ISSN 2409-7284 (Online)

Enhancing Object Detection in Smart Logistics: Integration of IFrustum-Pointnets Model with Industrial Internet of Things

Sangbing Tsai
International Engineering and Technology Institute, Hong Kong
*Corresponding Author Email: klj0418@gmail.com
Received: October,5,2025；Revised: January,3,2026；Accepted: February,12,2026

ABSTRACT. The Industrial Internet of Things (IIoT) is driving unprecedented transformation across the production sector. Nevertheless, the field of intelligent logistics object detection continues to face challenges in achieving high accuracy and robustness in object recognition. To tackle these issues, we propose the IFrustum-Pointnets model. Specifically, we optimized the threshold selection strategy by adopting a more suitable threshold to enhance both the stability and accuracy of target detection. Additionally, we refined the parallel attention mechanism and replaced the original loss function with Focal Loss to mitigate class imbalance in the dataset, thereby improving the model's overall performance and robustness. We also introduce a novel IoT framework designed to enhance the intelligent logistics system, which comprises four main layers: the Big Data Layer, the Edge Data Layer, the IoT Layer, and the Deployment Layer. The proposed algorithm leverages both the Big Data and Edge Data Layers to collect and process data in real time, enabling more effective deployment of the IFrustum-Pointnets model. Experimental results on the KITTI 3D Object Detection Benchmark dataset show that IFrustum-Pointnets outperforms traditional methods across all evaluation metrics. These findings not only demonstrate the potential applications of IFrustum-Pointnets in intelligent logistics but also provide strong support for the advancement of IIoT technologies.
Keywords: Industrial Internet of Things, Intelligent Logistics, Frustum-Pointnets, Focal Loss, Bidirectional Attention Extraction Mechanism

1. Introduction

The Industrial Internet of Things (IIoT) is a revolutionary technology that significantly enhances the automation and intelligence of industrial production and operations by interconnecting sensors, machinery, and computing devices[1, 2]. Intelligent logistics management, as a key application area of IIoT, leverages these interconnected devices to track goods, monitor inventory in real-time, and automate the scheduling of transportation resources, thereby optimizing the efficiency and responsiveness of the entire supply chain. However, despite the certain level of automation and monitoring achieved by intelligent logistics systems, they still lack sufficient precision in path planning and goods handling, which limits their potential maximization[3]. In this context, the introduction of three-dimensional object detection technology becomes crucial for enhancing the accuracy and efficiency of intelligent logistics management systems. By providing high-precision

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

spatial data and object recognition, it offers an effective solution to the challenges of precise positioning in traditional logistics systems[4].

In recent years, three-dimensional object detection technology has been widely applied in the field of intelligent logistics management to enhance system efficiency and accuracy[5]. Specifically, 3D object detection based on LiDAR utilizes laser radar to generate high-precision point cloud data, enabling precise spatial positioning and size measurement of objects. This technology is extensively used in autonomous logistics vehicles and automated loading and unloading systems. Grid-based 3D object detection, on the other hand, involves dividing space into regular grids to facilitate faster processing and recognition of objects in three-dimensional space by computer vision algorithms[6, 7]. This method has proven effective in sorting items within warehouses and in robotic navigation. Additionally, Point-Voxel based 3D object detection combines the advantages of point clouds and voxels. By converting point cloud data into voxel representations, it further enhances processing speed and recognition accuracy, particularly when handling large-scale point cloud data.

With the advancement of deep learning, a multitude of current object detection algorithms have achieved significant success[8]. Initially, PointNet directly learns features from point cloud data for classification and segmentation tasks. While it performs well in simple scenarios, its capabilities are limited in complex environments where it struggles to capture finer object details. Next, the SECOND algorithm enhances point cloud processing efficiency through the use of sparse convolutions, notably excelling in detecting large objects like vehicles, though its accuracy in sparse data regions still needs improvement[9]. PointPillars then speeds up data processing and enhances 3D detection performance by converting point cloud data into columnar voxels, but it still lacks in capturing spatial relationships, which may affect the recognition of complex intersecting objects[10]. 3DSSD introduces an anchor-free detection framework, increasing the flexibility and speed of detection, but this method shows unstable results when detecting small-sized objects. PV-RCNN combines the advantages of voxels and point clouds with a complex network structure to enhance the accuracy of detecting small objects at a distance, yet this also results in higher computational costs, limiting its feasibility for real-time applications[10]. Additionally, VoteNet uses an innovative voting-based strategy to directly generate 3D bounding boxes from unordered point clouds. Although it performs excellently in indoor scenarios, it still needs further improvement in its adaptability to external environments and noise tolerance. Finally, Frustum-PointNets introduces a revolutionary framework that integrates three-dimensional point clouds with two-dimensional image data, effectively capturing and recognizing objects from multiple perspectives. This method significantly enhances detection accuracy and efficiency in complex scenarios, particularly in dynamic environments for detecting vehicles and pedestrians, thanks to the guidance of 3D point cloud processing from 2D detection results }[11]. This multimodal data fusion not only strengthens the model's spatial understanding of objects but also enhances the overall robustness of the system. However, despite its effective handling of various object types, its performance still requires improvement when dealing with very small or heavily occluded objects[12].

To address these challenges, we propose a novel network framework named IFrustum-PointNets.

ISSN 2409-7284 (Online)

This model enhances the original Frustum-PointNets architecture by broadening the threshold for mask definition, thereby reducing the adverse effects of inaccurate mask predictions commonly encountered in existing systems. Moreover, we introduce an optimized parallel attention mechanism tailored to the Frustum-PointNets structure, which significantly boosts network performance by emphasizing critical features across different data modalities. This enhancement supports a more efficient and precise object detection process. Furthermore, to mitigate the prevalent issue of class imbalance, we integrate Focal Loss as part of the new loss function. This adjustment prioritizes hard-to-classify instances, ensuring that minority yet critical objects receive appropriate attention without being dominated by more frequent but less consequential categories.

