenarios. For instance, [15] explicitly distills knowledge from a complex 3D voxel-based teacher to a real-time BEV-based student. Similarly, [82] creates a pseudo-LiDAR representation from images, which is then enhanced by a teacher trained on real LiDAR data. Furthermore, 2D projections serve as a critical bridge for transferring knowledge from powerful 2D foundation models to the 3D domain. Frameworks like [25] and [83] use projections to create correspondences between image pixels and 3D points, enabling distillation from models like CLIP without 3D labels. The paper [84] utilizes a front-view projection to enable a teacher-student exchange between 2D and 3D modalities. This principle is also used to bridge sensor gaps, as shown in [10] which distills knowledge from high-resolution to low-resolution LiDAR sensors, a process often facilitated by their 2D projected representations. Finally, works like [85] and [69] focus on improving the projection step itself within the distillation pipeline to better align features across modalities.

5.4 Results and Discussion

Table 5. Representative quantitative comparison of key knowledge distillation methods for LiDAR-centric 3D object detection and mapping. For each method, we list representation and modality path, dataset, task, and the approximate gain Δ = (Student + KD) − Student. Values are approximated, aggregated from the original papers.

Method	Represen-tation	Dataset	Task	Δ (metric)
[42]	PC/intra (LiDAR→ LiDAR)	Waymo, KITTI	3D OD	2-3 mAP
[70]	PC/intra (LiDAR→ LiDAR)	Waymo, KITTI	3D OD	2-4 mAP
[10]	BEV/intra (Hi→ Lo beam)	KITTI	3D OD	3-4 mAP
[26]	PC/intra (Agg→ Single)	Waymo, KITTI-like	3D OD	3-6 mAP
[11]	BEV/cross (LiDAR→ Cam)	nuScenes, Waymo	3D OD	3-4 mAP
[43]	BEV/cross (LiDAR→ Cam)	nuScenes	3D OD	2-3 mAP
[14]	BEV/cross (LiDAR→ Cam)	nuScenes	3D OD	3-4 mAP
[16]	BEV/cross (Fusion→ Cam)	nuScenes	3D OD	3-4 mAP
[17]	BEV/cross (LiDAR→ Cam)	nuScenes	3D OD	2-3 mAP
[18]	BEV/cross (LiDAR+ Cam → Cam+Radar)	nuScenes	3D OD	2-3 mAP
[67]	BEV/cross (LiDAR→ Mono)	KITTI, nuScenes	3D OD	3-5 mAP
[71]	BEV/cross (Cam→ LiDAR, online KD)	nuScenes-like	Semantic mapping	2-4 mIOU

To make the surveyed methods more accessible to practitioners, we complement the qualitative discussion in this section with three summary tables that organize LiDAR KD approaches by representation, modality path, and task, and report representative quantitative gains. We also include the datasets used in each respective paper: Waymo [82], KITTI [26], nuScenes [86], ShapeNetPart [42], NYUv2 [87], SUN-RGBD [88].

(Table 5) focuses on 3D object detection and semantic mapping. It collects both intra-modal LiDAR-to-LiDAR KD and cross-modal BEV KD, together with their main datasets (Waymo, KITTI, nuScenes), metrics (3D mAP, NDS), and approximate improvements of the student over its non-distilled baseline.

(Table 6) focuses on 3D / BEV semantic segmentation and related 3D tasks. It includes intra-modal LiDAR segmentation KD as well as cross-modal segmentation KD from images or foundation models. For each method we report the main dataset, metric, and representative gains.

Finally, (Table 7) targets 2D range views and other 2D projections of LiDAR data, which are particularly relevant for resource-constrained platforms. It summarizes KD methods that operate on projected representations.

