

RVSL: Robust Vehicle Similarity Learning in Real Hazy Scenes Based on Semi-supervised Learning

Wei-Ting Chen^{1*}, I-Hsiang Chen^{2*}, Chih-Yuan Yeh², Hao-Hsiang Yang², Hua-En Chang², Jian-Jiun Ding², and Sy-Yen Kuo²

¹ Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan

² Department of Electrical Engineering, National Taiwan University, Taiwan
{f05943089, f09921058, f09921063}@ntu.edu.tw,
islike8399@gmail.com, {r10921a35, jjding, sykuo}@ntu.edu.tw
[https://github.com/Cihsaing/
rvsl-robust-vehicle-similarity-learning--ECCV22](https://github.com/Cihsaing/rvsl-robust-vehicle-similarity-learning--ECCV22)

Abstract. Recently, vehicle similarity learning, also called re-identification (ReID), has attracted significant attention in computer vision. Several algorithms have been developed and obtained considerable success. However, most existing methods have unpleasant performance in the hazy scenario due to poor visibility. Though some strategies are possible to resolve this problem, they still have room to be improved due to the limited performance in real-world scenarios and the lack of real-world clear ground truth. Thus, to resolve this problem, inspired by CycleGAN, we construct a training paradigm called **RVSL** which integrates ReID and domain transformation techniques. The network is trained on semi-supervised fashion and does not require to employ the ID labels and the corresponding clear ground truths to learn hazy vehicle ReID mission in the real-world haze scenes. To further constrain the unsupervised learning process effectively, several losses are developed. Experimental results on synthetic and real-world datasets indicate that the proposed method can achieve state-of-the-art performance on hazy vehicle ReID problems. It is worth mentioning that although the proposed method is trained without real-world label information, it can achieve competitive performance compared to existing supervised methods trained on complete label information.

Keywords: Hazy Vehicle Similarity Learning, Semi-supervised Learning, Image Dehazing

1 Introduction

Vehicle similarity learning, also called vehicle re-identification (ReID), is a crucial technique for intelligent surveillance systems in a smart city. It is to track the vehicles with the same identity within a set of images captured by multiple cameras and various viewpoints. With the development of the deep convolutional neural network (DCNN),

*Indicates equal contribution.

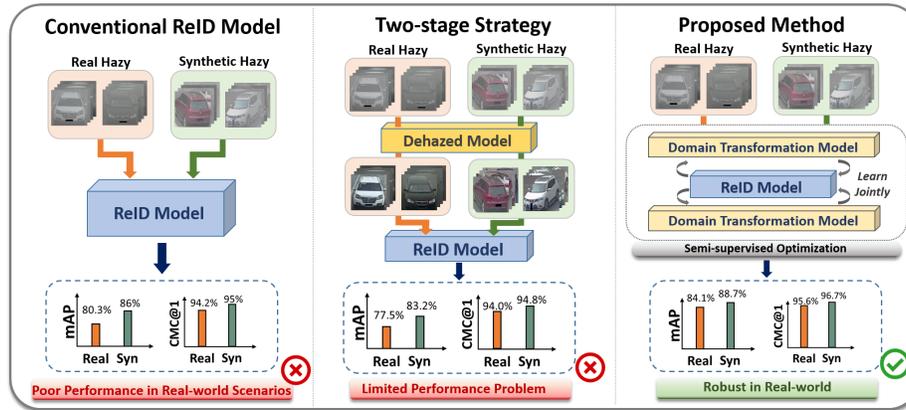


Fig. 1: **Illustration of different strategies to solve hazy vehicle ReID problem.** One can see that our method outperforms other existing methods in terms of the mean average precision (mAP) and CMC@1. Moreover, other strategies may have limited performance problem in real-world scenarios. We adopt CAL [33] and MPR-Net [39] as the ReID model and the dehazed model, respectively.

several approaches for vehicle ReID [13,18,31,41] have been proposed and achieved impressive performance. Similar to other high-level vision applications such as object detection and semantic segmentation [5], although existing methods can handle vehicle ReID effectively on normal images, they have limited performance under inclement weather, especially in hazy scenario. Haze is a common and inevitable weather phenomenon that leads to poor visual appearances and causes the loss of discriminative information by deteriorating the contents of images for vehicle ReID. Thus, this field still has room for improvement.