Here are the three main contributions of this paper:

- This paper introduces a new IoT framework designed for smart logistics, which enhances data processing efficiency through its structured multi-layer approach. Additionally, the framework has the potential for application in other areas, such as smart healthcare, smart transportation, and smart homes, among others.
- The IFrustum-PointNets framework innovatively expands the threshold for mask definition, addressing a common problem in existing intelligent logistics systems—imprecise mask predictions that lead to errors in object detection and tracking. By refining this aspect, IFrustum-PointNets considerably reduces inaccuracies, thereby increasing the reliability and effectiveness of logistics operations, especially in environments requiring high precision such as automated warehouses and transportation systems.
- The enhanced parallel attention mechanism integrated into IFrustum-PointNets improves the network's capacity to capture and interpret critical features from both 3D point clouds and 2D RGB images. By leveraging this dual-modality focus, the system achieves significantly stronger detection performance, enabling more accurate and robust object identification and localization. This improvement is especially valuable in dynamic operational settings, where rapid and precise object recognition is essential for ensuring smooth and efficient logistics processes.
- Incorporating Focal Loss into the IFrustum-PointNets framework addresses the pervasive issue of class imbalance in object detection. This contribution is critical as it helps the network focus more on hard-to-classify but critical objects, ensuring that these are not overshadowed by more frequent but less significant items. This enhancement is essential for maintaining operational integrity and accuracy in logistics systems, where overlooking critical but less frequent items could lead to significant disruptions.

Here is the structure for the remainder of the work. Section 2 presents some of the latest research in the field of industrial logistics networks and target detection AI. Section 3 introduces our method. Section 4 provides experimental evaluations and discussions. Section 5 is the conclusion.

ISSN 2409-7284 (Online)

2. Literature Review

2.1 Industrial Internet of Things

The development of the Industrial Internet of Things (IIoT) has made significant progress, connecting sensors, machinery, and computing devices, thereby greatly enhancing the automation and intelligence of industrial production and operations. Intelligent logistics, as one of the key application areas of IIoT, focuses on automating goods tracking, inventory monitoring, and transportation resource scheduling through connected devices[13]. In the research of intelligent logistics, several common methods are employed:Firstly, data mining and predictive analytics utilize big data analysis techniques to uncover patterns and trends in historical data, predicting future goods flow and demand to optimize logistics scheduling and resource allocation. Secondly, intelligent sensor technology application involves deploying smart sensors to monitor the real-time location, temperature, humidity, and other information of goods, enabling precise monitoring and management of the logistics process[14]. Thirdly, machine vision and image recognition utilize computer vision technology to analyze and recognize images and videos in logistics scenarios, enabling goods identification, classification, and tracking, thereby enhancing the automation level of logistics operations. Additionally, intelligent optimization algorithms utilize mathematical modeling and optimization theory to design intelligent algorithms for optimizing logistics networks, including path planning, inventory management, and transportation scheduling, to improve logistics efficiency and reduce costs. Lastly, artificial intelligence and machine learning leverage AI and ML technologies to analyze and learn from logistics data, discovering patterns and optimizing strategies to enable intelligent decision support and adaptive adjustments[15].

These methods combine IIoT technology with modern information technology, providing crucial support and assurance for achieving logistics automation and intelligence[16]. With the continuous development and innovation of technology, intelligent logistics will play an increasingly important role in enhancing logistics efficiency, reducing costs, and meeting the growing demands of logistics[17].

2.2 3D Object Detection

In intelligent logistics management, precise and efficient 3D object detection methods are integral components of the Industrial Internet of Things (IIoT). Researchers have proposed various methods, each building upon the previous one, to meet the demand for accurate object recognition in logistics environments. Firstly, LiDAR-based object detection methods[18], utilizing laser detection and ranging (LiDAR) technology, can generate high-precision point cloud data, providing accurate spatial positioning and size measurement of objects. However, their real-time performance in dynamic logistics environments still needs improvement. Secondly, grid-based object detection methods partition space into regular grids[19], employing computer vision algorithms for faster processing and recognition speed. Yet, they may overlook small or irregularly shaped objects in complex logistics scenes. Additionally, point cloud and voxel-based object detection methods convert

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

point cloud data into voxel representations, improving processing speed and recognition accuracy, but still face challenges when dealing with occluded or overlapping objects[20].

Deep learning-based object detection methods, such as convolutional neural networks (CNNs), enhance detection performance in various logistics environments by learning complex features from 3D data. PointNet[21], as a point cloud-based deep learning method, directly processes raw point cloud data without the need for voxel or grid representations, capturing global features and distinguishing between different objects effectively. PointNet++[22], an extension of PointNet, further improves feature learning and capturing capabilities by introducing hierarchical structures, resulting in enhanced accuracy in object detection and classification. Frustum-PointNets combines point cloud and image information for object detection[23], generating frustums from image data captured by cameras or sensors and then inputting the point cloud data within these frustums into PointNet or other deep learning models for object detection and recognition. This approach maximizes the utilization of both image and point cloud information, thereby improving the accuracy and robustness of object detection. However, despite the potential of these new methods for intelligent logistics management, challenges remain, such as the efficiency of handling large-scale data, real-time processing, and the generalization ability of models, which require further research and improvement.

3. Method

In this paper, we propose an IoT framework specifically designed for smart logistics systems, aimed at establishing a comprehensive and effective architecture. As shown in Figure 1, this framework consists of four key layers, each optimized for specific needs within smart cities and logistics systems:

- Big Data Layer: The primary responsibility of this layer is to process and analyze massive volumes of data. It not only stores a vast amount of information but also employs advanced data analytics technologies, such as machine learning and data mining, to uncover patterns and trends within the data, thereby providing a scientific basis for high-level decision-making.
- Edge Data Layer: Located at the network edge, this layer primarily handles real-time data close to the data sources. Processing data near its origin significantly reduces latency and speeds up data processing, while also alleviating the load on central data storage and processing facilities. This is particularly crucial for logistics operations that require quick decision-making and responsiveness.
- IoT Layer: Composed of various interconnected physical devices, sensors, and actuators, this layer continuously collects critical data throughout the logistics chain. The design of this layer enables real-time monitoring of logistics activities, from vehicle tracking to inventory management, with all devices aimed at optimizing logistics and supply chain management through precise data exchange.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

- Deployment Layer: This layer is primarily responsible for deploying and maintaining the IFrustum-Pointnets Network, an innovative network model designed to process and analyze complex data related to logistics. The deployment of this model not only enhances the system's processing capabilities but also, due to its versatility, can be applied in other areas that require precise data analysis and real-time decision-making support.