Although this work is a survey and does not introduce a new model with its own experimental results, the methods we review report relatively consistent empirical trends across datasets and tasks. In 3D object detection on benchmarks such as KITTI, Waymo, and nuScenes, intra-modal distillation ([42], [70], [10]) typically yields gains of approximately 2–4 mAP over non-distilled baselines, with larger improvements (up to about 5–6 mAP) observed when the teacher has access to richer information such as multi-frame aggregation or denser beams [26]. Cross-modal BEV distillation from LiDAR or fusion teachers to camera-only students ([11], [14], [16], [12]) shows similar relative improvements, usually in the range of 2–4 mAP, which is notable given that the students operate on cheaper sensors and often have stricter runtime budgets. For LiDAR semantic segmentation, both point-cloud and BEV-based KD methods ([21], [15], [24], [23]) report improvements on the order of 2–5 mIoU on SemanticKITTI, nuScenes lidarseg, and related datasets, with self-distillation and weak-label settings particularly benefiting from the extra regularization and soft supervision.

Across representations and modality paths, several qualitative patterns emerge. First, feature- and relation-based KD tend to provide the largest benefits in LiDAR settings, especially when the teacher and student operate on different resolutions or views (e.g., [21], [18]). Purely response-based KD remains simple and effective but often underutilizes geometric and structural cues that are critical for sparse point clouds. Second, cross-modal BEV distillation has become a practical mechanism to transfer geometric reliability from LiDAR to camera-only or camera-radar systems while preserving real-time performance, suggesting that BEV is an effective common space for multi-sensor KD. Third, methods targeting robustness, such as cross-weather KD [27], cross-sensor/beam distillation [10], and self-distillation for noisy or compressed point clouds [24], [76]—consistently reduce performance gaps under adverse conditions, even when absolute gains on standard benchmarks are modest. Finally, although only a few works explicitly address 2D range LiDAR with KD ([81] and related detectors on DROW, JRDB, FROG), they indicate that similar trends hold: compact models operating on single-plane scans can recover several points of F1/IoU when guided by stronger teachers, which is important for resource-constrained

Table 6. Representative quantitative comparison of knowledge distillation methods for LiDAR-centric 3D/BEV semantic segmentation and related 3D tasks. We report approximate the gain on the main metric reported in each work. Cross-modal methods (image/vision-model teachers to LiDAR) and self-distillation approaches are included to illustrate their benefit in low-label and robustness settings. Values are approximated, aggregated from the original papers.

Method	Representation	Dataset	Task	Δ (metric)
[21]	PC/intra (LiDAR→ LiDAR)	Sem. KITTI, nuScenes	Segm.	4.4 mIOU
[24]	PC/intra (LiDAR→ LiDAR)	nuScenes lidarseg	Segm.	2.2 mIOU
[23]	BEV/intra (Hi→ Lo beam)	Sem. KITTI, nuScenes	Segm.	3-5 mIOU
[15]	PC/intra (Agg→ Single)	Sem. KITTI, nuScenes	Segm. (weak)	3-4 mIOU
[72]	BEV/cross (LiDAR→ Cam)	Sem. KITTI, nuScenes	Segm.	3-4 mIOU
[25]	BEV/cross (LiDAR→ Cam)	nuScenes, Waymo	Segm. (downs.)	3-5 mIOU
[9]	BEV/cross (LiDAR→ Cam)	KITTI, nuScenes,	Seq. Segm.	+10 mIOU
[28]	BEV/cross (Fusion→ Cam)	ShapeNetPart	Part Segm.	5-10 mIOU
[19]	BEV/cross (LiDAR→ Cam)	KITTI, nuScenes	3D OD (struct. KD)	3-5 mAP
[20]	BEV/cross (LiDAR+ Cam→ Cam +Radar)	KITTI nuScenes	Det./ Segm.	2-4 mAP, mIOU
[57]	BEV/cross (LiDAR→ Mono)	Sem. KITTI, nuScenes	Segm.	3-4 mIOU
[22]	BEV/cross (Cam→ LiDAR, online KD)	Class., det. datasets	Class./ Det.	2-4 Acc., mAP
[56]	PC/intra (self, geom.)	cross-view datasets	Loc.	3-7 Recall
[9]	2D/cross (3D→ 2D)	NYUv2, SUN-RGBD	Indoor segm.	2-4 mIOU

mobile robots. From a practical standpoint, these results suggest several design guidelines.

