Self-supervised adaptive weighting for cooperative perception in V2V communications

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1109/TIV.2023.3345035

Publisher: IEEE

www.reading.ac.uk/centaur

Self-supervised Adaptive Weighting for Cooperative Perception in V2V Communications

Chenguang Liu, Jianjun Chen, Yunfei Chen, *Senior Member, IEEE*,
Ryan Payton, Michael Riley, and Shuang-Hua Yang, *Senior Member, IEEE*

Abstract—Perception of the driving environment is critical for collision avoidance and route planning to ensure driving safety. Cooperative perception has been widely studied as an effective approach to addressing the shortcomings of single-vehicle perception. However, the practical limitations of vehicle-to-vehicle (V2V) communications have not been adequately investigated. In particular, current cooperative fusion models rely on supervised models and do not address dynamic performance degradation caused by arbitrary channel impairments. In this paper, a self-supervised adaptive weighting model is proposed for intermediate fusion to mitigate the adverse effects of channel distortion. The performance of cooperative perception is investigated in different system settings. Rician fading and imperfect channel state information (CSI) are also considered. Numerical results demonstrate that the proposed adaptive weighting algorithm significantly outperforms the benchmarks without weighting. Visualization examples validate that the proposed weighting algorithm can flexibly adapt to various channel conditions. Moreover, the adaptive weighting algorithm demonstrates good generalization to untrained channels and test datasets from different domains.

Index Terms—Adaptive weighting, cooperative perception, self-supervised learning, V2V communications.

I. INTRODUCTION

Cooperative perception enabled by vehicular communications facilitates the exchange of complementary perceptual information among multiple connected and autonomous vehicles (CAVs), leading to a more comprehensive and integrated perception for autonomous driving [1]. By aggregating the information from multiple viewpoints, it provides an effective framework to alleviate the limitations of single-vehicle

Chenguang Liu is with the School of Engineering, University of Warwick, Coventry, UK, CV4 7AL. e-mail: chenguang.liu@warwick.ac.uk

Jianjun Chen is with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. e-mail: jianjun.chen@student.uts.edu.au

Yunfei Chen is with the Department of Engineering, University of Durham, Durham, UK, DH1 3LE. e-mail: yunfei.chen@durham.ac.uk

Ryan Payton and Michael Riley are with Oracle for Research. email: ryan.payton@oracle.com; michael.riley@oracle.com

Shuang-Hua Yang is with Shenzhen Key laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, Shenzhen, China, and also with Department of Computer Science, University of Reading, UK. e-mail: shuang-hua.yang@reading.ac.uk

perception, to detect occluded or distantly located objects. Such collaboration schemes highly rely on the vehicular communications system to transmit the shared information from CAVs to the ego vehicle. In practice, the reliability of vehicle-to-vehicle (V2V) communications is constrained by limited bandwidth and channel distortion. Therefore, addressing the practical limitations imposed by the V2V communications system on collaborative perception is vital, which has garnered considerable interest from researchers.

Based on the previous works in 3D detection backbones [2]–[4], related works have been widely conducted to develop efficient collaborative fusion schemes, including sharing raw point clouds (*i.e.* early fusion) [5], intermediate features (*i.e.* intermediate fusion) [6]–[14] and detection results (*i.e.* late fusion) [15]. Most of these works focus on improving perception accuracy and reducing bandwidth utilization, assuming ideal exchange of information. To consider more realistic V2V communication, some studies examined the effects of communications delay [12], pose error [7], and lossy communications [13]. However, they have still assumed perfect communications channel without any channel impairments, except [16], which considered Rician fading and free-space path loss by incorporating the learning-based communication channel into cooperative perception. Moreover, the work in [16] demonstrated that the intermediate features are more robust to channel impairments than raw point clouds and detection results. Also, the intermediate fusion can be further improved in the presence of channel impairments.

In realistic V2V communications, the information shared among vehicles could potentially enhance the detection performance, but if it suffers from severe channel distortion, it may instead manifest as interference that undermines detection accuracy. Especially for intermediate features, it is more challenging to identify the corrupted features than the raw point clouds and detection outputs since latent features do not have a physical detection range. Therefore, it is vital to prevent the severely distorted information from taking part in the fusion or collaboration.

On the other hand, the channel distortion in realistic V2V communication may exhibit arbitrary and dynamic characteristics, leading to signal strength fluctuation [17]. This non-stationarity of the V2V channel is mainly caused by the high velocity of the transmitter, receiver and moving scatterers in the V2V environment [18]. This necessitates the cooperative system to adapt to various levels of channel impairment. In this regard, relying solely on the supervised training with the well-labeled dataset is inefficient when dealing with vari-

ous communication environments, since labeling the received intermediate features is difficult due to their latent representation. Consequently, the cost of training and labeling would exponentially increase when scaling to a more complex communication environment or different levels of channel impairments. Supervised models might also encounter difficulties in generalizing effectively to unseen data that significantly differs from the training samples due to random distortion. Therefore, it is necessary to design a specialized loss function and training scheme to address these challenges.

Although the previous works have explored various fusion schemes for cooperative perception, several challenges remain. For example, although different fusion schemes for cooperative perception with V2V communications were studied in [16], the significant performance degradation for intermediate fusion due to channel distortion has yet to be addressed. Additionally, the works mentioned above mainly use supervised learning with well-labeled datasets, but this could be impractical for dynamic and random channel distortions. Motivated by these, this work will focus on mitigating the adverse effects of channel distortion on the intermediate fusion in cooperative perception. The main contributions are summarized below:

1) To address the severe performance degradation of intermediate fusion in [16], a self-supervised adaptive weighting model is proposed for the cooperative perception with intermediate fusion. A convolutional neural network (CNN)-based structure ending with Softmax is used for adaptive weighting. Self-supervised training with contrasting information is used to train this model without manual annotation. Different data domains, detection backbones, noise levels, and path loss factors are examined to validate its generalization and adaptive ability.

2) Unlike previous works that neglect realistic channel models, we evaluate the proposed adaptive weighting considering the WINNER II channel using orthogonal frequency-division multiplexing (OFDM). Different fusion methods and transmitted information are evaluated for different SNRs and path loss factors. A real-world dataset for cooperative perception is also considered.

3) Numerical results show that the proposed adaptive weighting algorithm performs better than the benchmarks without weighting. It is also validated that the proposed adaptive weighting can scale to PointPillars with different fusion schemes and real-world datasets. Visualization examples further validate that the proposed weighting algorithm can mitigate the adverse effects under severe channel conditions and simultaneously enhance the received shared features when the channel condition improves.

The rest of the paper is organized as follows. Section II will briefly review the related works. In Section III, we will introduce the system model of the cooperative perception system and describe the backbone algorithm. Section IV will present the adaptive weighting algorithm and the self-supervised optimization. Simulation results will be demonstrated in Section V. Finally, Section VI will conclude the work.

II. RELATED WORK

1) Single-vehicle perception: Due to the advancement of sensor technologies and learning-based algorithms, the performance of 3D object detection for autonomous driving has been continuously improved using 3D scanners, such as light detection and ranging (LiDAR). Several approaches to processing the LiDAR point clouds have been proposed in [2]–[4], [19]–[21]. The work in [2] proposed a novel method to convert point clouds into 3D voxels and extract essential features. This detection algorithm adopts an end-to-end learning structure, enabling integration with a learning-based communications system for the end-to-end global optimization. To reduce the computational complexity, SECOND was proposed in [4], where sparsely embedded convolutional layers were applied to 3D voxel features. On the other hand, the work in [3] proposed to convert the voxels along the z-axis into pillars to avoid the computation of 3D voxels. Meanwhile, the imbalance issue caused by anchors was addressed in [19]. Furthermore, a two-step detection framework was proposed in [20], where the rough estimation of proposals was generated first and refined in the second stage. In [21], the authors proposed to jointly leverage voxel-based and point-based detection methods.