By integrating these four layers, our framework not only improves data processing efficiency and system responsiveness but also enhances the scalability and flexibility of the system, providing solid technical support for the development of smart logistics. This multi-layered system design ensures that every step from data generation to decision implementation is effectively managed and optimized, thereby driving innovation and progress in the field of smart logistics.

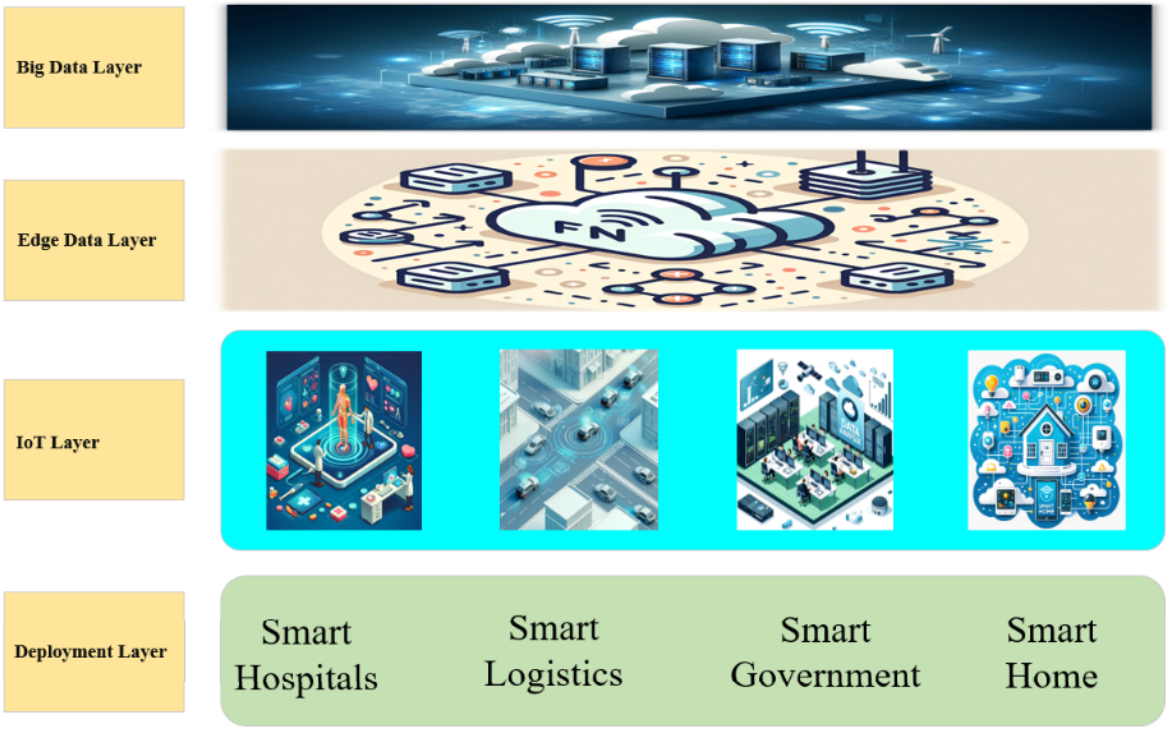

Figure 1. The proposed smart Logistics framework.

3.1 Overview of Our Network

In this paper, we improved the Frustum-Pointnets network and proposed IFrustum-Pointnets. IFrustum-Pointnets introduces new designs based on the original network structure, aimed at enhancing the network's detection accuracy and robustness. Specifically, in the mask prediction stage, we found that using a threshold of 0.5 to separate foreground and background was not ideal. Therefore, we adjusted the thresholding strategy to select a more suitable threshold, thereby improving the stability and accuracy of target detection.

Moreover, considering the large volume of point cloud data, we realized that using only cross-entropy loss was insufficient. Therefore, we proposed the idea of combining point cloud data with the 2D detection results of RGB images. Specifically, we used the point clouds of each target's frustum as a reference to obtain the mask of predicted objects and derive corresponding target frontal point

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

cloud data based on this mask. Subsequently, we input the acquired target foreground information and feature information into the 3D detection framework to achieve precise target detection and recognition.

Overall, a series of enhancements were implemented to improve model performance. First, the mask threshold definition was expanded to more effectively encompass target objects. Second, the parallel attention mechanism was optimized to boost both efficiency and performance in multitask processing. Finally, the original loss function was replaced with Focal Loss to mitigate class imbalance in the dataset, thereby strengthening the model's robustness and accuracy. The corresponding network structure is illustrated in Figure 2.

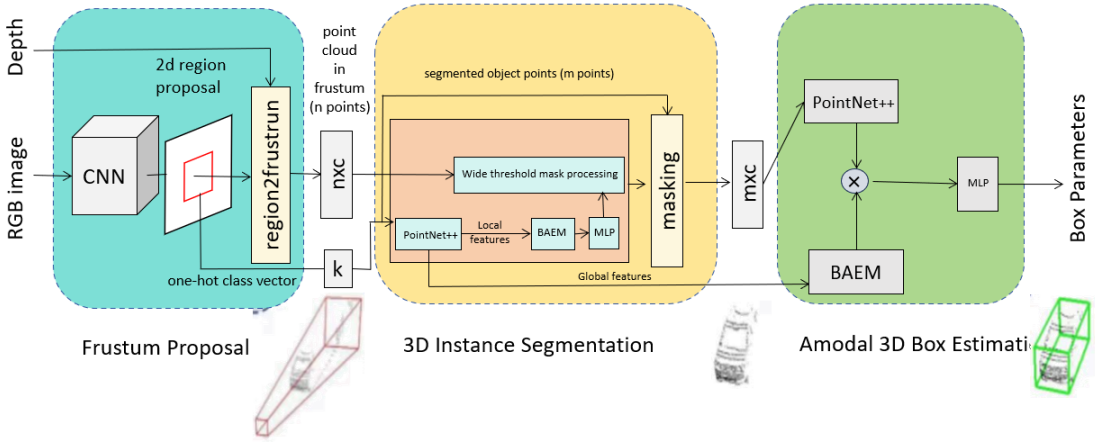

Figure 2. IFrustum-Pointnets Network Architecture Diagram.