For detection and segmentation on native point clouds, combining point- and voxel-level supervision with relational or affinity losses tends to yield more robust gains than logit matching alone. For cross-modal transfer, careful spatial alignment in BEV and the use of foreground-aware or uncertainty-aware distillation are crucial to avoid negative transfer from misaligned background regions. For projected views (range images and 2D scans), using a 3D teacher or multi-frame teacher to compensate for projection and sparsity artefacts appears particularly effective. Overall, the empirical evidence across the surveyed literature supports the central message of this paper: knowledge distillation is a consistently beneficial tool for trading teacher complexity for student efficiency in LiDAR perception, typically recovering several points of accuracy or robustness while enabling deployment on real-time, resource-constrained autonomous platforms.

Table 7. Representative quantitative comparison of knowledge distillation methods that operate on 2D range views or other 2D projections of LiDAR data. We report the main representation and modality path, dataset, task and the student with KD improvement in metric. Values are approximated, aggregated from the original papers.

Method	Representation	Dataset	Task	Δ (metric)
[9]	2D image/ cross(3D→ 2D)	NYUv2, SUN-RGBD	Indoor sem. segm. (RGB)	2-4 mIOU
[81]	2D range view/intra (LiDAR→ LiDAR)	Off-road LiDAR scans	Det.	3-5 IOU/F1
[84]	2D front-view /cross (2D↔ 3D)	Synthetic + real video scenes	Obj. discovery	3-6 F1/AR
[85]	2D proj. (image/RV)/ cross	nuScenes, Waymo	Det., segm.	2-4 mIOU

6. Conclusion

This review has systematically charted the landscape of knowledge distillation as applied to LiDAR data, revealing it as a critical enabling technology for deploying advanced 3D perception in resource-constrained autonomous systems. We have categorized the field along two primary axes: the distillation scheme—offline, online, and self-distillation—which dictates the training paradigm, and the source of knowledge—response, feature, and relation-based—which defines the nature of the information being transferred. Our analysis across different data representations—native point clouds, Bird's-Eye-View (BEV) grids, and projected 2D views—demonstrates a clear evolutionary path.

The field has progressed from direct mimicry of final predictions to sophisticated methods that transfer nuanced, high-order structural and geometric knowledge, which is essential for handling the unique sparsity and irregularity of point cloud data.

A key synthesis from this review is the emergence of specialized techniques tailored to each data representation. For native point clouds, the most effective methods have moved beyond simple feature matching to focus on distilling local geometric structures and relational affinities, thereby respecting the data's unstructured nature. For cross-modal applications, the BEV representation has become the "lingua franca," providing a unified space to bridge the significant modality gap between geometrically precise LiDAR and semantically rich cameras. Meanwhile, for highly efficient systems, distillation applied to 2D projections serves as a vital tool to compensate for the inherent information loss, pushing the accuracy-efficiency frontier.

Looking forward, the field faces several grand challenges that will define the next generation of research. The persistent modality gap, even within unified spaces like BEV, requires more advanced feature alignment and normalization techniques to be fully overcome. More critically, the community must move beyond perception-level metrics and embrace end-to-end, closed-loop evaluation in high-fidelity simulators to assess the true impact of distillation on driving safety and planning performance. This will necessitate the development of hardware-aware and planning-aware distillation frameworks that optimize for the entire autonomous stack.

Perhaps the most transformative trajectory lies in expanding the source of knowledge beyond specialized sensor models. The nascent trend of distilling from large-scale, pre-trained foundation models, such as Vision-Language Models, signals a paradigm shift from transferring perceptual patterns to transferring abstract, conceptual, and even predictive world knowledge. The goal will be to distill comprehensive, generative world models into efficient, real-time agents capable of proactive and robust decision-making.