Inspired by previous dehazing tasks [2], we can apply an atmospheric scattering model [23] to synthesize haze images and then train the vehicle ReID models based on the rendered images and the corresponding ID labels. Though this strategy can achieve decent performance on synthetic images, they have limited performance on real haze images due to the domain gap between synthetic and real-world images [3]. While this issue can be resolved by adopting real haze images in the training stage, collecting the real haze data and labeling the correct ground truths are difficult and troublesome.

Another possible baseline strategy is to adopt the existing dehazing approaches [4,40] or comprehensive image restoration method [39] as the pre-processing technique and then apply the ReID. Although the above strategies are shown to achieve promising results in haze removal, there is no guarantee that the selected pre-processing techniques would be able to improve ReID, since these two tasks are performed separately and existing dehazing methods are not designed for the purpose of ReID but for human perception. Moreover, most existing dehazing methods require pair data to train the model, but it is infeasible to attain the ground truths of haze images in real-world scenes. Though we can adopt synthetic data to train the network, the domain gap prob-

lem may still exist, which may generate undesired dehazed results in real haze scenes and further limit the performance of ReID.

By the above analysis, there are two reasons that hinder the development of ReID in real haze scenes: (i) the scarcity of labels of real-world data and (ii) the lack of appropriate guidance for real haze. In this paper, to mitigate these problems, we construct a novel training architecture based on the deep convolutional neural network (DCNN). Inspired by CycleGAN [21] which can transform images between any two domains, we introduce the domain transfer technique in the proposed network and combine it with the vehicle ReID. Specifically, the proposed method is trained on a semi-supervised paradigm in an end-to-end fashion and there are two parts in the training process: supervised training for synthetic data and unsupervised training for real-world data. For the former part, the network can learn the knowledge of transformation between two domains and extract more discriminative features for vehicle ReID in fully supervision by paired data (i.e., synthetic hazy images and the corresponding clear ground truths). For the latter parts, we only leverage two sets of unpaired data (i.e., real hazy images and clear images) to strengthen the robustness of the domain transformation and the ReID in real-world scenes in an unsupervised manner.

The idea of our method is that, the domain transformation network transfers the input image (i.e., hazy or clear images) between two domains with the same background information and the ReID network extracts the latent features for classifying from two images (i.e., input and transferred image). Inspired by the cycle consistency [21,37], two extracted embedding features should be identical since they are from the same vehicles. Thus, we can calculate the consistency of between two extracted features for optimizing the network.

Based on the semi-supervised training scheme, the utilization of the synthetic data can guide the unsupervised stage and prevent the network from unstable performance [21]. On the other hand, the use of real-world data can improve the generalization ability of our model to real data and further mitigate the domain gap problem when the synthetic data is applied in the training process [25]. Moreover, using the domain transformation network can assist the ReID network to learn more discriminative features for the ReID under real-world haze scenarios. By our design, the proposed method can perform vehicle ReID in hazy scenarios effectively without additional annotations in real hazy data which are usually hard to be obtained. Furthermore, our proposed training scheme can be also applied in the case that we have annotations of vehicle ID and achieve better performance.

The contribution of this paper is summarized as follows.

- A novel training paradigm based on semi-supervised learning and domain transformation is proposed to learn hazy vehicle ReID without the labels or clear ground truths of real-world data. We term it **Robust Vehicle Similarity Learning (RVSL)**. As depicted in Fig. 1, by combining domain transferring technique with the ReID network, the proposed method can achieve decent performance to learn discriminative features under real haze scenes without using ID label. Surprisingly, the proposed method achieves competitive performance compared with other existing methods trained with complete ID information.

- To constrain the unsupervised stage in the training process, we developed several loss functions such as embedding consistency loss, colinear relation constraint, and monotonously increasing dark channel loss to improve the performance. These loss functions enable the network to learn both domain transformation and ReID in an unsupervised way effectively. Experimental results prove the effectiveness of these loss functions.

