Towards Robust 3D Object Detection with LiDAR and 4D Radar Fusion in Various Weather Conditions

Yujeong Chae
KAIST
yujeong@kaist.ac.kr

Hyeonseong Kim
KAIST
brian617@kaist.ac.kr

Kuk-Jin Yoon
KAIST
kjyoon@kaist.ac.kr

Abstract

Detecting objects in 3D under various (normal and adverse) weather conditions is essential for safe autonomous driving systems. Recent approaches have focused on employing weather-insensitive 4D radar sensors and leveraging them with other modalities, such as LiDAR. However, they fuse multi-modal information without considering the sensor characteristics and weather conditions, and lose some height information which could be useful for localizing 3D objects. In this paper, we propose a novel framework for robust LiDAR and 4D radar-based 3D object detection. Specifically, we propose a 3D-LRF module that considers the distinct patterns they exhibit in 3D space (e.g., precise 3D mapping of LiDAR and wide-range, weather-insensitive measurement of 4D radar) and extract fusion features based on their 3D spatial relationship. Then, our weather-conditional radar-flow gating network modulates the information flow of fusion features depending on weather conditions, and obtains enhanced feature that effectively incorporates the strength of two domains under various weather conditions. The extensive experiments demonstrate that our model achieves SoTA performance for 3D object detection under various weather conditions.

1. Introduction

Detecting 3D objects, which aims to classify the objects and localize them in 3D coordinates, plays a crucial role in various applications such as autonomous driving, robotic, and drone systems [1, 8, 44]. Many attempts have been made to utilize various sensors, such as camera, LiDAR, and radar, for 3D object detection [9, 11, 16, 24, 31, 51, 53]. These methods are typically trained and tested in ideal autonomous driving scenarios, demonstrating satisfactory performance under normal conditions. Since real-world driving situations have diverse weather conditions, robust models operating in various conditions are needed.

*Code: https://github.com/yujeong-star/RL_3DOD.

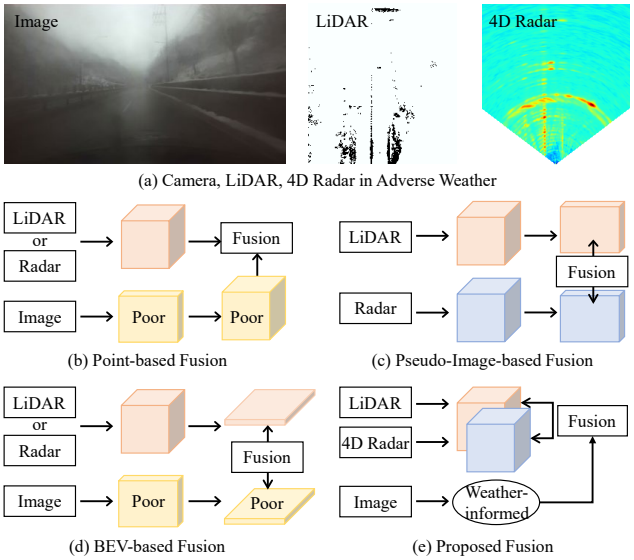

Figure 1. In adverse weather conditions, as depicted in (a), radar exhibits the highest robustness, followed by LiDAR, while the image is significantly degraded. Prior multi-modal 3D object detection research follows the fusion method in (b)-(d). They suffer from inaccurate information from images or sub-optimal performance due to the compression of critical 3D information from LiDAR and radar into BEV or pseudo-images for fusion. In contrast, our approach (e) effectively fuses LiDAR and 4D radar in 3D space, taking the strengths of each sensor through weather information, showing robust performance under adverse conditions.

Recently, several research has focused on addressing these challenges by employing radar sensors capable of handling various weather conditions [28], and has released datasets containing diverse weather environments [4, 28, 35]. Moreover, research on a new novel sensor, 4D radar, which includes height information, has been initiated [22, 28, 38, 43]. Since the radar relies on radio waves, it has the advantages of long-range detection and robustness under adverse weather conditions. However, it does not provide precise distance or detailed 3D maps and struggles with standalone deployment [12, 19, 43]. Therefore, ongoing

research endeavors are currently leveraging various multi-modal sensors to compensate for the limitations of radar and mitigate the shortcomings of LiDAR or image sensors in diverse weather conditions [6, 12, 14, 19, 25, 33]. While these studies have proposed new fusion methods, they still exhibit the following limitations.

The primary one is that, although each sensor shows robustness to specific weather conditions, prior works fuse multi-modal information without considering weather conditions and each sensor's characteristics. For instance, radar is mainly unaffected by various weather conditions, while cameras experience severe corruption in almost all adverse weather scenarios. LiDAR measurements become inaccurate in snow or rain due to changes in the reflection of laser signals, however, they are less impacted by overcast or severe light conditions. An adaptive fusion of semantic information of the camera and precise mapping capabilities of LiDAR, tailored to the climatic strengths of each sensor, would yield a more effective multi-modal 3D object detection model. The second limitation is that, despite the inherently 3D nature of the LiDAR and radar data, multi-modal fusion is not conducted in the 3D domain. As shown in Fig. 1, previous studies in adverse weather conditions transform LiDAR features to range view, pseudo-image and BEV features before fusing its information with other modalities. It causes the loss of crucial height information critical for effective 3D object detection. Besides, 4D radar aids in recognizing scene information in the 3D domain in conjunction with LiDAR [43]. However, there is currently a lack of research on fusing two domains for 3D object detection in adverse weather.

To overcome these limitations, we propose a novel method for a robust LiDAR- and 4D radar-based 3D object detection framework that considers the weather conditions and 3D domain fusion. Specifically, our framework first takes LiDAR point cloud and 4D radar tensor as input, and encodes voxel features for each modality to preserve 3D information. While extracting the features, our 3D LiDAR and 4D Radar Fusion (3D-LRF) module queries non-empty LiDAR voxels, groups the neighbor 4D radar voxels, and extracts the fusion feature at each layer. When there are few radar voxel features around the location of a LiDAR voxel feature, our 3D-LRF module discerns the corresponding LiDAR feature as imprecise, suppressing it. Conversely, when LiDAR has many neighbor radar features, our 3D-LRF module effectively fuses features from both domains. Moreover, to account for diverse weather conditions and take the strengths of each sensor, we propose weather-conditional radar-flow gating network (WRGNet). The camera can be easily corrupted in adverse conditions but has richer semantic information about the scene than other modalities. Therefore, our WRGNet takes non-empty LiDAR, neighbor 4D radar voxel features and a simple 1D feature from the pre-trained lightweight weather classification network trained on images. 1D weather-conditioned image feature and 4D radar voxel feature are fed into the gating layer to extract radar-flow gating feature. The radar-flow gating feature is multiplied to the fusion feature of 3D-LRF module to effectively control the information flow from 4D radar to LiDAR based on the weather conditions. After going through two novel fusion modules, the BEV encoder and detection head are employed to produce final 3D object detection results.