2) Cooperative fusion: Using the previous works on LiDAR-based 3D object detection backbones, efficient cooperative fusion schemes can be adopted for multiple CAVs. In [5], early fusion was first proposed to aggregate the raw point clouds shared among multiple CAVs, while the late fusion scheme was proposed in [15] to use independently detected results for late fusion.

To avoid the excessive bandwidth usage in early fusion, F-Cooper was proposed in [6] to use the intermediate features in cooperative perception. Furthermore, V2VNet in [7] used a graph neural network for the intermediate feature fusion considering the V2V communication with imperfect localization and time delay. To further reduce the bandwidth requirement, Who2com in [22] proposed a three-stage handshake communication mechanism to select collaborators based on their matching scores. It was optimized by end-to-end supervision using the ground truth of the target task without annotations for the best agents to communicate with. Based on this handshake communication framework, When2com was proposed in [8] to construct communication groups and learn when to communicate by using an asymmetric attention mechanism, while Where2comm [9] proposed a spatial-confidence-aware communication scheme to only transmit the crucial area of the perceptual information. In [10], a collaboration weighting graph was proposed for multi-vehicle perception, which requires fewer communication rounds than When2com and V2VNet. Furthermore, in [11], the attention mechanism was used to aggregate the intermediate features from CAVs, while the work in [12] proposed a fusion framework based on vision Transformer to leverage the features from vehicles and infrastructures. Then, a hybrid object detection and tracking frame was proposed in [23], where historical tracking information was leveraged to enhance the inference for object detection with a spatial-temporal deep neural network. In [24], a data acquisition and analysis framework was proposed to collect

and process the advanced sensor data from multiple CAVs for vehicle trajectory extraction, reconstruction, and evaluation, in which the Kalman filter and the Chi-square test were used to reduce the noise and outlier in the trajectories. In [14], a spatial-wise adaptive fusion was proposed for intermediate features. In [13], attention-based modules were adopted to enhance the interaction between the ego vehicle and other CAVs, and a repair network was used to mitigate the adverse effects of lossy communications. Most of these works focus on improving perception accuracy and reducing bandwidth utilization, assuming ideal communications. Very few have considered more realistic V2V communications, such as time delay [12], pose error [7], [22], lossy communication [13] and Rician fading with channel impairments [16]. Among the existing works for cooperative fusion, [10], [14], and [22] are more relevant to our work, where different weighting algorithms were investigated for fusion. However, these works are trained with well-annotated datasets for detection and neglect the realistic channel impairments.
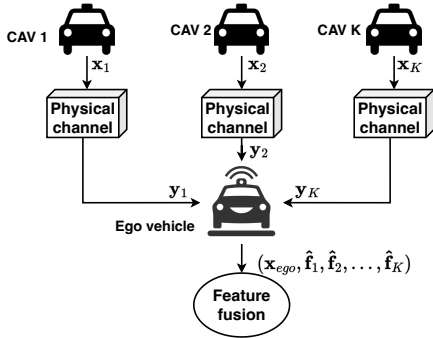
III. SYSTEM MODEL

Fig. 1: Cooperative perception via V2V communication.

A. V2V communications channel model

Consider a cooperative perception system where one ego vehicle is collaborating with K CAVs. The shared information is transmitted from the CAVs to the ego vehicle, which is demonstrated in Fig. 1.

1) Rician fading: Consider line-of-sight (LOS) signal propagation with path loss, the received signal at the ego vehicle from the k-th CAV is

$$\mathbf{y}_k = \sqrt{\frac{p_0}{d_k^n}} h_k \mathbf{x}_k + \mathbf{w}_k, \tag{1}$$

where $\sqrt{\frac{p_0}{d_k^n}}$ is the path loss with parameter p_0 determined by antennas and channel characteristics, distance d_k between transmitter and receiver, and path loss factor n, h_k denotes the Rician fading channel following $\mathcal{CN}(\mu, \sigma_h^2)$, $\mathbf{y}_k \in \mathbb{C}^{L \times 1}$ denotes the complex-valued received signals from the k-th CAV, $\mathbf{x}_k \in \mathbb{C}^{L \times 1}$ is the transmitted signal, and $\mathbf{w}_k$ denotes the additive white Gaussian noise following $\mathcal{CN}(0, \sigma^2)$.

2) Multi-path: In realistic communication, the communication channel may also suffer from multi-path fading. In addition to the communication model in (1), a multi-path channel model using OFDM is also considered. The received symbols at i-th sub-carrier from the k-th CAV are

$$\boldsymbol{Y}_k[i] = \boldsymbol{H}_k[i]\boldsymbol{X}_k[i] + \boldsymbol{W}_k[i], \tag{2}$$

where $\boldsymbol{H}_k[i]$ denotes the channel frequency response, $\boldsymbol{W}_k[i]$ denotes the additive white Gaussian noise, and $\boldsymbol{X}_k[i]$ denotes the transmitted symbol.

To cooperatively leverage the shared information from multiple CAVs, denote the aggregation of the shared information at the ego vehicle as

$$\mathbf{f}_{agg} = \mathcal{F}_{fusion}(\mathbf{f}_{ego}, \hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, ..., \hat{\mathbf{f}}_K), \tag{3}$$

where $\mathcal{F}_{fusion}(\cdot)$ denotes the fusion algorithm to combine the shared information received from CAVs, $\mathbf{f}_{ego}$ denotes the information sensed at the ego car itself, and $\hat{\mathbf{f}}_k$ denotes the shared information of the k-th CAV, reshaped from the recovered received signals $\mathbf{y}_k$ and $\boldsymbol{Y}_k$.

The sensed information from CAVs could provide informative viewpoints to improve the perception precision and range. However, when there are severe channel impairments, the recovered information $\hat{\mathbf{f}}_k$ suffers from distortion, thus compromising the aggregated feature $\mathbf{f}_{agg}$. In order to explore the effects of channel impairments, we consider the Rician fading channel with free-space path loss as in (1) and the WINNER II channel model [25] with multi-path fading as in (2) to simulate the practical channel. Imperfect CSI with a Gaussian disturbance and pilot-based least-square channel estimation are adopted, accounting for channel estimation error. A zero-forcing detector is used to recover the signal $\hat{\mathbf{f}}_k$ from the received $\mathbf{y}_k$ or $\boldsymbol{Y}_k$.