2 Related Works

Vehicle Re-identification. With the great effort of data collection and annotation, several large-scale benchmarks for vehicle ReID such as VehicleID [28], VeRi-776 [29], VERI-Wild [30], and Vehicle-1M [12]) are proposed. Based on these well-developed benchmarks, several approaches [19,9,17,26,33,35] have been developed and most of them rely on DCNN. We can divide them into the following categories.

1) *Meta-information-based methods* which integrate meta-information for feature learning. For example, Zheng *et al.* [42] leveraged the additional information such as the camera view, and the vehicle type and color to guide the network. Shen *et al.* [34] integrated the visual-spatio-temporal path proposals and spatial-temporal relations to a Siamese-CNN+Path-LSTM network. Rao *et al.* [33] proposed the attention mechanism with counterfactual causality which enables the network to learn more useful attention for fine-grained features for ReID.

2) *Local information-based methods*: Meng *et al.* [31] adopted the common region information extracted by a vehicle part parser to improve the mutual representation information between different viewpoints. Khorramshahi *et al.* [22] applied Variational Auto-Encoder (VAE) to find crucial detailed information which can be regarded as the pseudo-attention map for highlighting discriminative regions. He *et al.* [13] combined local and non-local information based on a part-regularized mechanism. These strategies can preserve the variance from near-duplicate vehicles to improve the performance of vehicle ReID. Zhao *et al.* [41] applied Cross-camera Generalization Measure technique and integrated region-specific features and cross-level features together to improve the performance of ReID.

3) *Generative Adversarial Network (GAN) based methods*: Zhou *et al.* [43] applied the conditional multi-view generative network to extract global feature representation from various viewpoints and then adopted adversarial learning to facilitate feature generation. Lou *et al.* [30] designed the FDA-Net to generate hard examples in the feature space based on the GAN to improve the robustness of ReID. Yao *et al.* [38] proposed to adopt a 3D graphic engine to reduce the content gap between the existing datasets to suppress the domain gap problem.

4) *Vision Transformer (ViT) based Methods*: He *et al.* [18] leveraged the ViT to encode input images as a vector for embedding representation. To further improve representation learning, the jigsaw patch module and side information were adopted in the training scheme.

Though the above methods can achieve decent vehicle ReID performance on clear images, they are still limited in real-world hazy image scenarios.

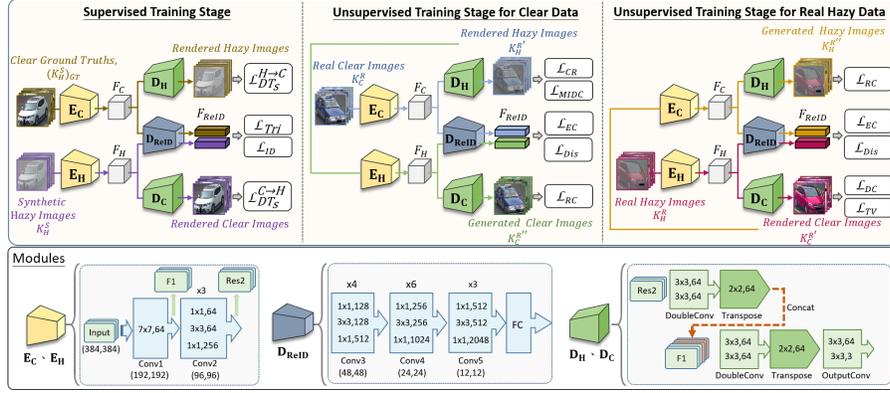


Fig. 2: **The architecture of the proposed semi-supervised hazy vehicle ReID network.** Our method consists of supervised and unsupervised training stages for synthetic and real-world data.

Single Image Haze Removal. Based on Koschmieder’s model [23], the formation of haze can be modeled by:

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (1)$$

where $I(x)$ is the hazy image, $J(x)$ is the haze-free image and A is the global atmospheric light. $t(x) = e^{-\beta d(x)}$ is the medium transmission map where β is the scattering coefficient and $d(x)$ is the depth from the camera to the object. There are numerous haze removal methods proposed in past decades. They can be classified into prior-based and deep learning-based methods. The former class is to explore the prior knowledge between hazy and haze-free images. For example, He *et al.* [14] proposed the dark channel prior, Zhu *et al.* [44] developed the color attenuation prior, and Berman *et al.* proposed the haze-line [1] to estimate the dehazed results. The other class is to apply the DCNN. For instance, Qu *et al.* [32] proposed multi-resolution generators and discriminators for dehazing in a coarse-to-fine way. Dong *et al.* [8] used the strengthen-operate-subtract boosting strategy to improve the dehazing network. Wu *et al.* [36] proposed an auto-encoder-like framework with additive mixup operation and a dynamic feature enhancement module to improve the quality of extracted features for dehazing. Zamir *et al.* [39] proposed a multi-stage architecture that can encode a diverse set of features simultaneously to restore accurate outputs. Chen *et al.* [6] proposed a unified architecture which can learn multiple adverse weather based on a single architecture.

3 Proposed Method

3.1 Overview of the Proposed Method

As shown in Fig. 2, there are five modules in the proposed network. That is, two encoders for hazy and clear scenes (E_H and E_C), and three decoders for hazy, clear, and

ReID (\mathbf{D}_H , \mathbf{D}_C , and \mathbf{D}_{ReID}), respectively. These modules can be combined to two sub-networks called domain transformation network and re-identification network. The features extracted by \mathbf{E}_H and \mathbf{E}_C are termed F_H and F_C , respectively.

As mentioned in section 1, due to the lack of real haze data, in this paper, inspired by semi-supervision [25], we apply both synthetic data and real data simultaneously in the learning process. At the supervised training stage, the application of the synthetic data can learn the transformation from various domains stably. At the unsupervised stage, we first take the real clear images and then take real hazy images as inputs, respectively. This operation enables our network to learn real-world information of hazy scenes, clear scenes, and ReID through the domain transformation network and ReID network simultaneously. We illustrate the details in the following subsections.

3.2 Domain Transformation Network.

The goal of the domain transformation network (DT-Net) is to help the ReID network to learn haze-invariant features via transforming the domain of the input data. The detailed illustration of this network is as follows.

Architecture. The DT-Net consists of two encoders (\mathbf{E}_H and \mathbf{E}_C) and two decoders (\mathbf{D}_H and \mathbf{D}_C). Given a hazy input, the decoder \mathbf{D}_C generates the corresponding clear image based on the features F_H extracted by encoder \mathbf{E}_H . On the other hand, the decoder \mathbf{D}_H takes the features F_C extracted from the encoder \mathbf{E}_C to produce the hazy image. The features (i.e., F_C and F_H) extracted by two encoders pass through a double convolution block and a deconvolution block for dimension matching. Then, the up-sampled features are concatenated with the features extracted by the first convolution blocks in the encoders to improve the feature diversity. This operation is based on the fact that the features in the shallow layer of the network contain more fruitful spatial and contextual information which can benefit domain transformation [7,20] while the deeper layers usually consist of more high-level vision. The concatenated features are passed through a double convolution block and a deconvolution block to reconstruct the final domain transformation results. The quality of domain transformation is crucial for the ReID network since it may affect the feature extraction of the input image. Thus, for the synthetic data, we adopt the supervised loss \mathcal{L}_{DT_s} to optimize the networks. For real data, we adopt unsupervised losses \mathcal{L}_{DT_u} .

Supervised Training Stage. At this stage, we can train the network in a fully supervised way since the corresponding clear ground truths and ID labels are available. First, we adopt synthetic image pairs to train the DT-Net (i.e., \mathbf{E}_H , \mathbf{E}_C , \mathbf{D}_H , and \mathbf{D}_C). Specifically, the synthetic haze image K_H^S and the corresponding clear ground truth $(K_H^S)_{GT}$ are fed into the domain transformation network to calculate the domain transformation loss \mathcal{L}_{DT_s} for synthetic data. This operation aims to constrain the distance between the predicted results (i.e., the rendered hazy images and the rendered clear images) and the corresponding ground truths. The domain transformation loss \mathcal{L}_{DT_s} can be formulated as follows.