We evaluate the performance of the proposed framework on K-Radar dataset [28], which includes 4D radar, LiDAR, and images captured under various weather conditions. The experimental results demonstrate that our method shows superior performance in detecting "Sedan" compared to previous approaches, providing evidence of effective consideration of sensor characteristics and weather information.

In summary, our main contributions are four-fold: (I) We propose a pioneering approach that fuses LiDAR and 4D radar for 3D object detection in various weather conditions. (II) We propose 3D-LRF module that effectively fuses LiDAR and 4D radar features in the 3D domain, considering the characteristics of each sensor. (III) We propose WRGNet, which modulates the flow of fusion features depending on weather conditions. (IV) We conduct extensive experiments on K-Radar dataset, showing our superior performance and validating the effects of each component.

2. Related Works

2.1. Radar-based 3D Object Detection

3D Radar-based Methods. 3D radar is composed of power measurements along azimuth (A), range (R), and Doppler (D) dimensions. Since 3D radar lacks elevation information, limited research has utilized it for 3D object detection. [3] generates point-based region proposals and refines the boxes with confidence score to predict 3D bounding boxes. [31] applies 3D CNN on radar cube, [24] applies GNN on radar tensor, [5] applies ConvLSTM to consider temporal information multi-frame radars.

4D Radar-based Methods. 4D radar provides additional elevation dimension information compared to 3D radar, which is crucial for 3D object detection. Due to the recent development of 4D radar sensors and the late release of datasets, a limited amount of research has been conducted so far. [32, 38] released new 4D radar-based 3D object detection datasets and evaluated the performance of [16] on their datasets. However, the algorithm does not consider radar's characteristics, and datasets do not include adverse weather conditions where 4D radar could demonstrate its advantages. [28, 29] encodes the 4D radar tensor using 3D sparse convolution for object detection in K-Radar dataset, which includes adverse weather conditions. [22] recently

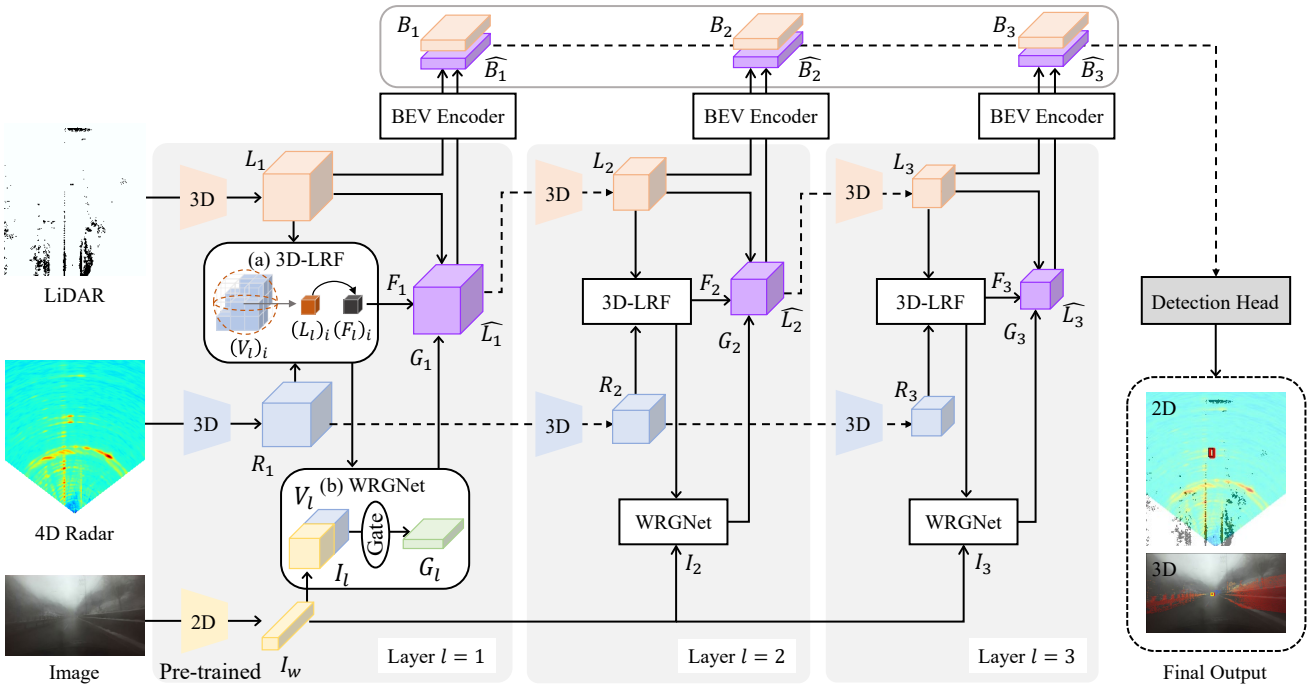

Figure 2. Overall scheme of our multi-modal 3D object detection framework in adverse weather. Our framework consists of a three-layer multi-modal fusion network, BEV encoder and detection head. In each l-th multi-modal fusion layer, proposed 3D-LRF module extract fusion feature F_l from LiDAR L_l and 4D radar feature R_l in 3D domain, considering the characteristics of each sensor. Our WRGNet further extracts enhanced LiDAR feature $\hat{L}_l$ by modulating the information flow of fusion features based on the weather conditions.

developed a method that encodes radar points with point and pillar encoders. Currently, there is a lack of literature on fusing 4D radar with other modalities while considering adverse weather conditions.

2.2. Multi-modal 3D Object Detection

Without Radar. LiDAR and camera are the two main sensors used for multi-modal 3D object detection. Two domains are diversely fused through point-based fusion [17, 41, 47, 48, 52], voxel-/pseudo-image-based fusion [2, 30, 37, 46, 49], BEV-based fusion [23, 40] and comprehensive fusion incorporating these methods [18, 36]. These models significantly degrade in adverse weather conditions.

With Radar. 3D object detection models with image and radar typically encode radar similarly to LiDAR and focus on multi-view fusion (range view of image and top view of radar) [10, 12–14, 26]. LiDAR and radar domains are fused with early fusion [27, 50] or self-/cross-attention between domains [19, 33, 43]. Besides, [26, 42, 50] focus on enhancing detection accuracy by incorporating the velocity head. InterFusion [43] utilizes 4D radar for fusion with LiDAR, however, the model is designed to encode both domains with identical pseudo-image-based structures. It did not effectively fuse the strength of each 3D domain and discern the advantages in adverse weather conditions.