In conclusion, knowledge distillation is no longer just a tool for model compression; it is a fundamental methodology for knowledge transfer that will be central to building the next generation of intelligent, safe, and scalable autonomous systems.

Conflict of Interest-The authors declare no potential conflicts of interest.

Funding Source-This work was supported by REXASI-PRO H-EU project, call HORIZON-WIDERA-2023-ACCESS-06, Grant agreement no. 101159179. (Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.)

Authors' Contributions -Author-1 researched literature and prepared the study. After several discussions, both authors came up with the unified taxonomy for KD methods for LiDAR data.

Author-1 created a first draft on the manuscript, whereas the second author was involved in refining the structure of the material.

All authors reviewed and edited the manuscript and approved the final version of the manuscript.

References

[1] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[2] C. Bucilua, R. Caruana, and A. Niculescu-Mizil, "Model compression," *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp.535-541, 2006.

[3] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for thin deep nets," arXiv preprint, 2015.

[4] A. M. Mansourian, A. Jalali, R. Ahmadi, and S. Kasaei, "Attention-guided feature distillation for semantic segmentation," arXiv preprint arXiv:2403.05451, 2024.

[5] W. Park, D. Kim, Y. Lu, and M. Cho, "Relational knowledge distillation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.3967–3976, 2019.

[6] A. M. Mansourian, R. Ahmadi, M. Ghafouri, A. M. Babaei, E. B. Golezani, Z. Y. Ghamchi, V. Ramezanian, A. Taherian, K. Dinashi, A. Miri, et al., "A comprehensive survey on knowledge distillation," arXiv preprint arXiv:2503.12067, 2025.

[7] BhavyaShree P, Sharon Thomas Takri, R. Gurunath, "*A Comparative Study of Enabling Technologies for Autonomous Vehicles*", *International Journal of Computer Sciences and Engineering*, Vol.07, Special Issue.09, pp.9-11, 2019.

[8] Nupur Choudhury, Rupesh Mandal2, "A survey on Robotic Vehicles in Autonomous Navigation", *International Journal of Computer Sciences and Engineering*, Vol.3, Issue.5, pp.79-84, 2015.

[9] Z. Liu, X. Qi, and C.-W. Fu, "3D-to-2D distillation for indoor scene parsing," *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE/CVF, pp.4462–4472, 2021.

[10] Y. Wei, Z. Wei, Y. Rao, J. Li, J. Zhou, and J. Lu, "LiDAR distillation: Bridging the beam-induced domain gap for 3D object detection," *European Conference on Computer Vision*, Springer, pp.179–195, 2022.

[11] Z. Wang, D. Li, C. Luo, C. Xie, and X. Yang, "DistillBEV: Boosting multi-camera 3D object detection with cross-modal knowledge distillation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.8637–8646, 2023.

[12] S. Zhou, W. Liu, C. Hu, S. Zhou, and C. Ma, "UniDistill: A universal cross-modality knowledge distillation framework for 3D object detection in bird's-eye view," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.5116–5125, 2023.

[13] J. Li, M. Lu, J. Liu, Y. Guo, L. Du, and S. Zhang, "BEV-LGKD: A unified LiDAR-guided knowledge distillation framework for BEV 3D object detection," arXiv preprint arXiv:2212.00623, 2022.

[14] T. Jiao, Y. Chen, Z. Zhang, C. Guo, and J. Song, "MMDistill: Multi-modal BEV distillation framework for multi-view 3D object detection," *Computers, Materials & Continua*, Vol.81, Issue.3, 2024. https://doi.org/10.32604/cmc.2024.058238.

[15] F. Jiang, H. Gao, S. Qiu, H. Zhang, R. Wan, and J. Pu, "Knowledge distillation from 3D to bird's-eye-view for LiDAR semantic segmentation," *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pp.402–407, 2023.

[16] H. Zhao, Q. Zhang, S. Zhao, Z. Chen, J. Zhang, and D. Tao, "SimDistill: Simulated multi-modal distillation for BEV 3D object detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, Vol.38, pp.7460–7468, 2024.