3.2 Frustum-Pointnets

The Frustum-PointNets method innovatively combines 2D CNN and 3D point cloud processing, providing effective support for intelligent logistics in the context of Industrial IoT[24]. Firstly, leveraging 2D Convolutional Neural Networks (CNNs), Frustum-PointNets extract two-dimensional regions from RGB images and classify the contents within these regions, accurately identifying objects in the image and determining their positions. Subsequently, these two-dimensional regions are extended into three-dimensional space, forming what are known as "frustums." By combining the object position and orientation information from images with point cloud data, these frustums represent the position and orientation of objects in three-dimensional space. This approach is directly applicable to intelligent logistics management, facilitating precise identification and tracking of goods' positions and statuses in industrial settings. By integrating information from both two-dimensional images and three-dimensional point clouds, Frustum-PointNets provide crucial support for intelligent logistics systems, enabling real-time monitoring and positioning of goods, thereby enhancing the intelligence of logistics management and optimizing supply chain efficiency and responsiveness.

We replaced the feature extractor of Frustum-PointNets with PointNet++, which can further improve the efficiency and accuracy of processing 3D point cloud data. As shown in Figure 3,

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

PointNet++ is a deep learning network designed for point cloud data, capable of effectively learning features from point clouds to provide more accurate and rich representations. By replacing the feature extractor of Frustum-PointNets with PointNet++, we can better capture key features in 3D point cloud data, thereby enhancing the accuracy of object recognition and localization. Additionally, PointNet++ exhibits good scalability and generalization capabilities, enabling it to adapt to various scene and object recognition tasks. Consequently, Frustum-PointNets can achieve excellent performance in various industrial IoT application scenarios. This replacement of the feature extractor further strengthens the application prospects of Frustum-PointNets in intelligent logistics management, bringing more efficient and intelligent solutions to industrial production and operations.

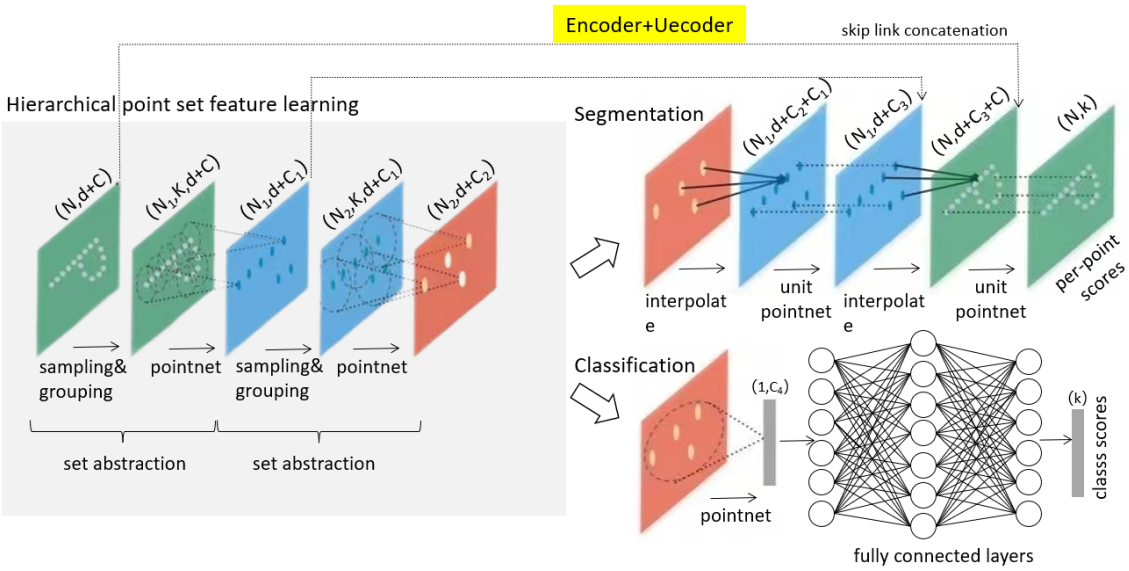

Figure 3. Bidirectional Attention Extraction Mechanism.

3.3 Frustum-Pointnets

We introduce the Bidirectional Attention Extraction Mechanism (BAEM), a novel approach that achieves significant breakthroughs in the fusion of 2D image and 3D point cloud data. First, a 2D Convolutional Neural Network (CNN) is employed to extract features from RGB images, capturing essential visual information. By incorporating channel and spatial attention mechanisms, the model effectively emphasizes semantically salient regions in the images, enhancing discriminative feature extraction. These attention mechanisms are then transferred to the corresponding 3D point cloud data, enabling focused extraction of critical features within the 3D domain. This transfer allows informative image features to enrich the representation of 3D structures. Finally, the feature representations from both 2D and 3D modalities are fused into a unified representation, providing more comprehensive and discriminative features for downstream object detection tasks. The proposed BAEM effectively leverages cross-modal correlations between imaging and point cloud data, leading to notable improvements in detection accuracy and robustness.