2.3. 3D Object Detection in Adverse Weather

One line of research focuses on directly training the network with data on adverse weather conditions. Using only LiDAR, [20] suppresses the noisy LiDAR points under snow. Several works adaptively fuse multi-modal features through convolution layers [6, 12, 19, 33] or attention [14, 25]. These studies do not take into account the climatic strengths of individual sensors. Another line of research attempts to enable the model trained under normal conditions, to perform robustly in adverse weather. [7, 39, 45] propose unsupervised/semi-supervised domain adaptation and knowledge distillation framework for LiDAR-based 3D object detection, respectively. However, these works show under-par performance compared to supervision-based research. Unlike prior works, we propose a novel LiDAR- and 4D radar-based 3D object detection framework considering weather and effectively fusing 3D features in supervision.

3. Methods

3.1. Framework Overview

The overall scheme of our framework is shown in Fig. 2. Our framework takes LiDAR point cloud, 4D radar tensor and image as input. Specifically, the LiDAR point cloud is $L \in \mathbb{R}^{N_0 \times 3}$, the 4D radar tensor is $R \in \mathbb{R}^{M_0 \times 3}$, and

the image is $I \in \mathbb{R}^{H \times W \times 3}$, where N_0, M_0, H and W are the number of LiDAR points and 4D radar points, the height and width of image, respectively. The sparse 3D convolution network is employed as the feature extraction backbone of LiDAR and 4D radar to preserve their 3D information. L and R first pass through an input layer individually, to map the input tensor to a higher dimension of voxel features $L_0 \in \mathbb{R}^{N_0 \times C_0}$ and $R_0 \in \mathbb{R}^{M_0 \times C_0}$. Then, voxel features are fed to each three-layer sparse 3D convolution network. Each layer extracts layer-wise voxel features $L_l \in \mathbb{R}^{N_l \times C_l}$ and $R_l \in \mathbb{R}^{M_l \times C_l}$, where l is the layer index and $l \in \{1, 2, 3\}$. They are effectively fused to obtain fusion feature $F_l \in \mathbb{R}^{N_l \times C_l}$ with our 3D-LRF module (see Sec. 3.2). The image is encoded separately with a lightweight three-layer 2D convolutional network pre-trained for weather classification. Utilizing 1D weather-conditioned image feature $I_w \in \mathbb{R}^{1 \times C_l}$ together with L_l and R_l, our WRGNet controls the information flow from 4D radar to LiDAR modality through gating and gets enhanced LiDAR feature $\hat{L}_l \in \mathbb{R}^{N_l \times C_l}$ (see Sec. 3.3). While $\hat{L}_l$ and R_l serve as input of the next layer, $\hat{L}_l$ and L_l are compressed through individual BEV encoders. The concatenation of all layers' BEV features outputs 3D detection results through the detection head (see Sec. 3.4).

3.2. 3D LiDAR and 4D Radar Fusion

In this section, we introduce our 3D LiDAR and 4D Radar fusion method. Based on the fact that 3D information plays a key role in 3D object detection, we fuse 3D LiDAR and 4D Radar in 3D space. Given LiDAR features $L_l \in \mathbb{R}^{N_l \times C_l}$ and radar features $R_l \in \mathbb{R}^{M_l \times C_l}$ from l-th sparse 3D convolution layer, our goal is to enhance the LiDAR feature with the aid of the radar features. Specifically, for each LiDAR voxel feature, we first find the K_l nearest neighbor radar voxel features, $V_l \in \mathbb{R}^{N_l \times K_l \times C_l}$, within the radius r_l, to allow the model to focus on areas of interest, considering the 3D spatial relationship. As in Eq. (1), we use different numbers of neighbors K_l and radius r_l for each layer l to consider the actual size of the detection target.

$$K_l = \lfloor \frac{64}{2^{l-1}} \rfloor, \quad r_l = \lfloor \frac{8}{2^{l-1}} \rfloor. \tag{1}$$

Detailed explanations of design choices are in supplementary material. After obtaining the K_l nearest neighbor radar voxel features V_l, the LiDAR feature L_l is fused with V_l through the 3D-LRF module.

3D-LRF module. The 3D-LRF module aims to efficiently integrate LiDAR and 4D radar domains, and to achieve this, it is essential to first understand the characteristics of both domains. LiDAR, utilizing laser reflections, provides precise 3D mapping of the environment under normal conditions. However, it is susceptible to noise during adverse weather, such as heavy snow or rain. On the other hand, radar employs waves for measurement, offering robustness in inclement weather. Nevertheless, it falls short of providing accurate object positions. Therefore, by discerning whether each position is an object in the scene or noise caused by adverse weather through surrounding radar, activating or suppressing the corresponding LiDAR accordingly becomes possible.

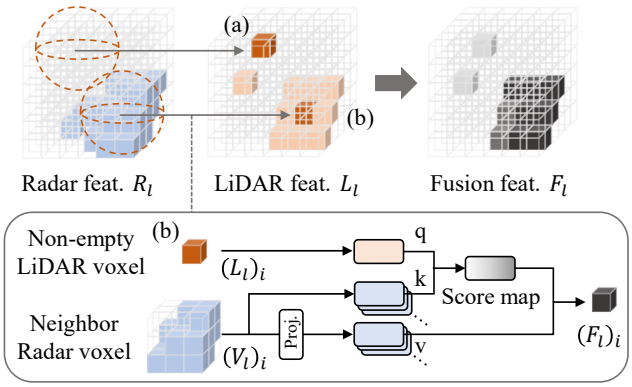

Figure 3. Illustration of proposed 3D-LRF module. Non-empty LiDAR voxel and neighbor radar voxel go through attention mechanism to enhance or suppress the LiDAR feature. Case (a) close to white indicates suppression, while case (b) close to black signifies activation.

Thus, we first get an attention map of non-empty LiDAR voxel $(L_l)_i$ as a query and pair $(V_l)_i$ as a key to calculate their activation relation.

$$attn((L_l)_i, (V_l)_i) = softmax((L_l)_i (V_l)_i^T) \tag{2}$$

where $(\cdot)_i$ refers to the i-th value corresponding to the LiDAR voxel feature and $softmax$ is a softmax function. As shown in Fig. 3, if there are many non-empty radar voxels in $(V_l)_i$, such as case (b), the attention value will be set to enhance $(L_l)_i$. Conversely, in the absence of neighbor radar voxels, such as case (a), the attention value will be set to suppress $(L_l)_i$.