$$\mathcal{L}_{DT_s}^{H \rightarrow C} = \frac{1}{M} \sum_{i=1}^M \|\mathbf{D}_C[\mathbf{E}_H[K_H^S(i)]] - (K_H^S)_{GT}(i)\|_1 \quad (2)$$

$$\mathcal{L}_{DT_s}^{C \rightarrow H} = \frac{1}{M} \sum_{i=1}^M \|\mathbf{D}_H[\mathbf{E}_C((K_H^S)_{GT}(i))] - K_H^S(i)\|_1 \quad (3)$$

where $\|\cdot\|_1$ presents the L_1 norm and M indicates the number of images. $\mathcal{L}_{DT_s}^{C \rightarrow H}$ and $\mathcal{L}_{DT_s}^{H \rightarrow C}$ denote the domain transformation loss for 'clear to haze' and 'haze to clear', respectively. The \mathcal{L}_{DT_s} loss is the summation of $\mathcal{L}_{DT_s}^{C \rightarrow H}$ and $\mathcal{L}_{DT_s}^{H \rightarrow C}$.

Unsupervised Training Stage for Clear Data. At this stage, to train the DT-Net without the hazy image ground truths, first, we adopted the cycle-consistency mechanism. The input clear image K_C^R is fed into the DT-Net to render the hazy image $K_H^{R'}$, where $K_H^{R'} = \mathbf{D}_H(\mathbf{E}_C(K_C^R))$. Then, we further take the rendered image to the DT-Net to generate the rendered clear image $K_C^{R''}$, where $K_C^{R''} = \mathbf{D}_C(\mathbf{E}_H(K_H^{R'}))$. In the same time, several loss functions are adopted to optimize the network. The loss at this stage $\mathcal{L}_{DT_{uc}}$ consists of the rendering consistency loss (\mathcal{L}_{RC}), the monotonously increasing dark channel loss (\mathcal{L}_{MIDC}), the colinear relation constraint (\mathcal{L}_{CR}), and the discriminative loss (\mathcal{L}_{Dis}). We illustrate each of them as follows.

(i) *Rendering Consistency Loss.* This loss is to constrain the learning process of the domain transformation network (i.e., \mathbf{E}_H , \mathbf{E}_C , \mathbf{D}_H , and \mathbf{D}_C). We adopt the pixel-wise difference between the clear input image K_C^R and the rendered clear image $K_C^{R''}$ to ensure that the domain transformation process can be conducted in two different domains robustly. This loss is formulated as follows.

$$\mathcal{L}_{RC} = \frac{1}{M} \sum_{i=1}^M \|K_C^R(i) - K_C^{R''}(i)\|_1 \quad (4)$$

(ii) *Monotonously Increasing Dark Channel Loss.* To further improve the image quality of rendered haze images, inspired by dark channel prior (DCP) [14], we propose the monotonously increasing dark channel loss \mathcal{L}_{MIDC} . The DCP demonstrates that for most natural clear images, the dark channel values may be close to zero. Specifically, it can be defined as:

$$DC(J)(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r, g, b\}} J^c(y) \right) \approx 0, \quad (5)$$

where $DC(\cdot)$ is the operation of the dark channel, $J^c(y)$ is the intensity in the color channel c , and $\Omega(x)$ is a local patch with a fixed size centered at x . With this prior, it can be further extended that, given an image deteriorated by haze, its dark channel value may be higher than that of the original clear image (i.e., $DC(I)(x) \geq DC(J)(x)$). Based on this idea, we proposed \mathcal{L}_{MIDC} which is determined as:

$$\mathcal{L}_{MIDC} = \frac{1}{M} \sum_{i=1}^M DM(i) \|DC(K_H^{R'}(i)) - DC(K_C^R(i))\|_1 \quad (6)$$

where $DM(i)$ is a binary map that identifies the region where the dark channel values of the clear image K_C^R are higher than that of its rendered haze result $K_H^{R'}$ (i.e., $DC(K_H^{R'})(x) < DC(K_C^R)(x)$). With it, we can prevent the rendered pixels from irrational results and further improve the robustness of domain transformation.

(iii) *Colinear Relation Constraint.* Due to the inaccessibility of the ground truth of real-world hazy images, the domain transformation process may generate undesired fake image contents without appropriate training. Although we can leverage the synthetic data to guide the network at the training stage of the synthetic data, it is not applicable at the training stage of the real-world data. To further strengthen the robustness of transformation (i.e., \mathbf{E}_C and \mathbf{D}_H), inspired by the haze-line prior [1,37], we develop colinear relation constraint \mathcal{L}_{CR} .