The BAEM network architecture is illustrated in Figure 4.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

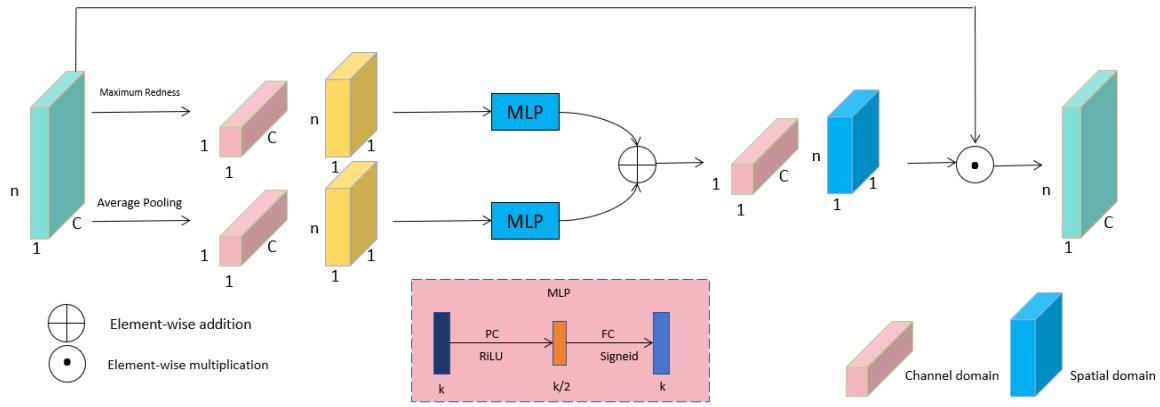

Figure 4. BAEM Network Architecture Diagram.

BAEM follows the process outlined below.

$$M_c = \sigma(W_c * X) \odot X \quad \text{[Formular 1]}$$

where W_c represents the channel attention map, X denotes the input feature map, σ denotes the sigmoid activation function, W_c denotes the learnable weight matrix, and $\odot$ denotes the element-wise multiplication operation.

$$M_s = \text{softmax}(W_s * X) \quad \text{[Formular 2]}$$

where W_s represents the spatial attention map, and W_s denotes the learnable weight matrix.

$$X = M_c \odot M_s \odot X \quad \text{[Formular 3]}$$

where X represents the attended feature map.

$$W_i = \text{softmax}(V_{\text{att}} \cdot \tanh(U_{\text{att}} * X_{3D})) \quad \text{[Formular 4]}$$

where W_i represents the attention weight vector, U_{att} and V_{att} are the learnable weight matrices, $\cdot$ denotes the dot product operation, and $\tanh$ denotes the hyperbolic tangent activation function.

$$X_{att} = \sum_i W_i \odot X_{3D} \quad \text{[Formular 5]}$$

where X_{att} represents the attention-enhanced feature map.

$$X_{fuse} = [X, X_{att}] \quad \text{[Formular 6]}$$

where X_{fuse} denotes the fused feature map, and $[X, X_{att}]$ denotes the concatenation operation.

3.4 Focal Loss

The introduction of Focal Loss marks the adoption of a new strategy in our mask prediction task. Compared to traditional cross-entropy loss functions, Focal Loss offers a more flexible and effective approach to addressing class imbalance issues. In traditional cross-entropy loss functions, the inadequate attention given to hard-to-classify positive samples leads to poor classification performance, especially when there is a large number of easily classifiable negative samples present. Focal Loss dynamically adjusts the weights of the loss function to focus more on hard-to-classify samples, thereby effectively enhancing the model's learning ability for these samples. In the mask

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

prediction task, this mechanism is particularly important as it enables the model to better handle complex spatial configurations and occlusion scenarios. By introducing Focal Loss, our model can better learn from these challenging samples, thereby improving the accuracy of mask prediction. This innovative loss function allows our model to more accurately delineate object boundaries and capture fine details, thereby providing a significant performance boost for the overall object detection task.The calculation of Focal Loss is as follows.

$$L_{mask} = L_{FL} = \begin{cases} -\alpha(1-p)^{\gamma}\log(p), & if - y = 1 \\ -(1-\alpha)p^{\gamma}\log(p), & otherwise \end{cases} \quad \text{[Formular 7]}$$

In this context, when $y = 1$, it signifies that the point is classified as foreground, with p denoting the probability assigned to this classification. Here, α serves as the weight factor, adjusting the emphasis placed on positive and negative samples, while γ acts as the modulation factor, controlling the rate of decrease in sample weights.

3.4 3D Detection Box Regression Network

We have improved the method for three-dimensional bounding box detection by utilizing point cloud data and global features obtained from a three-dimensional instance segmentation network. Firstly, we reposition the origin of the point cloud coordinates to the center of the masked point cloud data, enabling better alignment with the position and orientation of the target objects. Subsequently, we adopted PointNet++ as the feature extractor and utilized MLP for regression to predict the parameters of the three-dimensional bounding boxes. PointNet++ is effective in handling point cloud data, thereby extracting high-quality features suitable for the detection task. Through this approach, we can more accurately localize and describe the targets, thereby enhancing the overall detection performance of the system.

4. Experiments

4.1 Dataset

We utilized the KITTI 3D Object Detection Benchmark dataset[25], widely employed for evaluating the performance of 3D object detection algorithms. This dataset comprises a vast collection of real-world images and point cloud data. Specifically, in this experiment, the training set consists of approximately 7,481 images/point clouds, while the testing set comprises around 7,518 images/point clouds. Additionally, we partitioned about 3,712 images/point clouds from the training set to form the validation set, aiding in a more thorough evaluation of the model's performance. Furthermore, the target objects in the dataset are primarily categorized into three classes: Cars, Pedestrians, and Cyclists. These categories encompass the most common objects in urban traffic scenarios, allowing for a comprehensive assessment of algorithm performance across various contexts. Partial data visualization examples are shown in Figure 5.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