The attention map is then multiplied to w_l^v, a value function used to extract value features of radar features. We use the attention mechanism to make the model further focus on the relevant radar information and obtain fusion features for a pair of $(L_l)_i$ and $(V_l)_i$, $(F_l)_i$. The equation can be written as Eq. (3):

$$(F_l)_i = attn((L_l)_i, (V_l)_i) w_l^v (V_l), \tag{3}$$

The final fusion feature F_l is obtained by aggregating $(F_l)_i$ using the indices of (L_l).

3.3. Weather-conditional Radar-flow Gating

The fusion features F_l obtained from the 3D-LRF module contain important information about how much to en-

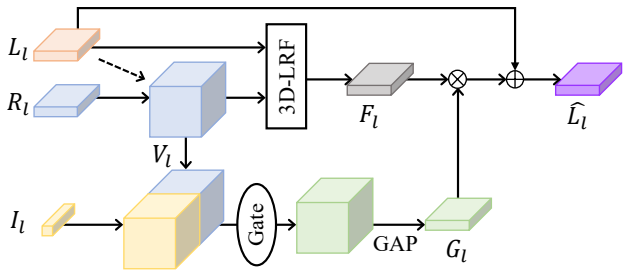

Figure 4. Illustration of proposed WRGNet. Weather-conditioned image feature and radar feature are gated to compute G_l, which modulates the flow to F_l and outputs enhanced LiDAR features $\hat{L}_l$.

hance or suppress LiDAR features. Therefore, fusing F_l to the original LiDAR features L_l can enhance the L_l, having the potential to improve the detection performance. The straightforward fusing strategy to obtain the enhanced LiDAR features $\hat{L}_l$ is as follows:

$$\hat{L}_l = L_l + F_l \tag{4}$$

Then, $\hat{L}_l$ is fed into the next convolution layer $l+1$ of the LiDAR stream. In the meantime, $\hat{L}_l$ and L_l fed into the separate BEV encoders to obtain the BEV features B_l and $\hat{B}_l$, respectively. These BEV features are then used in the detection head to make the predictions (Sec. 3.4).

While the straightforward fusing strategy can improve the performance in certain weather conditions, for example, the radar features in heavy snow conditions can enhance the LiDAR features, we observed that this fusing strategy causes a trade-off between performances in various weather conditions. The direct fusion of radar into LiDAR may degrade the performance in normal weather conditions due to the radar's limited precision in localization. Therefore, we propose to modulate the information flow of fusion features F_l depending on the weather conditions, *i.e.* gating F_l conditioned on the weather. To inject the weather conditions into the gating process, we choose to use image features I_l pre-trained by the weather classification task. The basic idea is that image is most affected by weather conditions, so the quality of data for 3D detection is low, but it is very advantageous for understanding weather conditions. The image feature is then used in the weather-conditional radar-flow gating network (WRGNet) to generate gating features G_l, modulating the information flow F_l into L_l.

WRGNet. Given the pre-trained image feature I_l, we first concatenate it with K_l nearest neighbor voxel features V_l by repeating I_l. Then, a single gate layer w_l^g is applied, followed by a global average pooling (GAP) layer to obtain the gating feature G_l as follows:

$$G_l = GAP(w_l^g([V_l, repeat(I_l)])) \tag{5}$$

where $[\cdot,\cdot]$ refers to a concatenation operation. The obtained G_l is used to gate the fusion feature F_l before fusing it into the LiDAR feature L_l as,

$$\hat{L}_l = L_l + G_l \otimes F_l, \tag{6}$$

where $\otimes$ denotes element-wise multiplication. The overall scheme of WRGNet is illustrated in Fig. 4. Through the proposed gated fusion strategy, we can control the information flow from the 3D-LRF module depending on the weather conditions, effectively alleviating the trade-off between performances in various weather conditions.

3.4. BEV Encoder and Detection Head

BEV encoder of each layer takes the LiDAR voxel feature L_l and enhanced LiDAR feature $\hat{L}_l$. In the BEV encoder, L_l and $\hat{L}_l$ are encoded separately with one sparse 3D convolution block, dense block, and transpose 2D convolution block to extract BEV features B_l and $\hat{B}_l$, respectively. The transpose 2D convolution block is designed to ensure that all layers' BEV features have the same dimension size $\mathbb{R}^{H_d \times W_d \times D}$, where H_d, W_d, D are the height, width, channel of BEV feature, respectively. After the network extracts BEV features from all layers, we concatenate all BEV features to make the final BEV feature $B \in \mathbb{R}^{H_d \times W_d \times 6D}$, and it is fed into the detection head.

The detection head is composed of the classification head and regression head, and they extract the classification and regression output of each grid, respectively, to estimate the center point, object size, and rotation [28, 34]. Each head consists of one convolution block. Focal loss [21] is adopted to train the classification and smooth L1 loss minimizes the regression error. The overall objective is the straightforward sum of the Focal loss and smooth L1 loss.

4. Experiments

4.1. Experimental Setup

Dataset and Metrics. The K-Radar dataset is a large-scale autonomous driving benchmark that contains 17,458 training and 17,536 testing scenes. It covers various time conditions (day, night), weather conditions (*e.g.* normal, rain, fog, snow, sleet), road structures (*e.g.* urban, highway, mountain) and sensor measurements (*e.g.* 4D radar, LiDAR, camera, GPS). The K-radar dataset is the only benchmark that provides 4D radar tensors in adverse weather conditions. Following the protocol of the original paper, we adopt two evaluation metrics for 3d object detection: AP_{3D} and AP_{BEV} of the class "Sedan" at IoU=0.3. For a more in-depth analysis, we additionally report the results at IoU=0.5.

Implementation Details. For the K-Radar dataset, we preprocess the 4D radar tensor by selecting only the top 10% of points with high power measurement as in [28]. We

Table 1. Quantitative results of LiDAR and 4D radar-based 3D object detection methods on K-Radar dataset [28]. We present the modality of each method (L: LiDAR, 4DR: 4D radar) and detailed performance for each weather condition. Best in **bold**, second in underline.