Based on the physical model of haze illustrated in (1), Berman *et al.* [1] observe that the clear image J , the hazy image I , and the atmospheric light A are colinear in the RGB space (i.e., $I(x) - A = t(x)(J(x) - A)$). We can adopt this relation to constrain the training process of the network for real-world scenarios. We define the colinear relation constraint as follows.

$$\mathcal{L}_{CR} = \frac{1}{M} \sum_{i=1}^M \left[1 - \phi(K_C^R(i) - A(i), K_H^{R'}(i) - A(i)) \right], \quad (7)$$

where $A(i)$ is the atmospheric light estimated by the rendered hazy image $K_H^{R'}(i)$ and $\phi(\cdot)$ means the cosine similarity. Different from [37], we adopt the atmospheric light estimation method in [14] in this loss. With this loss, the consistency of structure and color can be further constrained.

(iv) *Discriminative Loss.* To further constrain unsupervised domain transformation, we adopt the discriminative loss [10] in the training process to distinguish whether the rendered hazy image $K_H^{R'}$ is real or fake. In our method, we adopt the saturating discriminative loss [11].

Unsupervised Training Stage for Real Hazy Data. At this stage, the real hazy images are adopted to optimize the network without clean ground truths and ID labels. Like the previous stage, the hazy images are fed into the DT-Net (i.e., \mathbf{E}_H and \mathbf{D}_C) to generate the clear images $K_C^{R'}$. Subsequently, the rendered clear images are fed into the DT-Net to obtain the rendered hazy images $K_H^{R''}$. To optimize our framework, apart from the monotonously increasing dark channel loss (\mathcal{L}_{MIDC}) and colinear relation constraint (\mathcal{L}_{CR}), the rest of losses in $\mathcal{L}_{DT_{uc}}$ are adopted. Moreover, to improve the predicted clear images by the DT-Net, we introduce two losses: the dark channel loss \mathcal{L}_{DC} to curb the residual haze and total variation loss \mathcal{L}_{TV} to prevent the noise generation. They can be formulated as follows.

$$\mathcal{L}_{DC} = \frac{1}{M} \sum_{i=1}^M \|DC(K_C^{R'}(i))\|_1, \quad (8)$$

$$\mathcal{L}_{TV} = \frac{1}{M} \sum_{i=1}^M \|\nabla_x K_C^{R'}(i)\|_1 + \|\nabla_y K_C^{R'}(i)\|_1, \quad (9)$$

where ∇_x and ∇_y denote the gradient operations along the horizontal and vertical directions, respectively.

3.3 Re-identification Network

The re-identification network (ReID-Net) aims to extract discriminative features to search the images with the same identification in the gallery. The details of architecture and training are illustrated as follows.

Architecture. The ReID-Net consists two encoders (i.e., \mathbf{E}_H and \mathbf{E}_C) and one decoder (i.e., \mathbf{D}_{ReID}). It adopts ResNet-50 [16] as the backbone, where we apply the first two convolution blocks as the architecture of two encoders. As shown in Fig. 2, the extracted features F_H or F_C are fed to the decoder \mathbf{D}_{ReID} to generate the ReID results where the decoder consists of the rest of convolution blocks in ResNet-50 and extracted features are down-scaled by global average pooling (GAP) and batch normalization (BN) to generate 2048-d embedding features F_{ReID} . Last, we adopt the fully connected layer (FC layer) to match the number of identities for the classification. For the supervised learning of synthetic data, since we have the corresponding ID label, we adopt the triplet loss \mathcal{L}_{Tri} and the ID loss \mathcal{L}_{ID} . For the unsupervised learning stage, due to lack of the ID label, the embedding consistency loss \mathcal{L}_{EC} is adopted to constrain the network. This architecture enables our two encoders to learn domain adaptive features because the features extracted by two encoders working on different domains are fed into the same decoder.