B. Cooperative perception backbone

Fig. 2 demonstrates the system model for cooperative perception with learning-based communications, the same as that in [16]. SECOND [4] is adopted as the backbone algorithm for 3D detection. Based on the end-to-end learning structure of VoxelNet [2], SECOND incorporates sparsely embedded convolutional layers to improve the computational efficiency. This enables the integration with learning-based communications system for the end-to-end global optimization. SECOND has three main components voxel feature extractor (VFE), sparse convolution middle layer (SpConv), and region proposal network (RPN). Firstly, iterative conversion of raw point clouds into voxel representations is conducted by assigning the points to their corresponding voxels. Then, a VFE layer is employed to extract point-wise features from voxels. Subsequently, 3D sparse convolution is conducted for these point-wise features by only applying convolution to those non-zero elements of the sparse embeddings. This selective processing significantly reduces computational costs. The RPN adopts a residual structure, which involves downsampling and upsampling operations. To achieve downsampling, multiple layers of 2D convolution are applied to the features, along with batch normalization and rectified linear unit (ReLU) activation.
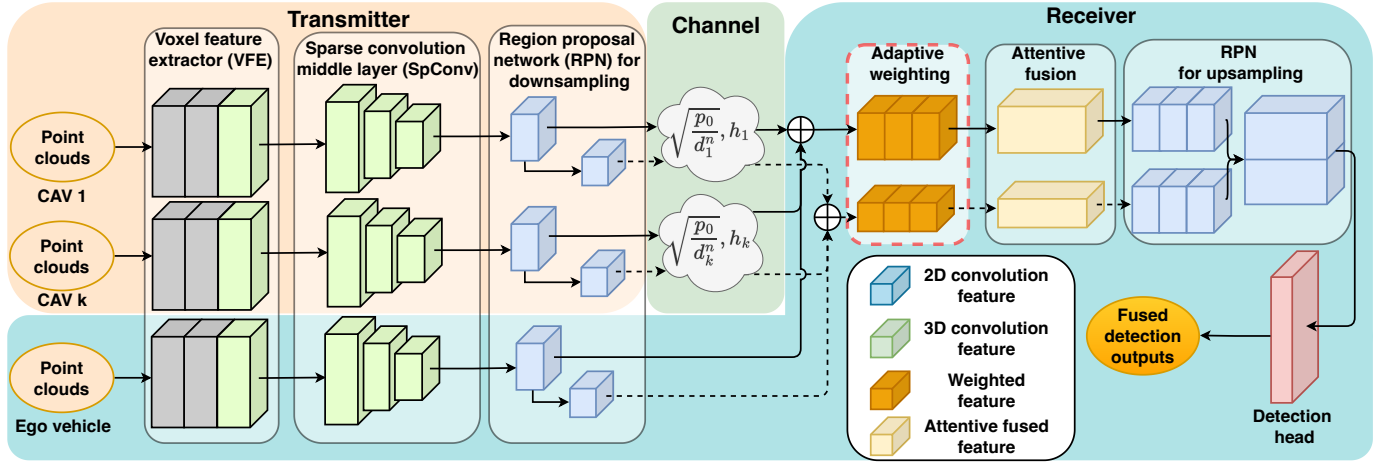
Fig. 2: The system model of cooperative perception with learning-based communications system.

On the other hand, upsampling is achieved by performing a 2D deconvolution on each level of downsampled features. Then, these deconvoluted features are concatenated to reconstruct the convolution feature, which serves as the input to the detection head. Finally, a single-shot detector (SSD) is used to output the classification results of objects and the regression results of their box localization.

To enable collaboration, multiple CAVs are connected to the ego vehicle. In this scenario, attentive fusion [11] is employed as an intermediate fusion method to aggregate the received downsampled features. Attentive fusion process the features of each CAV via attention-based neural network.

C. Supervised end-to-end training

Supervised end-to-end training is adopted for cooperative perception with V2V communications, as the end-to-end trainable structure of SECOND allows any part of the information to be transmitted through the learning-based communications system. In this case, the learning-based communication and the detection algorithm can be trained together to consider signal distortion in communication using the loss function in [3], [4]. Define the boxes and anchors for object detection as

$$(x, y, z, w, l, h, \theta), \tag{4}$$

where x, y and z denote the coordinates of the box center, w, l and h denote the width, length and height of the box, θ denotes the rotation angle around z-axis. The localization residuals for elements in (4) can be expressed by

$$\Delta_x = \frac{x_{gt} - \hat{x}}{\hat{d}}, \Delta_y = \frac{y_{gt} - y_{pred}}{d}, \Delta_z = \frac{z_{gt} - z_{pred}}{\hat{d}},$$
$$\Delta_w = \log(\frac{w_{gt}}{w_{pred}}), \Delta_l = \log(\frac{l_{gt}}{l_{pred}}), \Delta_h = \log(\frac{h_{gt}}{h_{pred}}), \tag{5}$$
$$\Delta_\theta = \sin(\theta_{gt} - \theta_{pred}),$$

where Δ is the residual between the ground truth and the prediction, the superscript gt and $pred$ denote the ground truth

and the prediction from the model, respectively. The regression loss of the box localization is computed by

$$\mathcal{L}_{reg} = \sum_{\lambda \in (x,y,z,w,l,h,\theta)} \mathcal{F}_{SmoothL1}(\Delta_\lambda), \tag{6}$$

$$\mathcal{F}_{SmoothL1}(x) = \begin{cases} 0.5x^2 & \text{if} |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}. \tag{7}$$

To address the imbalance of the object samples in dataset, the focal loss in [26] is adopted as the classification loss of target objects, which can be expressed by

$$\mathcal{L}_{cls} = \mathcal{L}_{focal} = -\alpha(1 - q_{pred})^\gamma \log(q_{pred}) \tag{8}$$

where q_{pred} is the model's estimated probability, α and γ denote the parameters for the focal loss. Therefore, the overall loss can be calculated by

$$\mathcal{L}_{total} = \frac{1}{N}(\beta_{reg}\mathcal{L}_{reg} + \beta_{cls}\mathcal{L}_{cls}) \tag{9}$$

where N denotes the number of positive anchors, β_{reg} and β_{cls} are the parameters for the regression loss and the classification loss, respectively.

The cooperative perception model is trained by the Adam optimizer with adaptive learning rate and decay weight to minimize the total loss $\mathcal{L}_{total}$. This optimization process has been widely utilized by previous works as a solid framework for training an effective cooperative perception model, which will be adopted as the baseline for comparison in this work.

IV. ADAPTIVE WEIGHTING

In this section, the proposed CAV-level adaptive weighting algorithm will be discussed first. Then, the self-supervised training algorithm without labeled data will be presented.

A. CAV-level adaptive weighting

As shown in Fig. 2, the adaptive weighting is employed at the ego vehicle after receiving the features from other CAVs. This weighting algorithm is designed to exclude severely distorted information from certain CAVs or outliers in the fusion while leveraging the multiple viewpoints from other

CAVs, with less distorted information having higher weights. This requires the weighting algorithm to generate weights for features from different CAVs. The weighted feature from the k-th CAV can be computed by

$$\mathcal{W}_k = \mathcal{F}_{weighting}(\mathbf{f}_{ego}, \hat{\mathbf{f}}_k) \qquad (10)$$

$$\tilde{\mathbf{f}}_k = \mathcal{W}_k \hat{\mathbf{f}}_k \qquad (11)$$

where $\mathbf{f}_{ego}$ and $\hat{\mathbf{f}}_k \in \mathbb{R}^{B \times C \times H \times W}$ denote the intermediate features from the ego vehicle and the k-th CAV, respectively, $\mathcal{F}_{weighting}$ denotes the weighting algorithm, $\mathcal{W}_k \in \mathbb{R}^{B \times 1 \times 1 \times 1}$ denotes the output weight and $\tilde{\mathbf{f}}_k$ denotes the weighted feature. Fig. 3 demonstrates the working flow of the adaptive weighting
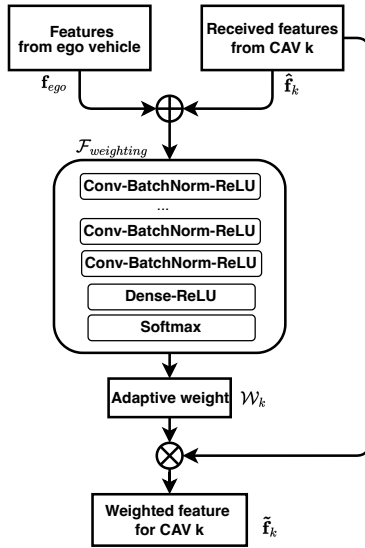

Fig. 3: The illustration of adaptive weighting for the received features from other CAVs.

for the received features from other CAVs. The input for the weighting algorithm is constructed by concatenating the features from the CAV and the ego vehicle. The underlying motivation behind this design is to effectively utilize the contrasting information in these two features. Despite having distinct viewpoints, the ego vehicle and CAV typically demonstrate similar data distribution and scales for the intermediate feature, particularly under optimal communication channel conditions. However, in scenarios where the communication channel is severely corrupted, there is a substantial difference between the CAV feature and the ego vehicle feature. Therefore, the contrasting characteristics between these two features could offer valuable insights for the weighting algorithm to generate adaptive weights.