Methods	Modality	IoU	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
RTNH [28]	4DR	0.3	AP_{BEV}	41.1	41.0	44.6	45.4	32.9	50.6	81.5	56.3
			AP_{3D}	37.4	37.6	42.0	41.2	29.2	49.1	63.9	43.1
		0.5	AP_{BEV}	36.0	35.8	41.9	44.8	30.2	34.5	63.9	55.1
			AP_{3D}	14.1	19.7	20.5	15.9	13.0	13.5	21.0	6.36
RTNH*	L	0.3	AP_{BEV}	76.5	76.5	88.2	86.3	77.3	55.3	81.1	59.5
			AP_{3D}	72.7	73.1	76.5	84.8	64.5	**53.4**	80.3	52.9
		0.5	AP_{BEV}	66.3	65.4	87.4	83.8	73.7	**48.8**	78.5	48.1
			AP_{3D}	37.8	39.8	46.3	**59.8**	28.2	**31.4**	50.7	24.6
PointPillars [16]	L	0.3	AP_{BEV}	51.9	51.6	53.5	45.4	44.7	54.3	81.2	55.2
			AP_{3D}	47.3	46.7	51.9	44.8	42.4	45.5	59.2	55.2
		0.5	AP_{BEV}	49.1	48.2	53.0	45.4	44.2	45.9	74.5	53.8
			AP_{3D}	22.4	21.8	28.0	28.2	27.2	22.6	23.2	12.9
InterFusion [43]	4DR+L	0.3	AP_{BEV}	57.5	57.2	60.8	81.2	52.8	27.5	72.6	57.2
			AP_{3D}	53.0	51.1	58.1	80.9	40.4	23.0	71.0	55.2
		0.5	AP_{BEV}	52.9	50.0	59.0	80.3	50.0	22.7	72.2	53.3
			AP_{3D}	17.5	15.3	20.5	47.6	12.9	9.33	56.8	25.7
Ours	4DR+L	0.3	AP_{BEV}	**84.0**	**83.7**	**89.2**	**95.4**	**78.3**	**60.7**	**88.9**	**74.9**
			AP_{3D}	**74.8**	**81.2**	**87.2**	**86.1**	**73.8**	49.5	**87.9**	**67.2**
		0.5	AP_{BEV}	**73.6**	**72.3**	**88.4**	**86.6**	**76.6**	47.5	**79.6**	**64.1**
			AP_{3D}	**45.2**	**45.3**	**55.8**	51.8	**38.3**	23.4	**60.2**	**36.9**

set the point cloud range as [0m, 72m] for the X axis, [-6.4m, 6.4m] for the Y axis, and [-2m, 6m] for the Z axis setting the same environment with [28]. The voxel size is set to (0.4m, 0.4m, 0.4m). For the training strategy, we first pre-train the image-based weather classification network for 43 epochs with cross-entropy loss. The classification accuracy on seven weather conditions achieved 91%. Then, our framework loads the pre-trained weight for initialization and trains the entire network with the loss in Sec. 3.4 for 11 epochs. We use batch size 4 and Adam optimizer [15] with lr= 1e-3, β_1 = 0.9, β_2 = 0.999. More implementation details are in the supplementary material.

4.2. Main Results

We compare our method with LiDAR only, 4D radar only, and LiDAR-4D radar fusion-based 3D object detection methods: PointPillars [16], RTNH [28] and InterFusion [43]. The variant of RTNH*, which has same model structure with [28] and takes LiDAR as input, is additionally adopted for comparison.

Table 1 shows the quantitative comparison results. Our method outperforms single- and multi-modal 3D object detection models under all metrics. Specifically, ours surpasses the second best model, RTNH*, by around 19% increase in AP_{3D} with IoU threshold 0.5. This result demonstrates that the proposed 3D-LRF module effectively fuses LiDAR and 4D radar in the 3D domain. Moreover, our method shows favorable performance in all weather conditions, confirming the validity of the proposed WRGNet. Under heavy snow, our model demonstrates performance improvements ranging from 10% to 43% compared to the second-best scores across various metrics. Under sleet and fog, RTNH* shows slightly higher results for several metrics than ours. This might be because our model heavily relies on radar under sleet and fog (the inefficiency of LiDAR in sleet conditions is evident from its inability to capture even the ground surface as shown in Fig. 5), leading to a greater impact on false positives than incorporating less scene information using LiDAR. Our method shows the least degradation in performance when increasing the IoU threshold from 0.3 to 0.5. This indicates that our model excels at accurately placing 3D bounding boxes in locations with objects compared to other models.

The visual comparison is shown in Fig. 5. Our method predicts the most accurate 3D bounding boxes under various weather conditions, while LiDAR-based methods encounter true negatives influenced by adverse weather and 4D radar-based methods suffer from false positives (especially under sleet) due to the characteristics of the sensor. Interfusion [43] shows better results than single modality-based models in normal scenes. However, in adverse conditions, it fails to outperform them as it struggles to effectively fuse features from both modalities.

4.3. Model Analysis

We analyze our model to validate the effects of each component and visualize the objectness score map calculated from the classification head for in-depth analysis. We also discuss the effect of using LiDAR/4D radar modality and compare our method with various fusion methods.