[17] S. Xu, F. Li, P. Huang, Z. Song, and Z.-X. Yang, "TIGDistill-BEV: Multi-view BEV 3D object detection via target inner-geometry learning distillation," arXiv preprint arXiv:2412.20911, 2024.

[18] L. Zhao, J. Song, and K. A. Skinner, "CRKD: Enhanced camera-radar object detection with cross-modality knowledge distillation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.15470–15480, 2024.

[19] L. Zhang, R. Dong, H.-S. Tai, and K. Ma, "PointDistiller: Structured knowledge distillation towards efficient and compact 3D detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.21791–21801, 2023.

[20] L. T. Hai, T. D. Le, Z. Ding, Q. Tian, and T.-S. Hy, "Topology-guided knowledge distillation for efficient point cloud processing," arXiv preprint arXiv:2505.08101, 2025.

[21] Y. Hou, X. Zhu, Y. Ma, C. C. Loy, and Y. Li, "Point-to-voxel knowledge distillation for LiDAR semantic segmentation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.8479–8488, 2022.

[22] S. J. Sanjay, J. Akash, A. S. Dimple, et al., "Adversarial learning based knowledge distillation on 3D point clouds," *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp.2932–2941, 2025.

[23] Y. Zhang, Y. Qu, Y. Xie, Z. Li, S. Zheng, and C. Li, "Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.15520–15528, 2021.

[24] J. Li, H. Dai, and Y. Ding, "Self-distillation for robust LiDAR semantic segmentation in autonomous driving," *European Conference on Computer Vision*, Springer, pp.659–676, 2022.

[25] C. Sautier, G. Puy, S. Gidaris, A. Boulch, A. Bursuc, and R. Marlet, "Image-to-LiDAR self-supervised distillation for autonomous driving data," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.9891–9901, 2022.

[26] A. Gambashidze, A. Dadukin, M. Golyadkin, M. Razzhivina, and I. Makarov, "Weak-to-strong 3D object detection with X-ray distillation," arXiv preprint, 2024.

[27] X. Huang, H. Wu, X. Li, X. Fan, C. Wen, and C. Wang, "Sunshine to rainstorm: Cross-weather knowledge distillation for robust 3D object detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, Vol.38, pp.2409–2416, 2024.

[28] A. Umam, C.-K. Yang, M.-H. Chen, J.-H. Chuang, and Y.-Y. Lin, "PartDistill: 3D shape part segmentation by vision-language model distillation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.3470–3479, 2024.

[29] D. Jia, A. Hermans, and B. Leibe, "DR-SPAAM: A spatial-attention and auto-regressive model for person detection in 2D range data," *International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020.

[30] H. Yang, Y. Yang, C. Yao, C. Liu, and Q. Chen, "Li2Former: Omni-dimension aggregation transformer for person detection in 2-D range data," *IEEE Transactions on Instrumentation and Measurement*, Vol.73, 2024. 10.1109/TIM.2024.3420353.

[31] L. Beyer, A. Hermans, and B. Leibe, "DROW: Real-time deep learning based wheelchair detection in 2D range data," *IEEE Robotics and Automation Letters (RA-L)*, Vol.2, Issue.2, pp.585-592, 2016. 10.1109/LRA.2016.2645131.

[32] L. Beyer, A. Hermans, T. Linder, K. O. Arras, and B. Leibe, "Deep person detection in 2D range data," *IEEE Robotics and Automation Letters (RA-L)*, Vol.3, Issue.3, pp.2726-2733, 2018. 10.1109/LRA.2018.2835510.

[33] R. Martin-Martin, M. Patel, H. Rezatofighi, A. Shenoi, J. Gwak, E. Frankel, A. Sadeghian, and S. Savarese, "JRDB: A dataset and benchmark of egocentric robot visual perception of humans in built environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, Issue.6, pp.6748–6765, 2021. 10.1109/TPAMI.2021.3070543.