In order to exploit these contrasting features, we use 2D CNN combined with batch normalization (BatchNorm) and ReLU activation to process the input. Multiple layers of CNN-BatchNorm-ReLU could enhance feature learning and allow for effective gradient propagation through deep networks. Moreover, BatchNorm and ReLU are essential due to its capability of preventing gradient vanishing. Instead of generating pixel-wise and channel-wise fusion weights, as in [10] and [14], we adopt CAV-level weighting that generates a value between 0 to 1 for each CAV for computational

efficiency, which is similar to Who2com [22] but focuses on mitigating the channel distortion. To achieve this, we flatten the multi-dimensional latent features processed by the CNN-BatchNorm-ReLU blocks. Then, linear dense layers and ReLU are used to facilitate fast convergence and mitigate the gradient vanishing. For the output layer, Softmax is employed to generate the regression results instead of using Sigmoid, as Sigmoid leads to severe gradient vanishing in self-supervised models without ground truth labels. Softmax is often utilized in multi-classification tasks to produce a probability distribution where the sum of the probabilities of all classes equals 1. In this work, we define two classes representing the positive or negative impact on the performance. The probability of the positive class is used as the output weight for the feature of the CAV.

B. Self-supervised optimization

The primary purpose of this weighting model is to mitigate the negative effect of channel distortion when it is severe, while minimizes its influences when the channel conditions are good. This requires the weighting model to adapt to the shared information with different levels of communication channel distortion. Unlike the works in [10], [14], and [22], which rely on the supervision of annotated datasets to determine weights for cooperative perception, we adopt self-supervised training that only requires the shared information and its augmentations for scalability and data efficiency. Specifically, a tailored self-supervised loss to optimize the weighting model $\mathcal{F}_{weighting}$ is proposed, which can be expressed by

$$\mathcal{L}_{self} = \frac{1}{K}(\lambda_{pos}\mathcal{L}_{pos} + \lambda_{neg}\mathcal{L}_{neg}), \qquad (12)$$

$$\mathcal{L}_{pos} = \sum_{k=1}^{K} \mathcal{D}_{kl}[\mathcal{S}(\mathcal{W}_k^+ \mathbf{f}_k^+)||\mathcal{S}(\mathbf{f}_k)], \qquad (13)$$

$$\mathcal{L}_{neg} = \sum_{k=1}^{K} \mathcal{D}_{kl}[\mathcal{S}(\mathcal{W}_k^- \mathbf{f}_k^-)||\mathcal{S}(\mathbf{f}_k)], \qquad (14)$$

where K is the number of CAVs, $\mathcal{L}_{pos}$ and $\mathcal{L}_{neg}$ denote the positive loss and negative loss with hype-parameter λ_{pos} and λ_{neg}, respectively, $\mathcal{S}$ denotes the Softmax function, $\mathcal{D}_{kl}(\cdot)$ denotes the Kullback–Leibler (KL) divergence, $\mathbf{f}_k^+$ and $\mathbf{f}_k^-$ denote the positive and negative augmentation experiencing minor and critical channel distortions, respectively, and $\mathcal{W}_k^+$ and $\mathcal{W}_k^-$ denote the corresponding adaptive weights.

KL divergence enables the comparison between two probability distributions with different scales, offering a means to quantify the extent to which the distribution of the weighted feature deviates from that of the transmitted feature. Instead of depending on external ground-truth annotations for object detection as in supervised learning, self-supervised learning leverages the shared information itself to generate internal supervisory signals for training models. By simulating light and severe channel distortions for the positive and the negative augmentation, respectively, their contrastive information can enhance the adaptability to various channel conditions. Specifically, minimizing both positive and negative loss in the same iteration allows the weighting model to generate

adaptive weights for features with different distortion levels. Ideally, the output weight for the less distorted feature should be close to 1, while the output for the severely polluted features should be close to 0 to minimize its negative effects. Furthermore, the self-supervised method is independent of the supervised training, allowing for a flexible extension to diverse communication environments without any modifications to the backbone network.

C. Training strategy

Three training schemes are considered to evaluate the performance of cooperative perception with V2V communications. *Training Scheme 1* assumes the ideal communication for the cooperative perception without an adaptive weighting module, which is widely used in the benchmarks in [11], [28]. Different from *Training Scheme 1* that ignores channel distortion in training, *Training Scheme 2* adopts a distortion-in-the-loop training strategy by incorporating a communication channel model into the cooperative perception system. Since we cannot assume the realistic channels in the training stage, this communication channel is usually mathematically simulated. In this case, the cooperative perception is trained in an end-to-end manner by using the supervised loss function in (9) and applying a simulated distortion to the shared information to enhance the model robustness, which is adopted in [16].

Unlike *Training Scheme 1* and *2* focus on the supervised optimization for the cooperative perception without weighting module, the proposed weighting module is introduced and optimized by *Training Scheme 3*. One prerequisite for this optimization is a pre-trained cooperative perception by *Training Scheme 2*, which generates the shared information for self-supervised training. However, the ground-truth annotations for the detection tasks are not required for training the adaptive weighting module. The proposed adaptive weighting is trained solely with the self-supervised loss in (12) using the shared information and its simulated augmentations, while the parameters of this pre-trained model will not be updated at this stage. The purpose of this design is to dynamically address the variations of channel distortion by adaptive weighting without affecting the cooperative detection systems. Additionally, the self-supervised training without annotated datasets contributes to data and computation efficiency, enabling flexible adaption to diverse communications environments.

In this work, the performance of different training schemes will be evaluated in Section V-D, and the training details will be given in Section V-A. *Training Scheme 2* will be adopted as benchmarks without adaptive weighting for comparison in other scenarios in Section V.

V. NUMERICAL RESULTS AND DISCUSSION

In this section, the simulation settings will be presented first. Then, the proposed adaptive weighting model will be evaluated for various effects.

A. Simulation settings

The simulation settings are as below:

1) Dataset: To conduct the training and evaluation, we utilize the OPV2V dataset proposed in [11], which is constructed using OpenCDA simulation tool [27]. OPV2V consists of the default CARLA towns and the Culver city dataset. The default CARLA towns consists of 6,765 samples for training and 1980 samples for validation, while the Culver city has 550 samples to evaluate the domain adaptability of the proposed model as the test set for generalization ability. In addition to the simulation dataset, V2V4Real dataset [28] is adopted as the real-world dataset, which is collected by two vehicles considering diverse real-world scenarios.

2) Baseline: Attentive fusion [11] with SECOND detection backbone [4] is employed as the baseline method for the cooperative perception. As a benchmark, the single-vehicle perception using only the sensed information from the ego car itself is adopted. Additionally, PointPillars [3] with attentive fusion [11] and V2VNet [7] are implemented to evaluate the scalability of the proposed methods. Average precision (AP) is adopted as the performance metric, which computes the average precision according to the recall value at different thresholds of intersection over union (IoU).

3) Communication settings: Rician fading with Rician K factor of 1 and free space path loss are considered to simulate the LOS channel. A Gaussian disturbance with mean 0 and variance 0.1 is added to the CSI to simulate the imperfect CSI for the Rician fading channel. Additive white Gaussian noise is considered with the signal-to-noise ratio (SNR) from -10 dB to 30 dB. Moreover, various path loss factors from 1 to 3 are also considered. To account for the realistic communication channel, we use the WINNER II channel [25] in an OFDM system with 64 sub-carriers to consider the multi-path effects. The carrier frequency is 2.6 GHz, the number of paths is 24, and the maximum delay is 16. To consider the channel estimation error, the least-square channel estimator with different pilots is also taken into account. It is important to note that the simulated Rician fading channel is used in training, while the WINNER II channel is only applied in inference to validate the performance of the proposed method as unseen realistic distortions.