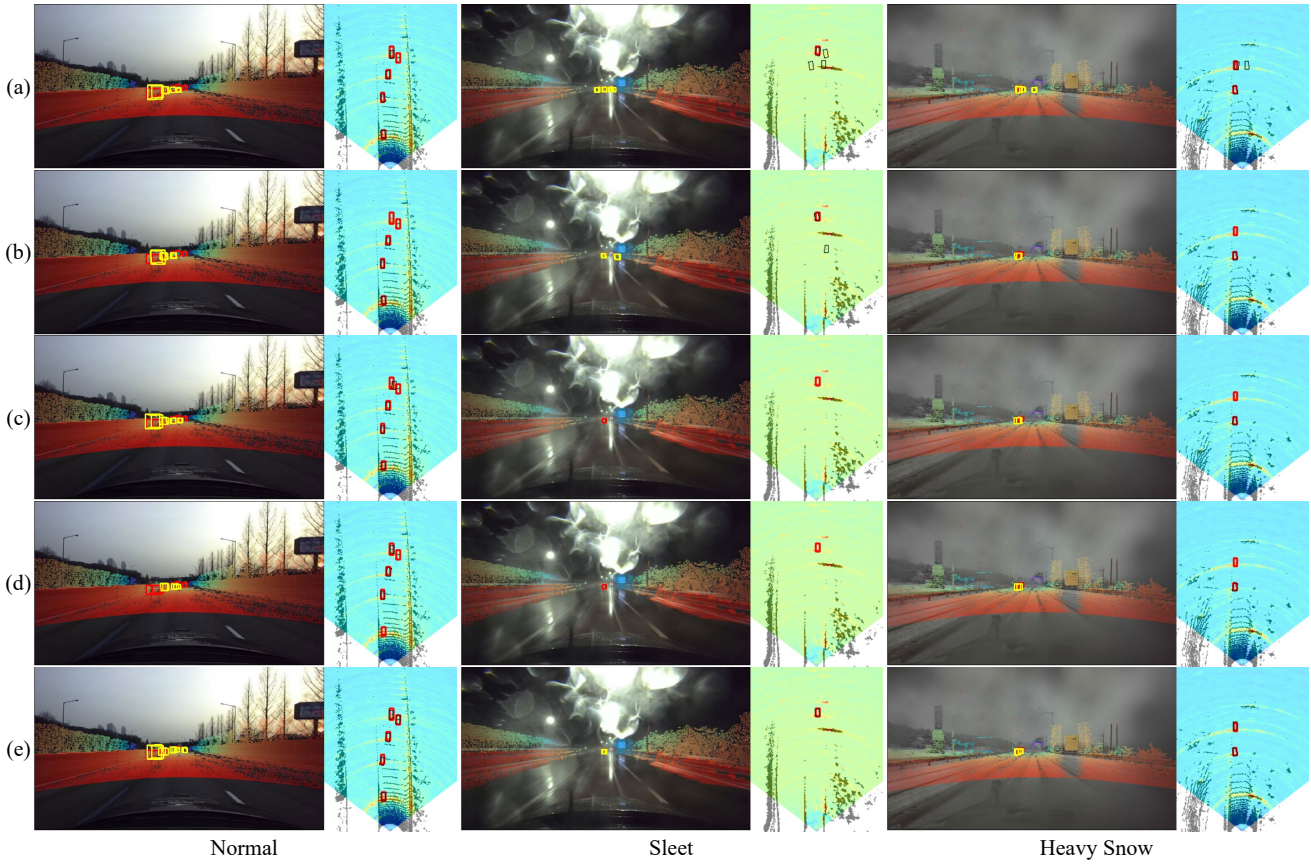

Figure 5. Visual results of 3D object detection in range view and bird-eye-view. The results in the range view show the image and projected LiDAR with red GT boxes and yellow predicted boxes. The results in bird-eye view show top-view LiDAR and 4D radar heatmap with red GT boxes and black predicted boxes. Each row means weather conditions: normal, sleet, and heavy snow (from left to right). Each column means the 3D object detection model: (a) RTNH [28], (b) RTNH*, (c) PointPillar [16], (d) InterFusion [43], (e) ours. More results with all weather conditions are in supplementary material. Best viewed when zoomed in with colors.

Effect of Each Modality. Table 2 shows the effect of using LiDAR and 4D radar domains. As the baseline of our model is RTNH [28], the performance of RTNH is reported in (a), RTNH* in (b), and ours in (e). While radar measurements exhibit robustness in adverse weather, the ability to precisely determine the exact location of objects is superior with LiDAR. Hence, currently, utilizing LiDAR for 3D object detection yields better performance. In cases like heavy snow, radar leverages its strengths and achieves performance comparable to LiDAR (see Table 1). Our model effectively takes advantage of both 4D radar and LiDAR in the 3D domain based on weather information, leading to optimal performance.

Component Analysis. The rows in (c), (d), and (e) of Table 2 illustrate the impact of each component of our model. When WRGNet is added, AP_{BEV} significantly improves across all IoU thresholds compared to the results using only single modalities. This demonstrates that WRGNet effectively regulates the radar flow and fuses information from both modalities by incorporating weather information. On the other hand, since (c) concatenates the BEV features of each modality without effective fusion in the 3D domain, AP_{3D} shows a slight increase or decrease compared to (b). In (d), we observe an overall enhancement across all metrics. This demonstrates the effective fusion of both domains in 3D, leveraging the characteristics of LiDAR with precise information and radar sensors capable of detecting approximate object positions in adverse weather conditions. Notably, it achieved a substantial performance boost even without utilizing weather information and outperformed other fusion methods, including concatenation and attention-based results in Table 3. (e) is our final model with two previously validated components, and shows the optimal performance resulting from the effective integration of the two proposed ideas.

Fusion Comparison. We validate that our 3D-LRF module is not a mere outcome of combining two domains. Two fusion strategies, concatenation and cross-attention, are adopted for the comparison and their results are shown in Table 3. The results with cross-attention show better per-

Table 2. Effect of each modality and component of our model. Best in **bold**, second in underline.

Methods	Modality		Components		IoU=0.3		IoU=0.5	
	R	L	WRG	3D-LRF	AP_{3D} (↑)	AP_{BEV} (↑)	AP_{3D} (↑)	AP_{BEV} (↑)
(a) Radar only	✓				37.4	41.1	14.1	36.0
(b) LiDAR only		✓			72.7	76.5	37.8	66.3
(c) Ours without 3D-LRF	✓	✓	✓		73.9 (+0.8)	82.8 (+6.3)	35.9 (-1.9)	70.9 (+4.6)
(d) Ours without WRG	✓	✓		✓	74.7 (+2.0)	**84.6** (+8.1)	38.6 (+0.8)	72.7 (+6.4)
(e) Ours	✓	✓	✓	✓	**74.8** (+2.1)	84.0 (+7.5)	**45.2** (+7.4)	**73.6** (+7.3)

Table 3. Comparison with various multi-modal feature fusion methods. Best in **bold**, second in underline.

Fusion	IoU=0.3		IoU=0.5	
	AP_{3D} (↑)	AP_{BEV} (↑)	AP_{3D} (↑)	AP_{BEV} (↑)
Concat.	73.5	82.0	34.9	69.9
Attn.	74.3 (+0.8)	83.1 (+1.1)	35.1 (+0.2)	70.3 (+0.4)
Ours	**74.8** (+1.3)	**84.0** (+2.0)	**45.2** (+10.3)	**73.6** (+3.7)

formance than those with concatenation. However, despite utilizing both domains, it can be observed that the results at AP_{3D} with IoU threshold 0.5 are not superior to using LiDAR alone. In contrast, with the help of 3D-LRF module, the proposed method effectively fuses LiDAR and 4D radar features, showing a substantial performance improvement compared to other fusion methods.

Objectness Score Map Visualization. We visualize the objectness score from the classification head to investigate which part of the input the model primarily focuses on. Fig. 6 introduces two scenes, one categorized as day, heavy snow (I) and the other as night, normal (II). In the case of (I), (b) with only LiDAR fails to recognize the vehicles behind, while (c) with only radar successfully focuses on both vehicles under adverse weather conditions. Concatenation (d) or attention (e) still activates non-vehicle areas. In (f), due to adverse weather conditions, it is evident that radar information is more prominently incorporated. In (g), the model enriches the locations that LiDAR and radar both focus on and suppresses the locations where radar does not focus. In (h), only locations with sedans are distinctly activated. In the case of (II), in (f), a substantial amount of LiDAR information is incorporated as the scene is under normal conditions. In (g), despite the distinct areas of activation in each modality, it is evident that the features have been enriched to activate only in regions where vehicles are present. In (h), clear activation is observed where objects are present. These two examples validate that the proposed method aligns with its intended design and demonstrates its effectiveness compared to other methods.