[34] F. Amodeo, N. Pérez-Higueras, L. Merino, and F. Caballero,

"FROG: A new people detection dataset for knee-high 2D range finders," arXiv preprint arXiv:2306.08531, 2023.
[35] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo Open Dataset," arXiv preprint arXiv:1912.04838, 2019.
[36] B. Yang, T. Tao, W. Wu, Y. Zhang, X. Meng, and J. Yang, "MultiDistiller: Efficient multimodal 3D detection via knowledge distillation for drones and autonomous vehicles," *Drones*, Vol.9, Issue.5, pp.322, 2025. https://doi.org/10.3390/drones9050322.
[37] B. Zhao, Q. Cui, R. Song, Y. Qiu, and J. Liang, "Decoupled knowledge distillation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.11953–11962, 2022.
[38] Z. Chi, T. Zheng, H. Li, Z. Yang, B. Wu, B. Lin, and D. Cai, "NormKD: Normalized logits for knowledge distillation," arXiv preprint arXiv:2308.00520, 2023.
[39] J. Guo, M. Chen, Y. Hu, C. Zhu, X. He, and D. Cai, "Reducing the teacher-student gap via spherical knowledge distillation," arXiv preprint arXiv:2010.07485, 2020.
[40] B. B. Sau and V. N. Balasubramanian, "Deep model compression: Distilling knowledge from noisy teachers," arXiv preprint arXiv:1610.09650, 2016.
[41] M. Yuan, B. Lang, and F. Quan, "Student-friendly knowledge distillation," *Knowledge-Based Systems*, Vol.296, pp.111915, 2024. https://doi.org/10.1016/j.knosys.2024.111915.
[42] J. Yang, S. Shi, R. Ding, Z. Wang, and X. Qi, "Towards efficient 3D object detection with knowledge distillation," *Advances in Neural Information Processing Systems*, Curran Associates, Vol.35, pp.21300–21313, 2022.
[43] S. Kim, Y. Kim, S. Hwang, H. Jeong, and D. Kum, "LabelDistill: Label-guided cross-modal knowledge distillation for camera-based 3D object detection," *European Conference on Computer Vision*, Springer, pp.19–37, 2024.
[44] S. Zagoruyko, N. Komodakis, "Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer,", arXiv preprint arXiv: 1612.03928, 2016.
[45] B. Heo, M. Lee, S. Yun, and J. Y. Choi, "Knowledge transfer via distillation of activation boundaries formed by hidden neurons," *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, Vol.33, pp.3779–3787, 2019.
[46] Z. Huang and N. Wang, "Like what you like: Knowledge distill via neuron selectivity transfer," arXiv preprint arXiv:1707.01219, 2017.
[47] J. Kim, S. Park, and N. Kwak, "Paraphrasing complex network: Network compression via factor transfer," *Advances in Neural Information Processing Systems*, Curran Associates, Vol.31, 2018.
[48] B. Heo, J. Kim, S. Yun, H. Park, N. Kwak, and J. Y. Choi, "A comprehensive overhaul of feature distillation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.1921–1930, 2019.
[49] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "Channel-wise knowledge distillation for dense prediction," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.5311–5320, 2021.
[50] Z. Yang, Z. Li, M. Shao, D. Shi, Z. Yuan, and C. Yuan, "Masked generative distillation," *European Conference on Computer Vision*, Springer, pp.53–69, 2022.
[51] Z. Liu, Y. Wang, X. Chu, N. Dong, S. Qi, and H. Ling, "A simple and generic framework for feature distillation via channel-wise transformation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.1129–1138, 2023.
[52] T. Huang, Y. Zhang, M. Zheng, S. You, F. Wang, C. Qian, and C. Xu, "Knowledge diffusion for distillation," *Advances in Neural Information Processing Systems*, Curran Associates, Vol.36, pp.65299–65316, 2023.