4) Training: For *Training Scheme 1*, ideal communication is considered for the cooperative perception without adaptive weighting module, the same as the benchmarks in [11], [28]. For *Training Scheme 2*, a Rician fading channel with an SNR at 15 dB is simulated for the distortion-in-the-loop training, adopted as the baseline for cooperative perception without weighting. The reason for adopting a mild distortion in this training is to enhance the model's robustness to distortions while preventing the overfitting on the patterns of the distorted samples. To train the proposed adaptive weighting by *Training Scheme 3*, the pre-trained model by *Training Scheme 2* is used to generate the shared information without updating its parameters. Then, a mathematically simulated Rician fading channel with an SNR of 30 dB and -10 dB is applied to the shared information to generate the positive augmentation $\mathbf{f}_k^+$ in (13) and the negative augmentation $\mathbf{f}_k^-$ in (14), respectively. Lastly, the proposed adaptive weighting is trained solely with the self-supervised loss function in (12) based on the shared information and augmentations, where λ_{pos} is 1 and λ_{neg} is

0.0001.

For the proposed adaptive weighting model, we use 4 layers of CNN-BatchNorm-ReLU blocks to process the inputs, and then flatten the convoluted features by one layer of Dense-ReLU block. Finally, a Softmax is used to produce a probability distribution for two classes, positive and negative. The probability of the positive class is used as the output weight for the feature of the CAV. The computational resource for the implementation, training and evaluation of the cooperative perception is provided by Oracle Cloud Infrastructure with NVIDIA Tesla V100 GPUs.

B. Performance in Rician fading channel

Fig. 4 shows the performance of cooperative perception in the presence of Rician fading, path loss and noise, where both perfect and imperfect CSI are considered.

In Fig. 4(a), we evaluate the performance on the dataset of the default towns. For the cooperative perception without adaptive weighting, it experiences a significant decline in average precision from approximately 85% to 5% for IoU=0.7 and from 90% to 10% for IoU=0.3, when the SNR decreases from 30 to -10 dB, assuming the perfect CSI. This is because the fusion is degraded by outliers with very noisy shared information. For the single-vehicle perception, it has stable performance at about 60% for IoU=0.7 and 65% for IoU=0.3 regardless of the SNRs, as it only uses its own sensed information at the ego car to avoid the distortion incurred in shared information. However, this method cannot leverage the multiple viewpoints when the channel condition improves. The cooperative perception with adaptive weighting could mitigate the adverse effects of severe channel impairments and leverage the shared information when there is limited channel distortion. Specifically, it achieves an accuracy at about 60% for IoU=0.7 and 65% for IoU=0.3 when the SNR decreases to -10 dB while having a similar performance to the non-weighted cooperative perception when the SNR is larger than 10 dB. These observations demonstrate that the weighting model outperforms the single-agent perception and the non-weighted cooperation by intelligently adapting to different levels of channel impairments.

Furthermore, obtaining the perfect CSI is not practical in realistic scenarios; thus, a Gaussian disturbance is added to CSI to simulate the imperfect CSI. Without weighting, the cooperative perception experiences around 10% performance degradation due to the imperfect CSI. However, with the implementation of adaptive weighting, the performance is nearly unaffected by imperfect CSI because of the effective mitigation of the negative effects. In this case, the system aided by adaptive weighting outperforms the non-weighted cooperation and single-agent perception across all SNRs.

Fig. 4(b) shows the performance using the dataset of the Culver city. Similar to the performance on the default dataset, the weighted cooperative perception performs better than the non-weighted systems when the SNR is less than 10 dB. When the SNR is larger than 10 dB, the perception with weighting achieves a very close accuracy to the baseline without weighting, which outperforms the single-agent perception. This also
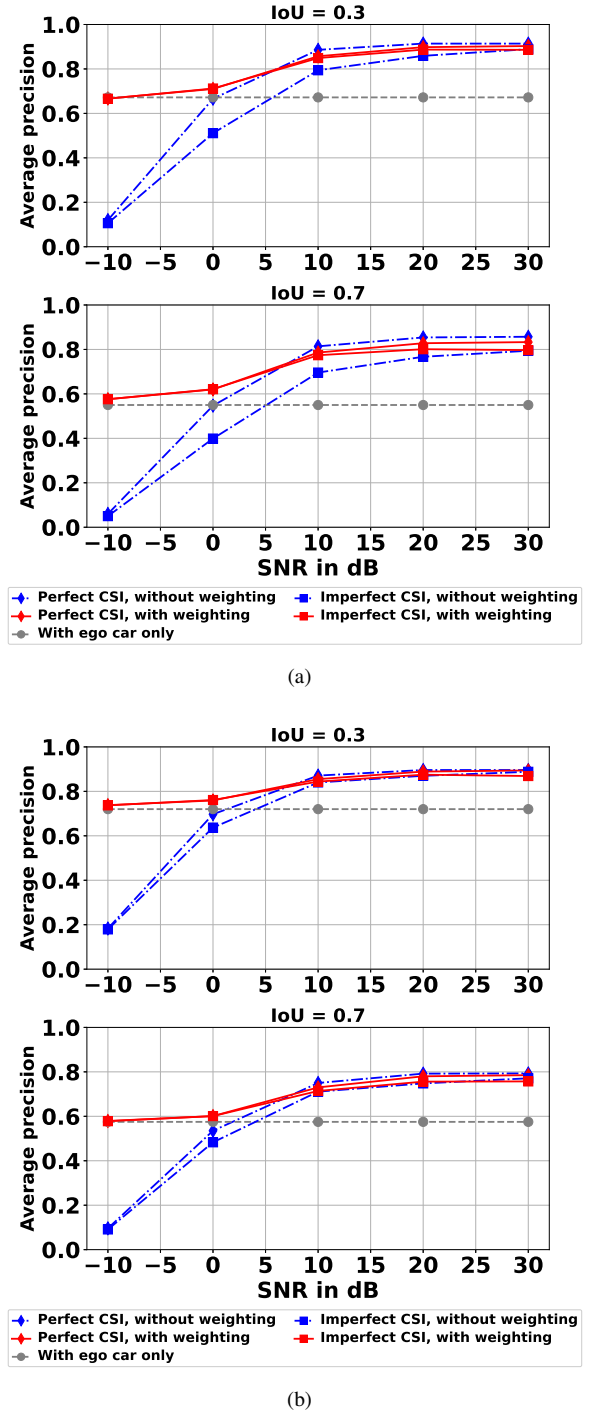

(a)

(b)

Fig. 4: Average precision with Rician fading, free path loss and noise. (a) Default towns. (b) Culver city.

validates that the proposed weighting model is well-trained and adaptive to datasets from diverse domains.