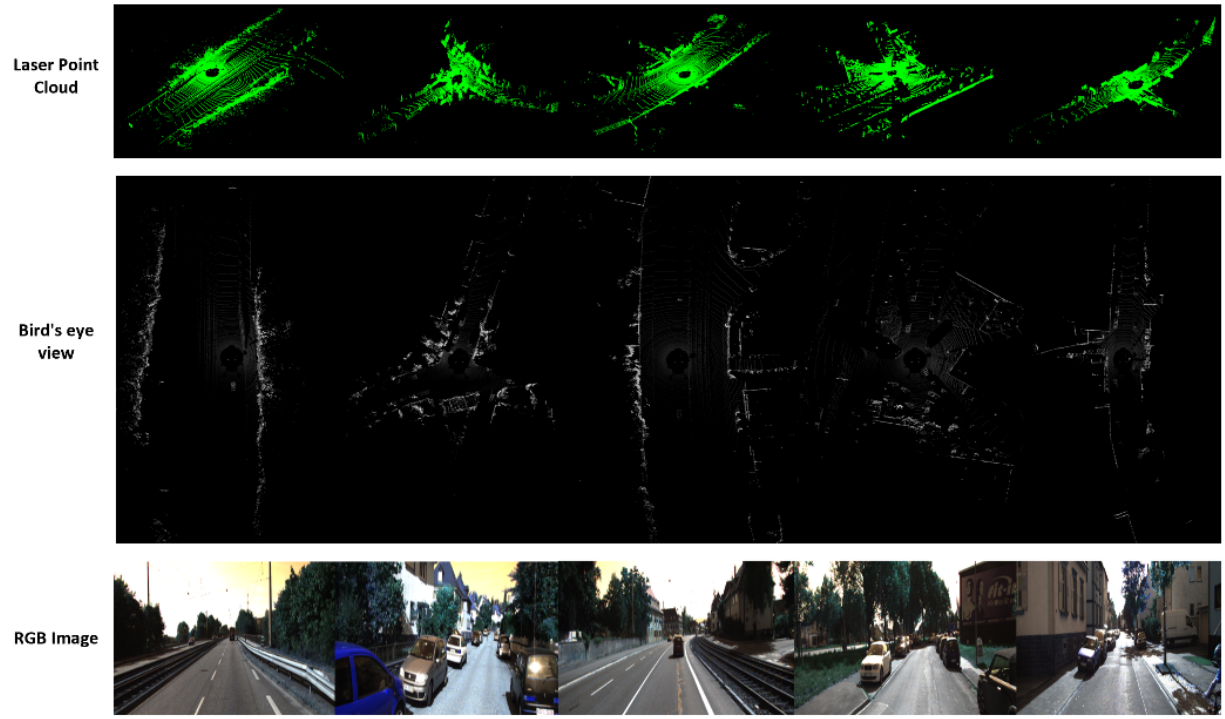

Figure 5. Example Demonstration of KITTI 3D Object Detection Benchmark Dataset, with Point Cloud Data on Top, Bird's-Eye View in the Middle, and RGB Image at the Bottom.

4.2 Comparison Method

Our comparison includes several algorithms that perform well in the category of cars. Among them, the image-based methods include Mono3D and 3DOP, which utilize a single image for object detection and localization. Additionally, LiDAR-based methods such as VeloFCN and 3D-FCN use LiDAR data for object detection. Furthermore, we also consider MV and Frustum-PointNets methods, which are multimodal approaches combining image and point cloud data for object detection. By comparing with these high-performing algorithms, we can more comprehensively evaluate the performance of our proposed method in the task of car object detection.

4.3 Evaluation metrics

To measure the accuracy of the models, this paper adopts the Average Precision (AP) as the primary metric and applies it to evaluate the average detection precision of individual categories by the model. AP is a commonly used evaluation metric for object detection, considering the precision performance of the model at different confidence thresholds and calculating a comprehensive accuracy score. By using AP as the evaluation metric, we can gain a more comprehensive understanding of the model's detection performance on different categories, enabling better comparison and evaluation of different methods.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \qquad \text{[Formular 8]}$$

where:True Positives : Number of correctly predicted positive instances.False Positives : Number of incorrectly predicted positive instances

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

[Formular 9]

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{[Formular 10]}$$

where:True Positives : Number of correctly predicted positive instances.False Positives : Number of incorrectly predicted positive instances

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \quad \text{[Formular 11]}$$

where:N : Number of classes.AP : Average Precision for class i.

4.4 Experimental Environment

The experimental environment for this experiment is shown in Table 1, which lists the hardware and software requirements for the experiment.

Table 1. Experimental Environment Details

Hardware/Software	Specification
CPU	Intel i9 7900X
GPU	A100
Memory	80G
Storage	12 $\times$ 5TB
Operating System	Ubuntu 22.1
Deep Learning Framework	PyTorch 1.10.1
Python Version	Python 3.8
CUDA	11.7

4.5 Results

Comparison with state-of-the-art results . As shown in Table 2, the performance comparison in terms of average precision (AP) for 3D detection on the KITTI validation set demonstrates outstanding results for Frustum-PointNets and the improved Frustum-PointNets (IFrustum-PointNets) across the categories of cars, pedestrians, and cyclists. Specifically, in the car detection category, IFrustum-PointNets achieved AP scores of 89.71%, 84.79%, and 79.41% for easy, moderate, and hard difficulty levels, respectively. Compared to other methods like Mono3D, 3DOP, and VeloFCN, which achieved maximum accuracies of 5.22%, 12.63%, and 40.14%, our method shows significant superiority. Even when compared to VoxelNet, which is also LiDAR-based, our method performs better on both the easy and hard difficulty levels, particularly outperforming VoxelNet by about 0.2 percentage points on the moderate difficulty level.

In the detection of pedestrians and cyclists, our method also exhibits exceptional performance.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Notably, in the pedestrian detection for moderate and hard difficulty levels, IFrustum-PointNets consistently achieved more than 61% AP, which is over 2.5 percentage points higher than the maximum accuracy of HC-baseline. For cyclist detection, our AP remains stable at over 50\%, marking a significant advance for detecting small objects in complex traffic scenarios.Overall, Frustum-PointNets and the enhanced IFrustum-PointNets show extremely high accuracy in 3D object detection, significantly outperforming other comparative methods, especially in moderate and hard detection tasks. These results prove the potential application of our method in smart logistics management systems, providing strong technical support for achieving more efficient and precise tracking and monitoring of goods.

Table 2. Performance comparison in 3D detection: average precision (in %) on KITTI validation set.