[53] T. Liu, C. Chen, X. Yang, and W. Tan, "Rethinking knowledge distillation with raw features for semantic segmentation," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, IEEE/CVF, pp.1155–1164, 2024.
[54] Y. Chen, Y. Xian, A. Koepke, Y. Shan, and Z. Akata, "Distilling audio-visual knowledge by compositional contrastive learning," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.7016–7025, 2021.
[55] P. Chen, S. Liu, H. Zhao, and J. Jia, "Distilling knowledge via knowledge review," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.5008–5017, 2021.
[56] S. Tong, Z. Xia, A. Alahi, X. He, and Y. Shi, "GeoDistill: Geometry-guided self-distillation for weakly supervised cross-view localization," arXiv preprint, 2025.
[57] S. Qiu, F. Jiang, H. Zhang, X. Xue, and J. Pu, "Multi-to-single knowledge distillation for point cloud semantic segmentation," arXiv preprint arXiv:2304.14800, 2023.
[58] J. Yim, D. Joo, J. Bae, and J. Kim, "A gift from knowledge distillation: Fast optimization, network minimization and transfer learning," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp.4133–4141, 2017.
[59] L. Liu, Q. Huang, S. Lin, H. Xie, B. Wang, X. Chang, and X. Liang, "Exploring inter-channel correlation for diversity-preserved knowledge distillation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.8271–8280, 2021.
[60] C. Wang, J. Zhong, Q. Dai, Y. Qi, Q. Yu, F. Shi, R. Li, X. Li, and B. Fang, "Channel correlation distillation for compact semantic segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.37, Issue.03, pp.2350004, 2023. https://doi.org/10.1142/S0218001423500040.
[61] F. Tung and G. Mori, "Similarity-preserving knowledge distillation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE/CVF, pp.1365–1374, 2019.
[62] C. Yang, H. Zhou, Z. An, X. Jiang, Y. Xu, and Q. Zhang, "Cross-image relational knowledge distillation for semantic segmentation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.12319–12328, 2022.
[63] Y. Feng, X. Sun, W. Diao, J. Li, and X. Gao, "Double similarity distillation for semantic image segmentation," *IEEE Transactions on Image Processing*, Vol.30, Issue.pp.5363–5376, 2021. 10.1109/TIP.2021.3083113.
[64] T. Huang, S. You, F. Wang, C. Qian, and C. Xu, "Knowledge distillation from a stronger teacher," *Advances in Neural Information Processing Systems*, Curran Associates, Vol.35, pp.33716–33727, 2022.
[65] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.2604–2613, 2019.
[66] L. Zhang and K. Ma, "Structured knowledge distillation for accurate and efficient object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, Issue.12, pp.15706–15724, 2023. 10.1109/TPAMI.2023.3300470.
[67] W. Yan, L. Xu, H. Liu, C. Tang, and W. Zhou, "High-order structural relation distillation networks from LiDAR to monocular image 3D detectors," *IEEE Transactions on Intelligent Vehicles*, Vol.9, Issue.2, pp.3593–3604, 2023. 10.1109/TIV.2023.3341981.
[68] C. B. Rist, M. Enzweiler, and D. M. Gavrila, "Cross-sensor deep domain adaptation for LiDAR detection and segmentation," *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp.1535–1542, 2019.
[69] S. Zhang, J. Deng, L. Bai, H. Li, W. Ouyang, and Y. Zhang, "HVDistill: Transferring knowledge from images to point clouds via unsupervised hybrid-view distillation," arXiv preprint, 2024.
[70] K. Ning, Y. Liu, Y. Su, and K. Jiang, "Diversity knowledge distillation for LiDAR-based 3-D object detection," *IEEE Sensors Journal*, Vol.23, Issue.11, pp.11181–11193, 2023. 10.1109/JSEN.2023.3241624.
[71] S. Wang, W. Li, W. Liu, X. Liu, and J. Zhu, "LiDAR2Map: In defense of LiDAR-based semantic map construction using online