C. Performance in multi-path fading channel

Fig.5 shows the performance of cooperative perception in WINNER II channel. An OFDM system with 64 sub-carriers is adopted. Pilot-based least-square channel estimation is used to account for the channel estimation error.
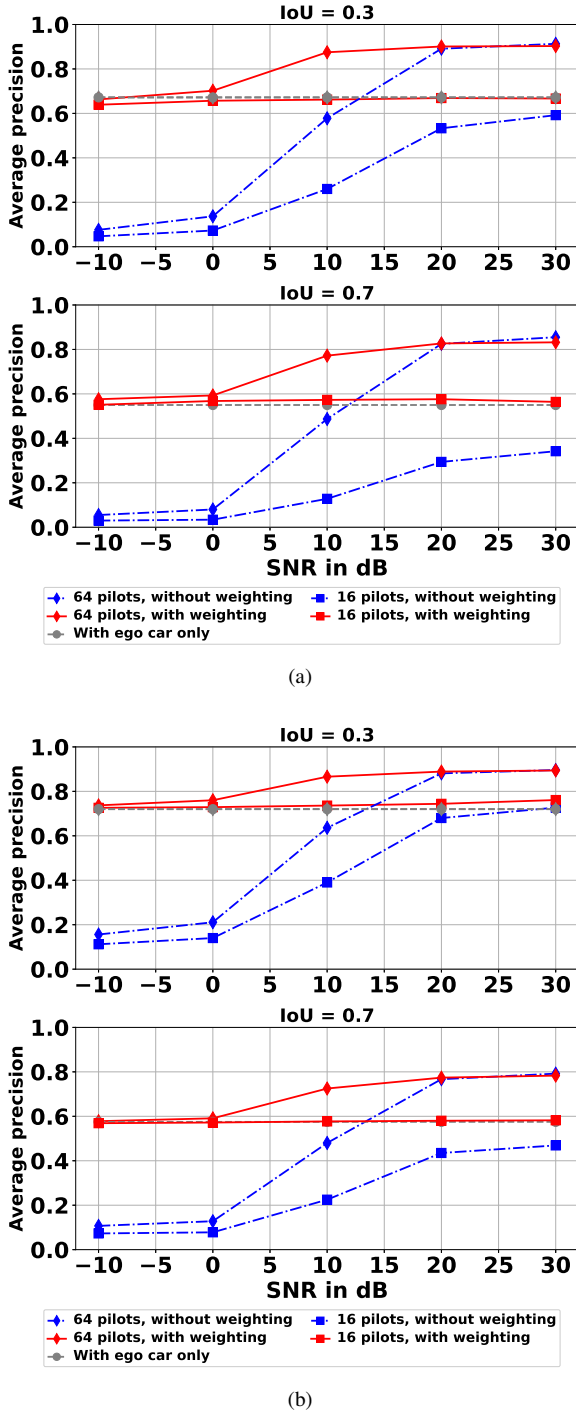
Fig. 5: Average precision with WINNER II channel. (a) Default towns. (b) Culver city.

In Fig. 5(a), when 16 pilots are used and the SNR is 30 dB, the baseline without weighting can only achieve an accuracy at approximately 30% for IoU=0.7 and 60% for IoU=0.3. However, stable performance at about 60% for IoU=0.7 and 65% for IoU=0.3 can be obtained by the proposed weighting model across all SNRs, which is almost identical to the performance with ego vehicle only but significantly outperforms the one without weighting.

If 64 pilots are used, the weighted cooperative perception

performs better than the non-weighted model when the SNR increases from -10 to 20 dB while maintaining a similar performance to the non-weighted cooperation for the SNR over 20 dB. Compared with the single-agent detection without collaboration, the non-weighted system can obtain a better accuracy only when the SNR is over 20 dB. In contrast, the weighted system performs better regardless of the SNRs. This demonstrates that the weighting algorithm works effectively on the realistic channel with multi-path fading. Furthermore, since the WINNER II channel is not used to train the weighting model, the proposed weighting model is also proved to generalize well to unseen communications channels. Similar observations can be made for the Culver city dataset in Fig. 5(b).

D. Performance for different training schemes

TABLE I: Performance of cooperative perception for different training schemes with WINNER II channel. (a) Default towns. (b) Culver city.

(a)

SNR	Training Scheme 1 AP@0.3	Training Scheme 1 AP@0.7	Training Scheme 2 AP@0.3	Training Scheme 2 AP@0.7	Training Scheme 3 AP@0.3	Training Scheme 3 AP@0.7
Ideal	0.899	0.831	**0.914**	**0.857**	0.903	0.833
30 dB	0.857	0.786	**0.913**	**0.855**	0.903	0.832
10 dB	0.089	0.054	0.578	0.487	**0.875**	**0.772**
-10 dB	0.042	0.023	0.076	0.055	**0.663**	**0.576**

(b)

SNR	Training Scheme 1 AP@0.3	Training Scheme 1 AP@0.7	Training Scheme 2 AP@0.3	Training Scheme 2 AP@0.7	Training Scheme 3 AP@0.3	Training Scheme 3 AP@0.7
Ideal	0.894	0.754	**0.897**	**0.794**	0.894	0.783
30 dB	0.862	0.720	**0.894**	**0.791**	0.894	0.783
10 dB	0.133	0.075	0.641	0.488	**0.866**	**0.725**
-10 dB	0.094	0.057	0.154	0.107	**0.737**	**0.578**

In this section, we compare the performance of cooperative perception using different training schemes. *Training Scheme 1* uses ideal communication for training, and *Training Scheme 2* adopts a distortion-in-the-loop training strategy. *Training Scheme 1* and *2* do not have an adaptive weighting module, while *Training Scheme 3* additionally applies adaptive weighting to a pre-trained model based on distortion-in-the-loop training. The performances are validated on the WINNER II channel with an SNR of 30 dB, 10 dB, and -10 dB for the unseen realistic channel with light, mild, and severe distortions, respectively.

Table I(a) presents the average precision of different training schemes on the default town dataset. *Training Scheme 1* can barely achieve an accuracy of 10% when the SNR is 10 dB and -10 dB. On the other hand, *Training Scheme 2* outperforms *Training Scheme 1* in all communication conditions, demonstrating enhanced robustness to channel distortions. Even in ideal communication scenarios, *Training Scheme 2* performs better than *Training Scheme 1*. This is because the simulated light distortion on the shared information serves as a form

of data augmentation, improving the system robustness and avoiding overfitting on the training samples. However, when there is severe channel distortion, the performance of *Training Scheme 2* experiences a significant degradation to below 10%. To address this, *Training Scheme 3* obtains the best accuracy among these three schemes when the SNR is 10 dB and -10 dB. This underscores the efficacy of *Training Scheme 3* in mitigating the adverse effects of severe channel distortion. Meanwhile, although *Training Scheme 3* has a very slight performance degradation compared with *Training Scheme 2* when the channel condition improves, *Training Scheme 3* can achieve over 90% for IoU=0.3 and 80% for IoU=0.7, outperforming *Training Scheme 1*. This observation highlights the adaptability of *Training Scheme 3* to the non-distorted communications scenarios, maintaining competitive performance to models trained under ideal communications and simulated distortions. Similar observations and conclusions can be obtained for the performance on the Culver city dataset in Table I(b), which validates the effectiveness of the proposed adaptive weighting model.
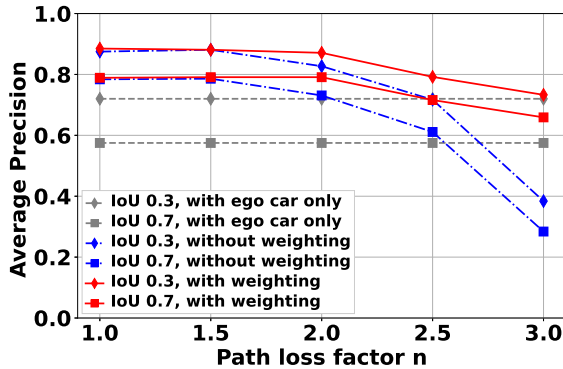
E. Performance for different path loss factors

Fig. 6: Average precision with different path loss factor n.