Method	Modality	Car			Pedestrian			Cyclist		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
Mono3D[26]	Mono	5.22	5.19	4.13	N/A	N/A	N/A	N/A	N/A	N/A
3DOP[27]	Stereo	12.63	9.49	7.59	N/A	N/A	N/A	N/A	N/A	N/A
VeloFCN[28]	LiDAR	40.14	32.08	30.47	N/A	N/A	N/A	N/A	N/A	N/A
MV (BV+FV) }[29]	LiDAR	86.18	77.32	76.33	N/A	N/A	N/A	N/A	N/A	N/A
MV (BV+FV+RGB) [29]	LiDAR+Mono	86.55	78.10	76.67	N/A	N/A	N/A	N/A	N/A	N/A
BriNet[30]	LiDAR	88.26	78.42	77.66	58.96	53.79	51.47	63.63	42.75	41.06
VoxelNet[31]	LiDAR	89.60	84.81	78.57	65.95	61.05	56.98	74.41	52.18	50.49
Frustum-Pointnets[24]	LiDAR	84.38	82.31	76.58	64.93	60.38	53.42	70.40	50.38	51.42
IFrustum-Pointnets	LiDAR	89.71	84.79	79.41	66.45	61.05	57.88	73.31	53.19	51.35

Visual demonstration. We present several examples of 3D detections in Figure 6. For better visualization, the 3D bounding boxes detected by LiDAR are projected onto the RGB images. As illustrated, IFrustum-Pointnets in combination with LiDAR provides highly accurate 3D bounding boxes across all categories. This approach allows us to see the location and size of objects in three-dimensional space and also provides an intuitive visual confirmation in two-dimensional images, greatly enhancing the intuitiveness and interpretability of the detection results. These examples clearly demonstrate that our method can achieve precise detection and positioning of various objects, regardless of their size, in complex urban traffic environments.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Figure 6. Visualization Showcase of iFrustrum-PointNets.

4.6 Ablation experiments

Margin Ablation Study. As shown in Table 3, during the ablation experiments conducted to assess the impact of different Margin values on model performance, we found that the model exhibited better performance across various metrics when the Margin was set to 0.2. This result reveals that an appropriate increase in the Margin has a positive effect on enhancing the model's discriminative ability. Specifically, with the Margin set at 0.2, the average precision (AP) for the vehicle category at easy, moderate, and hard difficulty levels were 83.04%, 70.95%, and 63.55%, respectively. For the pedestrian category, the APs were 67.05%, 59.28%, and 52.16%, and for the cyclist category, the APs were 76.33%, 57.15%, and 53.66%. Compared to other Margin values, the setting of 0.2 achieved higher precision in most cases, particularly standing out at the moderate difficulty level, indicating that fine-tuning the Margin is crucial for enhancing model performance.

Table 3. Performance Metrics Across Different Margins.

Margin	Car			Pedestrian			Cyclist		
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
0	82.73	69.28	62.54	65.84	58.49	51.25	74.29	55.46	52.65
0.1	82.85	69.62	62.59	61.80	55.40	49.33	73.56	55.98	53.01
0.2	83.04	70.95	63.55	67.05	59.28	52.16	76.33	57.15	53.66
0.3	83.21	70.63	63.20	65.06	57.43	50.79	74.02	56.07	52.82

Ablation Study of Different Components. As shown in Table 4, these experiments demonstrated how different combinations affect car detection performance. Our model utilized three key components to enhance its performance: Triplet Loss (with a margin of 0.2), BAEM, and Focal Loss. The addition or removal of each component had a significant impact on the model's accuracy.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Initially, we considered a baseline model without any advanced loss functions or regularization techniques, which achieved average precision (AP) scores of 82.15%, 68.55%, and 62.44% for the Easy, Moderate, and Hard difficulty levels, respectively. With the sole addition of Focal Loss, the model's performance improved across all difficulty levels, especially at Moderate and Hard levels, with increases of 11.4% and 11.01%, respectively. This confirmed the effectiveness of Focal Loss in dealing with hard-to-detect objects in imbalanced datasets. Next, we incorporated BAEM alone. This method also showed a relative improvement at the Moderate and Hard levels, with APs reaching 79.33% and 72.86%. Although the increases were not as significant as with Focal Loss, it still underscored the importance of regularization techniques in preventing overfitting. Considering the impact of Triplet Loss, its standalone use resulted in a slight improvement in detection performance at the Moderate level and a more substantial increase to 73.51% at the Hard level. Ultimately, by combining all three techniques, our model exhibited significant performance improvements across all difficulty levels, impressively reaching 84.35% on the Moderate level and 79.01% on the Hard level. These results clearly demonstrated the crucial role of integrating these technologies in enhancing the accuracy of the 3D detection model.

Table 4. Performance Metrics for Car Detection.

Triplet Loss (Margin=0.2)	BAEM	Focal Loss	Easy	Moderate	Hard
-	-	-	82.15	68.55	62.44
-	-	√	82.85	79.95	73.45
-	√	-	82.06	79.33	72.86
√	-	-	82.88	71.02	73.51
√	√	√	89.14	84.35	79.01

5. Conclusions

The IFrustum-Pointnets introduced in this paper is groundbreaking in the field of the Industrial Internet of Things (IIoT), particularly in enhancing the accuracy of object detection in smart logistics management. However, we recognize that there are areas where the model still falls short. First, the current model's robustness when processing extremely sparse point cloud data needs enhancement; second, the model's adaptability to objects of varying scales could be improved. In the future, we plan to focus on two main areas for optimization: first, developing more efficient algorithms to enhance the model's capability to handle sparse point clouds; second, exploring scale-adaptive mechanisms to improve the model's detection precision for objects of various sizes. Through these efforts, we hope to further advance the development of IIoT and provide solid technological support for achieving more intelligent industrial systems.

References

[1] R. A. Khalil, N. Saeed, M. Masood, Y. M. Fard, M.-S. Alouini, and T. Y. Al-Naffouri, “Deep

ISSN 2409-7284 (Online)

learning in the industrial internet of things: Potentials, challenges, and emerging applications," *IEEE Internet of Things Journal,* vol. 8, no. 14, pp. 11016-11040, 2021.

[2] P. K. Malik, R. Sharma, R. Singh, A. Gehlot, S. C. Satapathy, W. S. Alnumay, D. Pelusi, U. Ghosh, and J. Nayak, "Industrial Internet of Things and its applications in industry 4.0: State of the art," *Computer Communications,* vol. 166, pp. 125-139, 2021.