camera distillation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.5186–5195, 2023.
[72] X. Yan, J. Gao, C. Zheng, C. Zheng, R. Zhang, S. Cui, and Z. Li, "2Dpass: 2D priors assisted semantic segmentation on LiDAR point clouds," *European Conference on Computer Vision*, Springer, pp.677–695, 2022.
[73] H. Cho, J. Choi, G. Baek, and W. Hwang, "ITKD: Interchange transfer-based knowledge distillation for 3D object detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.13540–13549, 2023.
[74] Y. Guo, Y. Li, D. Ren, X. Zhang, J. Li, L. Pu, C. Ma, X. Zhan, J. Guo, M. Wei, et al., "LidarNet: A real-scanned 3D point cloud dataset for indoor scenes," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.21989–21999, 2024.
[75] C. Stearns, A. Fu, J. Liu, J. J. Park, D. Rempe, D. Paschalidou, and L. J. Guibas, "CurveCloudNet: Processing point clouds with 1D structure," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, pp.27981–27991, 2024.
[76] H. Jing, A. Wang, Y. Zhang, D. Bu, and J. Hou, "Reflectance prediction-based knowledge distillation for robust 3D object detection in compressed point clouds," arXiv preprint arXiv:2505.17442, 2025.
[77] J. Zhang and J. Liu, "Voxel-to-Pillar: Knowledge distillation of 3D object detection in point cloud," *Proceedings of the 4th European Symposium on Software Engineering*, pp.99–104, 2023.
[78] Y. Liu, L. Kong, J. Cen, R. Chen, W. Zhang, L. Pan, K. Chen, and Z. Liu, "Segment any point cloud sequences by distilling vision foundation models," *Advances in Neural Information Processing Systems, Curran Associates*, Vol.36, pp.37193–37229, 2023.
[79] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, "SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud," *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp.4376–4382, 2019.
[80] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet++: Fast and accurate LiDAR semantic segmentation," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp.4213–4220, 2019.
[81] N. Kim and J. An, "Knowledge distillation for traversable region detection of LiDAR scan in off-road environments," *Sensors*, Vol.24, Issue.1, pp.79, 2023. https://doi.org/10.3390/s24010079.
[82] E. Son, J. Choi, J. Song, Y. Jin, and S. J. Lee, "Monocular depth estimation from a fisheye camera based on knowledge distillation," *Sensors*, Vol.23, Issue.24, pp.9866, 2023. https://doi.org/10.3390/s23249866.
[83] A. Mahmoud, A. Harakeh, and S. Waslander, "Image-to-LiDAR relational distillation for autonomous driving data," *European Conference on Computer Vision*, Springer, pp.459–475, 2024.
[84] S. Lahlali, S. Kara, H. Ammar, F. Chabot, N. Granger, H. Le Borgne, and Q.-C. Pham, "XMoD: Cross-modal distillation for 2D/3D multi-object discovery from 2D motion," arXiv preprint arXiv:2503.15022, 2025.
[85] H. Govindarajan, M. K. Wozniak, M. Klingner, C. Maurice, B. R. Kiran, and S. Yogamani, "CleverDistiller: Simple and spatially consistent cross-modal distillation," arXiv preprint, 2025.
[86] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE/CVF, pp.11618–11628, 2020
[87] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," *Proceedings of the 12th European Conference on Computer Vision (ECCV)*, Springer, pp.746–760, 2012.
[88] S. Song, S. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp.567–576, 2015

AUTHORS PROFILE

Avesalon Razvan Marian earned his B. Tech., and M. Tech. in computer science at "Universitatea Transilvania", Brasov in 2023, and 2025, respectively. He is currently working at Siemens Industry Software in Brasov, Romania for more than 2 years in Computer Vision.

Miron Radu earned his B. Tech., M. Tech., and is currently enrolled in its Ph.D. in computer science at "Alexandru Ioan Cuza University" of Iasi in 2018, 2023, and 2025, respectively. He is currently working at Siemens Industry Software for more than 4 years on Machine Learning for Computer Vision.