Fig. 6 depicts the performance curves of the cooperative perception with various path loss factors when the SNR is 30 dB with imperfect CSI. When the path loss factor increases from 1 to 3, the non-weighted cooperative perception experiences a substantial decline in accuracy, decreasing from about 80% to 30% for IoU=0.7 and from 90% to 40% for IoU=0.3, respectively. For the cooperative perception with adaptive weighting, the accuracy ranges from approximately 80% to 65% for IoU=0.7 and 88% to 73%, which is more robust than the non-weighted system. Moreover, when the path loss factor ranges from 2 to 3, the weighted model performs better than both baselines with non-weighted cooperation and single-agent perception. Therefore, the proposed weighting model can adapt to different communications environments with different path loss factors and mitigate its negative effects. This is important for autonomous driving, as the car can move from city centre to highway with different n.
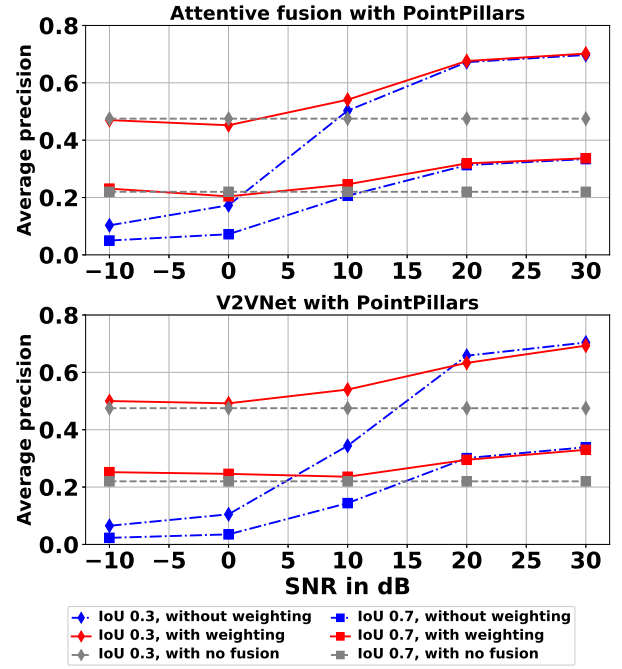
Fig. 7: Performance of the attentive fusion and V2VNet using the PointPillars backbone on the V2V4Real dataset in the WINNER II channel.

F. Performance on real-world datasets using other backbones

To evaluate the adaptive weighting in a real-world dataset and its scalability to different backbones, Fig. 7 demonstrates the performance of PointPillars [3] on the V2V4Real dataset [28] with WINNER II channel. V2VNet [7] and attentive fusion [11] are used to account for different intermediate fusion schemes.

In Fig. 7, the performance of the weighted attentive fusion with PointPillars ranges from around 47% to 70% for IoU=0.3 and 23% to 34% for IoU=0.7, which performs better than the non-weighted baselines and the single-vehicle detection. As the SNR decreases from 30 dB to -10 dB, the non-weighted collaboration experiences a significant performance degradation to be below 10%; nevertheless, the performance of the weighted model decreases slowly until it approaches the single-vehicle detection accuracy with no fusion. This indicates that the proposed weighting algorithm could effectively work with PointPillars and adapt to different channel conditions in real-world datasets. Similar observations can be made for the performance of the weighted V2VNet with PointPillars in Fig. 7. The weighted V2VNet with PointPillars outperforms the non-weighted collaboration and the single-vehicle detection by mitigating the negative effects of severe channel distortions and maintaining a close accuracy to the non-weighted collaboration when the SNR is over 20 dB. These observations demonstrate that the proposed adaptive weighting can not only generalize on the real-world dataset with communication channel distortion but is also compatible with different intermediate fusion schemes and detection backbones.
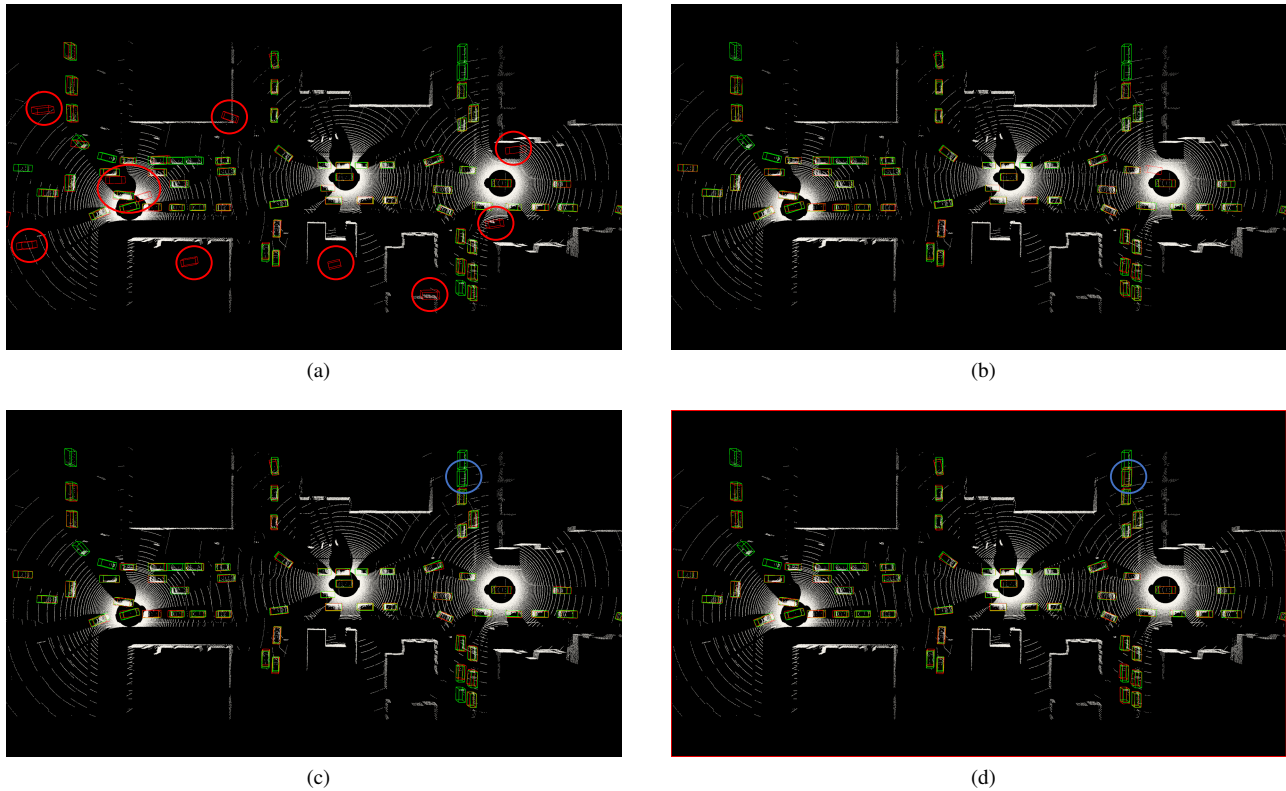
Fig. 8: Examples of cooperative perception with and without adaptive weighting on the WINNER II channel. The ground truth and predicted 3D bounding boxes are illustrated in green and red, respectively. Bold red circles highlight the false predictions of the cooperative perception without weighting when the SNR is 10 dB. Bold blue circles indicate the vehicle additionally detected by the weighted cooperative perception when the SNR is 30 dB. (a) SNR = 10 dB, without weighting. (b) SNR = 10 dB, with weighting. (c) SNR = 30 dB, without weighting. (d) SNR = 30 dB, with weighting.