[3] W. Z. Khan, M. Rehman, H. M. Zangoti, M. K. Afzal, N. Armi, and K. Salah, "Industrial internet of things: Recent advances, enabling technologies and open challenges," *Computers & electrical engineering,* vol. 81, pp. 106522, 2020.

[4] S. Latif, Z. Zou, Z. Idrees, and J. Ahmad, "A novel attack detection scheme for the industrial internet of things using a lightweight random neural network," *IEEE access,* vol. 8, pp. 89337-89350, 2020.

[5] T. Xie, and X. Yao, "Smart logistics warehouse moving-object tracking based on YOLOv5 and DeepSORT," *Applied Sciences,* vol. 13, no. 17, pp. 9895, 2023.

[6] R. C. Joshi, S. Yadav, M. K. Dutta, and C. M. Travieso-Gonzalez, "Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people," *Entropy,* vol. 22, no. 9, pp. 941, 2020.

[7] J. Guerrero - Ibañez, J. Contreras - Castillo, and S. Zeadally, "Deep learning support for intelligent transportation systems," *Transactions on Emerging Telecommunications Technologies,* vol. 32, no. 3, pp. e4169, 2021.

[8] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge computing in industrial internet of things: Architecture, advances and challenges," *IEEE communications surveys & tutorials,* vol. 22, no. 4, pp. 2462-2488, 2020.

[9] L. Du, X. Ye, X. Tan, J. Feng, Z. Xu, E. Ding, and S. Wen, "Associate-3Ddet: Perceptual-to-conceptual association for 3D point cloud object detection." pp. 13329-13338.

[10] D. Fernandes, A. Silva, R. Névoa, C. Simões, D. Gonzalez, M. Guevara, P. Novais, J. Monteiro, and P. Melo-Pinto, "Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy," *Information Fusion,* vol. 68, pp. 161-191, 2021.

[11] C. R. Qi, X. Chen, O. Litany, and L. J. Guibas, "Imvotenet: Boosting 3d object detection in point clouds with image votes." pp. 4404-4413.

[12] Z. Liu, X. Zhao, T. Huang, R. Hu, Y. Zhou, and X. Bai, "Tanet: Robust 3d object detection from point clouds with triple attention." pp. 11677-11684.

[13] S. Din, A. Paul, A. Ahmad, B. B. Gupta, and S. Rho, "Service orchestration of optimizing continuous features in industrial surveillance using big data based fog-enabled internet of things," *IEEE Access,* vol. 6, pp. 21582-21591, 2018.

[14] W. A. Jabbar, C. W. Wei, N. A. A. M. Azmi, and N. A. Haironnazli, "An IoT Raspberry Pi-based parking management system for smart campus," *Internet of Things,* vol. 14, pp. 100387, 2021.

[15] K. A. Abuhasel, and M. A. Khan, "A secure industrial internet of things (IIoT) framework for resource management in smart manufacturing," *IEEE Access,* vol. 8, pp. 117354-117364,

2020.

[16] A. Kumar, Z. J. Zhang, and H. Lyu, “Object detection in real time based on improved single shot multi-box detector algorithm,” *EURASIP Journal on Wireless Communications and Networking,* vol. 2020, no. 1, pp. 204, 2020.

[17] I. Ahmed, G. Jeon, and F. Piccialli, “A deep-learning-based smart healthcare system for patient’s discomfort detection at the edge of internet of things,” *IEEE Internet of Things Journal,* vol. 8, no. 13, pp. 10318-10326, 2021.

[18] C. Wisultschew, G. Mujica, J. M. Lanza-Gutierrez, and J. Portilla, “3D-LIDAR based object detection and tracking on the edge of IoT for railway level crossing,” *IEEE Access,* vol. 9, pp. 35718-35729, 2021.

[19] G. R. Venkatakrishnan, R. Ramasubbu, and R. Mohandoss, “An efficient energy management in smart grid based on IOT using ROAWFSA technique,” *Soft Computing,* vol. 26, no. 22, pp. 12689-12702, 2022.

[20] K. Kanna, K. A. Lachguer, and R. Yaagoubi, “MyComfort: An integration of BIM-IoT-machine learning for optimizing indoor thermal comfort based on user experience,” *Energy and Buildings,* vol. 277, pp. 112547, 2022.

[21] J. Xie, Y. Xu, Z. Zheng, S.-C. Zhu, and Y. N. Wu, "Generative pointnet: Deep energy-based learning on unordered point sets for 3d generation, reconstruction and classification." pp. 14976-14985.

[22] G. Qian, Y. Li, H. Peng, J. Mai, H. Hammoud, M. Elhoseiny, and B. Ghanem, “Pointnext: Revisiting pointnet++ with improved training and scaling strategies,” *Advances in neural information processing systems,* vol. 35, pp. 23192-23204, 2022.

[23] A. Paigwar, D. Sierra-Gonzalez, Ö. Erkent, and C. Laugier, "Frustum-pointpillars: A multi-stage approach for 3d object detection using rgb camera and lidar." pp. 2926-2933.

[24] E. Erçelik, E. Yurtsever, and A. Knoll, "Temp-frustum net: 3d object detection with temporal fusion." pp. 1095-1101.

[25] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The kitti dataset,” *The international journal of robotics research,* vol. 32, no. 11, pp. 1231-1237, 2013.

[26] C. Yan, and E. Salman, “Mono3D: Open source cell library for monolithic 3-D integrated circuits,” *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 65, no. 3, pp. 1075-1085, 2017.

[27] Y. Zheng, B. Shyrokau, and T. Keviczky, "3DOP: Comfort-oriented motion planning for automated vehicles with active suspensions." pp. 390-395.

[28] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving." pp. 1907-1915.

[29] Y. Zhou, and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection." pp. 4490-4499.

[30] J. Beltrán, C. Guindel, F. M. Moreno, D. Cruzado, F. Garcia, and A. De La Escalera, "Birdnet: a

ISSN 2409-7284 (Online)

3d object detection framework from lidar information." pp. 3517-3523.

[31] Y. Chen, J. Liu, X. Zhang, X. Qi, and J. Jia, "Voxelnext: Fully sparse voxelnet for 3d object detection and tracking." pp. 21674-21683.