5. Conclusion

We introduce an accurate and robust LiDAR and 4D radar-based 3D object detection framework under various weather conditions. Our method effectively fuses LiDAR and 4D radar with the 3D-LRF module, considering the advantages of each sensor in 3D domain. Moreover, WRGNet is proposed to control the flow from 4D radar to LiDAR according to the weather information. Extensive experiments demonstrate the validity of two novel ideas, and our model achieves remarkable improvement over the state-of-the-art LiDAR and 4D radar-based models in K-Radar dataset.

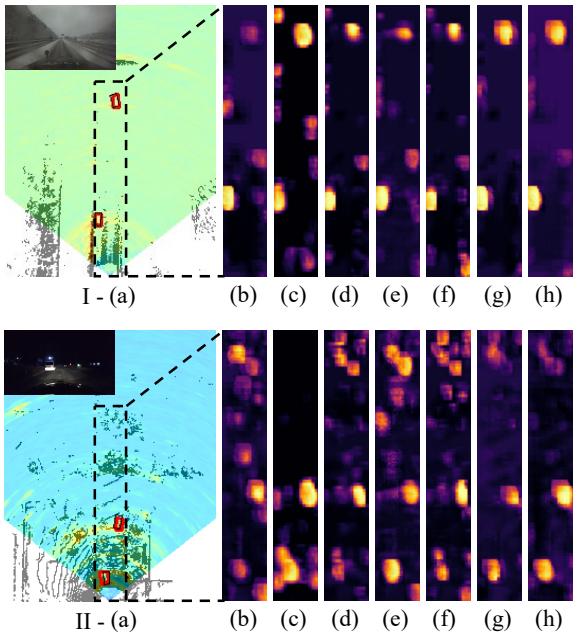

Figure 6. Visualized objectness score map of (b) RTNH*, (c) RTNH [28], (d) concatenation-based variant, (e) attention-based variant, (f) ours without 3D-LRF, (g) ours without WRGNet and (h) ours. (I: day scene under heavy snow, II: night normal scene)

Acknowledgement. This work was supported by the Technology Innovation Program (1415187329,20024355, Development of autonomous driving connectivity technology based on sensor-infrastructure cooperation) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea), the Challengeable Future Defense Technology Research and Development Program through the Agency For Defense Development(ADD) funded by the Defense Acquisition Program Administration(DAPA) in 2024(No.912768601) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF2022R1A2B5B03002636).

References

[1] Eduardo Arnold, Omar Y. Al-Jarrah, Mehrdad Dianati, Saber Fallah, David Oxtoby, and Alex Mouzakitis. A survey on 3d object detection methods for autonomous driving applications. *IEEE Transactions on Intelligent Transportation Systems*, 20(10):3782–3795, 2019. 1

[2] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1080–1089, 2022. 3

[3] Kshitiz Bansal, Keshav Rungta, Siyuan Zhu, and Dinesh Bharadia. Pointillism: accurate 3d bounding box estimation with multi-radars. *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020. 2

[4] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yuxin Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11618–11628, 2019. 1

[5] Colin Decourt, Rufin van Rullen, Didier Salle, and Thomas Oberlin. A recurrent cnn for online object detection on raw radar frames. *ArXiv*, abs/2212.11172, 2022. 2

[6] Florian Drews, Di Feng, Florian Faion, Lars Rosenbaum, Michael Ulrich, and Claudius Gläser. Deepfusion: A robust and modular 3d object detector for lidars, cameras and radars. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 560–567, 2022. 2, 3

[7] Deepti Hegde, Velat Kilic, Vishwanath Sindagi, A Brinton Cooper, Mark Foster, and Vishal M Patel. Source-free unsupervised domain adaptation for 3d object detection in adverse weather. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6973–6980, 2023. 3

[8] Keli Huang, Botian Shi, Xiang Li, Xin Li, Siyuan Huang, and Yikang Li. Multi-modal sensor fusion for auto driving perception: A survey. *ArXiv*, abs/2202.02703, 2022. 1

[9] Kuan-Chih Huang, Tsung-Han Wu, Hung-Ting Su, and Winston H. Hsu. Monodtr: Monocular 3d object detection with depth-aware transformer. In *CVPR*, 2022. 1

[10] Jyh-Jing Hwang, Henrik Kretzschmar, Joshua Manela, and Sean Rafferty. Cramnet: Camera-radar fusion with ray-constrained cross-attention for robust 3d object detection. *2022 European Conference on Computer Vision (ECCV)*, 2022. 3

[11] Zhang Jinqing, Zhang Yanan, Liu Qingjie, and Wang Yunhong. Sa-bev: Generating semantic-aware bird's-eye-view feature for multi-view 3d object detection. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 1

[12] Youngseok Kim, Jun Won Choi, and Dongsuk Kum. Grif net: Gated region of interest fusion network for robust 3d object detection from radar point cloud and monocular image. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10857–10864, 2020. 1, 2, 3

[13] Youngseok Kim, Sanmin Kim, Jun Won Choi, and Dongsuk Kum. Craft: Camera-radar 3d object detection with spatio-contextual fusion transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[14] Youngseok Kim, Juyeb Shin, Sanmin Kim, In-Jae Lee, Jun Won Choi, and Dongsuk Kum. Crn: Camera radar net for accurate, robust, efficient 3d perception. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2, 3

[15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 6

[16] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12689–12697, 2018. 1, 2, 6, 7

[17] Hao Li, Zehan Zhang, Xian Zhao, Yulong Wang, Yuxi Shen, Shiliang Pu, and Hui Mao. Enhancing multi-modal features using local self-attention for 3d object detection. *2022 European Conference on Computer Vision (ECCV)*, 2022. 3

[18] Xin Li, Tao Ma, Yuenan Hou, Botian Shi, Yuchen Yang, Youquan Liu, Xingjiao Wu, Qin Chen, Yikang Li, Yu Qiao, and Liang He. Logonet: Towards accurate 3d object detection with local-to-global cross-modal fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3

[19] Yu-Jhe Li, Jinhyung Park, Matthew O'Toole, and Kris Kitani. Modality-agnostic learning for radar-lidar fusion in vehicle detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 908–917, 2022. 1, 2, 3

[20] Jia Lin, Huilin Yin, Jun Yan, Wancheng Ge, Hao Zhang, and Gerhard Rigoll. Improved 3d object detector under snowfall weather condition based on lidar point cloud. *IEEE Sensors Journal*, 22(16):16276–16292, 2022. 3

[21] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017. 5

[22] Jianan Liu, Qiuchi Zhao, Weiyi Xiong, Tao Huang, Qing-Long Han, and Bing Zhu. Smurf: Spatial multi-representation fusion for 3d object detection with 4d imaging radar. *ArXiv*, abs/2307.10784, 2023. 1, 2