G. Visualization example of adaptive weighting

In order to visualize how adaptive weighting improves cooperative perception, Fig. 8 shows examples of cooperative perception with and without weighting. When the SNR is 10 dB, the baseline without weighting is affected by the channel distortion, thus leading to false predictions shown in Fig. 8(a). However, it is demonstrated in Fig. 8(b) that adaptive weighting could eliminate the falsely predicted bounding boxes. When the SNR is 30 dB, both weighted and non-weighted systems can significantly reduce the false predictions due to the improved channel conditions, which are demonstrated in Fig. 8(c) and Fig. 8(d). Nevertheless, the weighted cooperative perception detects an additional vehicle, highlighted by the blue circle in Fig. 8(d), which the non-weighted baseline neglects. This indicates that the proposed weighting algorithm can not only mitigate the adverse effects under severe channel conditions, but also enhance the received shared features when the channel condition improves.

VI. Conclusion

In this work, we have proposed a self-supervised adaptive weighting model for intermediate fusion to mitigate the adverse effects of distorted information caused by channel impairments. The performance of our proposed weighting algorithm has been evaluated in terms of different datasets, detection backbones, fusion schemes, noise levels and path loss factors. Numerical results and visualization examples have demonstrated that the proposed adaptive weighting algorithm performs better than the benchmarks without weighting under all conditions. It is shown that the proposed weighting algorithm can mitigate the adverse effects of severe signal distortion and enhance the received shared features when the channel condition improves. Moreover, the adaptive weighting algorithm has also demonstrated good generalization to unseen realistic communications channels and test datasets from different domains. It is also validated that the proposed weighting can flexibly scale to other detection backbones, fusion schemes and real-world datasets.

References

[1] L. Chen, *et al.*, "Milestones in autonomous driving and intelligent vehicles: Survey of surveys," *IEEE Trans. Intell. Veh.*, vol. 8, no. 2, pp. 1046–1056, 2023.

[2] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4490–4499.

[3] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12 689–12 697.

[4] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, 2018.

[5] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2019, pp. 514–524.

[6] Q. Chen, X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu, "F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds," in *Proc. IEEE/ACM Symp. Edge Comput.*, 2019, pp. 88–100.

[7] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2vnet: Vehicle-to-vehicle communication for joint perception and prediction," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 605–621.

[8] Y. Liu, J. Tian, N. Glaser, and Z. Kira, "When2com: Multi-agent perception via communication graph grouping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4105–4114.

[9] Y. Hu, S. Fang, Z. Lei, Y. Zhong, and S. Chen, "Where2comm: Communication-efficient collaborative perception via spatial confidence maps," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 4874–4886.

[10] Y. Li, S. Ren, P. Wu, S. Chen, C. Feng, and W. Zhang, "Learning distilled collaboration graph for multi-agent perception," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 29 541–29 552.

[11] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 2583–2589.

[12] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2x-vit: Vehicle-to-everything cooperative perception with vision transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 107–124.

[13] J. Li, R. Xu, X. Liu, J. Ma, Z. Chi, J. Ma, and H. Yu, "Learning for vehicle-to-vehicle cooperative perception under lossy communication," *IEEE Trans. Intell. Veh.*, vol. 8, no. 4, pp. 2650–2660, 2023.

[14] D. Qiao and F. Zulkernine, "Adaptive feature fusion for cooperative perception using lidar point clouds," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2023, pp. 1186–1195.

[15] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1852–1864, 2022.

[16] C. Liu, Y. Chen, J. Chen, R. Payton, M. Riley, and S.-H. Yang, "Cooperative perception with learning-based v2v communications," *IEEE Wirel. Commun. Lett.*, vol. 12, no. 11, pp. 1831-1835, 2023.

[17] C. F. Mecklenbrauker, *et al.*, "Vehicular channel characterization and its implications for wireless system design and performance," in *Proc. of the IEEE*, vol. 99, no. 7, pp. 1189-1212, 2011.

[18] C.-X. Wang, X. Cheng, D. I. Laurenson, "Vehicle-to-vehicle channel modeling and measurements: recent advances and future challenges," in *IEEE Commun. Mag.*, vol. 47, no. 11, pp. 96-103, 2009.

[19] Y. Wang, A. Fathi, A. Kundu, D. A. Ross, C. Pantofaru, T. Funkhouser, and J. Solomon, "Pillar-based object detection for autonomous driving," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 18–34.

[20] S. Shi, X. Wang, and H. Li, "Pointrcnn: 3d object proposal generation and detection from point cloud," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 770–779.

[21] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "Pv-rcnn: Point-voxel feature set abstraction for 3d object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10 526–10 535.

[22] Y. -C. Liu, J. Tian, C. -Y. Ma, N. Glaser, C. -W. Kuo and Z. Kira, "Who2com: Collaborative perception via learnable handshake communication," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 6876-6883.

[23] Z. Meng, X. Xia, R. Xu, W. Liu and J. Ma, "Hydro-3d: Hybrid object detection and tracking for cooperative perception using 3d lidar," *IEEE Trans. Intell. Veh.*, vol. 8, no. 8, pp. 4069-4080, 2023.

[24] X. Xia, *et al.*, "An automated driving systems data acquisition and analytics platform," *Transp. Res. Part C: Emerg. Technol.*, vol. 151, p. 104120, 2023.

[25] P. Kyosti, et al "Winner ii channel models," *IST, Tech. Rep. IST-4-027756 WINNER II D1.1.2 V1.2*, 2007. [Online]. Available: https://cir.nii.ac.jp/crid/1571698600936523008

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

[27] R. Xu, *et al.*, "The OpenCDA Open-Source Ecosystem for Cooperative Driving Automation Research," in *IEEE Trans. Intell. Veh.*, vol. 8, no. 4, pp. 2698-2711, 2023.

[28] R. Xu, *el al.*, "V2V4Real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 13712–13722.

Chenguang Liu Chenguang Liu received his B.E. degree in software engineering from Dalian University of Technology, Dalian, P.R.China, in 2016 and M.S. degree in advanced computer science from The University of Manchester, U.K., in 2017. He is currently studying as a Ph.D. student at the University of Warwick, U.K. His research interests include deep learning, wireless communications and cooperative perception.

Jianjun Chen Jianjun Chen is working toward the PhD degree in the Australian Artificial Intelligence Institute within the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. He received the BS degree in software engineering and MS degree in computer science from the Taiyuan University of Technology, China. His research interests include Federated Learning, Cooperative Perception and 3D Object Detection.
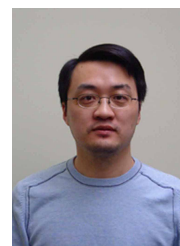
Yunfei Chen Yunfei Chen (S'02-M'06-SM'10) received his B.E.and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China,in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as a Professor at the University of Durham, U.K. His research interests include wireless communications, cognitive radios, wireless relaying and energy harvesting.

Ryan Payton Ryan Payton received his BSc degree in Environmental Geology in 2015 and PhD in Earth Sciences in 2022 from Royal Holloway, University of London. He is currently a Senior Research Advocate at Oracle within the Oracle for Research team, providing opportunities and technical support for researchers using Oracle Cloud Infrastructure across a range of scientific research disciplines.

Michael Riley Michael Riley received his BSc degree in Information Systems in 1988 form North Staffordshire Polytechnic. He has worked in various Consulting roles within Oracle for the past 25 years. He is currently a Cloud Solution Architect supporting research projects that are utilising Oracle Cloud Infrastructure for their research projects, working collaboratively as part of the Oracle for Research program.

Shuang-Hua Yang Shuang-Hua Yang (Senior Member, IEEE) received the B.S. degree in instrument and automation and the M.S. degree in process control from the China University of Petroleum, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree in intelligent systems from Zhejiang University, Hangzhou, China, in 1991. He is currently the director of the Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, and a Professor and the Head of the Department of Computer Science, University of Reading, U.K. His research interests include Internet of Things, industrial internet, Cyber-Physical System safety and security, Process Systems Engineering. Prof. Yang is a fellow of IET and InstMC, U.K. He is an Associate Editor of IET Cyber-Physical Systems: Theory and Applications.