[23] Zhijian Liu, Haotian Tang, Alexander Amini, Xingyu Yang, Huizi Mao, Daniela Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2023. 3

[24] Michael Meyer, Georg Kuschk, and Sven Tomforde. Graph convolutional networks for 3d object detection on radar data. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3053–3062, 2021. 1, 2

[25] Nguyen Anh Minh Mai, Pierre Duthon, Pascal Housam Salmane, Louahdi Khoudour, Alain Crouzil, and Sergio A. Velastin. Camera and lidar analysis for 3d object detection in

foggy weather conditions. In *2022 12th International Conference on Pattern Recognition Systems (ICPRS)*, pages 1–7, 2022. 2, 3

[26] Ramin Nabati and Hairong Qi. Centerfusion: Center-based radar and camera fusion for 3d object detection. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1526–1535, 2020. 3

[27] Felix Nobis, Ehsan Shafiei, Phillip Karle, Johannes Betz, and Markus Lienkamp. Radar voxel fusion for 3d object detection. *Applied Sciences*, 11(12), 2021. 3

[28] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 1, 2, 5, 6, 7, 8

[29] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. Enhanced k-radar: Optimal density reduction to improve detection performance and accessibility of 4d radar tensor-based object detection. *2023 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–6, 2023. 2

[30] Anshul Paigwar, David Sierra-Gonzalez, Özgür Erkent, and Christian Laugier. Frustum-pointpillars: A multi-stage approach for 3d object detection using rgb camera and lidar. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2926–2933, 2021. 3

[31] Andras Palffy, Jiaao Dong, Julian F. P. Kooij, and Dariu M. Gavrila. Cnn based road user detection using the 3d radar cube. *IEEE Robotics and Automation Letters*, 5:1263–1270, 2020. 1, 2

[32] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, Julian Kooij, and Dariu Gavrila. Multi-class road user detection with 3+1d radar in the view-of-delft dataset. *IEEE Robotics and Automation Letters*, 7:4961–4968, 2022. 2

[33] Kun Qian, Shilin Zhu, Xinyu Zhang, and Li Erran Li. Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 444–453, 2021. 2, 3

[34] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1137–1149, 2015. 5

[35] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew M. Wallace. Radiate: A radar dataset for automotive perception in bad weather. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7, 2020. 1

[36] Vishwanath A. Sindagi, Yin Zhou, and Oncel Tuzel. Mvx-net: Multimodal voxelnet for 3d object detection. *2019 International Conference on Robotics and Automation (ICRA)*, pages 7276–7282, 2019. 3

[37] Jiying Song, Haiyue Wei, Lin Bai, Lei Yang, and Caiyan Jia. Graphalign: Enhancing accurate feature alignment by graph matching for multi-modal 3d object detection. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 3

[38] Bin Tan, Zhixiong Ma, Xichan Zhu, Sen Li, Lianqing Zheng, Sihan Chen, Libo Huang, and Jie Bai. 3-d object detection for multiframe 4-d automotive millimeter-wave radar point cloud. *IEEE Sensors Journal*, 23:11125–11138, 2023. 1, 2

[39] Anh The Do and Myungsik Yoo. Lossdistillnet: 3d object detection in point cloud under harsh weather conditions. *IEEE Access*, 10:84882–84893, 2022. 3

[40] Kaicheng Yu Zhongyu Xia Zhiwei Lin Yongtao Wang Tao Tang Bing Wang Tingting Liang, Hongwei Xie and Zhi Tang. BEVFusion: A Simple and Robust LiDAR-Camera Fusion Framework. In *Neural Information Processing Systems (NeurIPS)*, 2022. 3

[41] Chunwei Wang, Chao Ma, Ming Zhu, and Xiaokang Yang. Pointaugmenting: Cross-modal augmentation for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11794–11803, 2021. 3

[42] Leichen Wang, Tianbai Chen, Carsten Anklam, and Bastian Goldluecke. High dimensional frustum pointnet for 3d object detection from camera, lidar, and radar. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1621–1628, 2020. 3

[43] Li Wang, Xinyu Zhang, Baowei Xv, Jinzhao Zhang, Rong Fu, Xiaoyu Wang, Lei Zhu, Haibing Ren, Pingping Lu, Jun Li, and Huaping Liu. Interfusion: Interaction-based 4d radar and lidar fusion for 3d object detection. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12247–12253, 2022. 1, 2, 3, 6, 7

[44] Li Wang, Xinyu Zhang, Ziying Song, Jiangfeng Bi, Guoxin Zhang, Haiyue Wei, Liyao Tang, Lei Yang, Jun Li, Caiyan Jia, and Lijun Zhao. Multi-modal 3d object detection in autonomous driving: A survey and taxonomy. *IEEE Transactions on Intelligent Vehicles*, 8(7):3781–3798, 2023. 1

[45] Yan Wang, Junbo Yin, Wei Li, Pascal Frossard, Ruigang Yang, and Jianbing Shen. Ssda3d: Semi-supervised domain adaptation for 3d object detection from point cloud. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. 3

[46] Hai Wu, Chenglu Wen, Shaoshuai Shi, and Cheng Wang. Virtual sparse convolution for multimodal 3d object detection. In *CVPR*, 2023. 3

[47] Liang Xie, Chao Xiang, Zhengxu Yu, Guodong Xu, Zheng Yang, Deng Cai, and Xiaofei He. Pi-rcnn: An efficient multi-sensor 3d object detector with point-based attentive cont-conv fusion module. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12460–12467, 2020. 3

[48] Xinli Xu, Shaocong Dong, Tingfa Xu, Lihe Ding, Jie Wang, Peng Jiang, Liqiang Song, and Jianan Li. Fusionrcnn: Lidar-camera fusion for two-stage 3d object detection. *Remote Sensing*, 15:1839, 2023. 3

[49] Junjie Yan, Yingfei Liu, Jianjian Sun, Fan Jia, Shuailin Li, Tiancai Wang, and Xiangyu Zhang. Cross modal transformer via coordinates encoding for 3d object dectection. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 3

[50] Bin Yang, Runsheng Guo, Mingfeng Liang, Sergio Casas, and Raquel Urtasun. Radarnet: Exploiting radar for robust perception of dynamic objects. In *European Conference on Computer Vision*, 2020. 3

[51] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. *CVPR*, 2021. 1

[52] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Multi-modal virtual point 3d detection. *NeurIPS*, 2021. 3

[53] Zixiang Zhou, Xiangchen Zhao, Yu Wang, Panqu Wang, and Hassan Foroosh. Centerformer: Center-based transformer for 3d object detection. In *ECCV*, 2022. 1