Supervised Training Stage. At this stage, we train the re-identification network (i.e., \mathbf{E}_C , \mathbf{E}_H , and \mathbf{D}_{ReID}) by adopting the triplet loss [19] \mathcal{L}_{Tri} and the ID loss \mathcal{L}_{ID} which can be defined as follows.

$$\mathcal{L}_{Tri} = \frac{1}{M} \sum_{i=1}^M \sum_k \left[\max_{z_p \in \mathcal{P}(z_i^k)} D(z_i^k, z_p) - \min_{z_n \in \mathcal{N}(z_i^k)} D(z_i^k, z_n) + \delta \right]_+ \quad (10)$$

$$\mathcal{L}_{ID} = -\frac{1}{M} \sum_{i=1}^M \sum_k \log \frac{\exp(\sigma_i^{y_i^k})}{\sum_{j=1}^C \exp(\sigma_i^j)} \quad (11)$$

where $k \in \{(K_H^S)_{GT}, K_H^S\}$. $\mathcal{P}(z_i^k)$ and $\mathcal{N}(z_i^k)$ denote the positive and negative sample sets, respectively. z_i^k represents the extracted embedding features from the i^{th} input sample (i.e., $((K_H^S)_{GT}(i) \text{ or } K_H^S(i))$). δ is the margin of the triplet loss, $D(\cdot, \cdot)$ is the Euclidean distance, and $[\cdot]_+$ equals to $\max(\cdot, 0)$. For \mathcal{L}_{ID} , σ_i^j is the output of the FC layer with the class j based on i^{th} input image. C presents the total number of the class, and y_i donates the ground truth class. The ReID loss \mathcal{L}_{ReID_s} at this stage is the combination of \mathcal{L}_{ID} and \mathcal{L}_{Tri} .

Unsupervised Training Stage. At this stage, we feed both real clear data and real hazy data separately. Due to the lack of labels about ID information, to train the ReID network (i.e., \mathbf{E}_C , \mathbf{E}_H , and \mathbf{D}_{ReID}) with real clear inputs, we develop *embedding consistency loss* (\mathcal{L}_{EC}) to calculate the distance of two embedding features extracted from the input clear image and the rendered haze image. Initially, given a clear image K_C^R , it is fed into the DT-Net to render the hazy image $K_H^{R'}$ and ReID-Net to extract embedding feature $(F_{\text{ReID}})_C^R$. Then, we further take the rendered image to the ReID-Net to produce the embedding feature $(F_{\text{ReID}})_H^{R'}$. We can calculate the loss between $(F_{\text{ReID}})_H^{R'}$

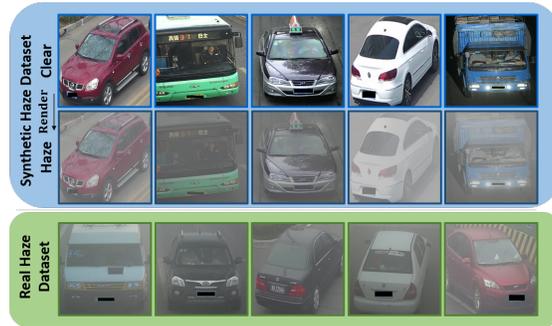


Fig. 3: Examples of the images in the synthetic dataset and the real-world datasets for vehicle ReID.

and $(F_{ReID})_C^R$ because they are the same vehicle. By using this loss, the haze-invariant features can be learned by the ReID-Net effectively. The mathematical expression of this loss is defined as follows.

$$\mathcal{L}_{EC} = \frac{1}{M} \sum_{i=1}^M \|[F_{ReID}(i)]_C^R - [F_{ReID}(i)]_H^R\|_1 \quad (12)$$

By contrast, when the input data is a real hazy image (i.e., K_H^R), the same mechanism is adopted.

4 Experiments

4.1 Dataset and Evaluation Protocols

Dataset Preparation. The proposed semi-supervised scheme is trained by both synthetic and real-world haze data. We select haze-free images from Vehicle-1M and VERI-Wild datasets. Subsequently, we apply the haze synthesis procedure proposed in [24] to synthesize these images. First, we adopt the method in [27] to estimate the depth map d . Then, we render the haze on these clear images by (1) with the predicted depth maps and set $\beta \in [0.4, 1.6]$ and $A \in [0.5, 1]$. Uniquely, each clear data generates a hazy image and all rendered images are divided into the training and the testing sets, respectively. For the real haze data, we survey all existing datasets and find that only Vehicle-1M and VERI-Wild datasets contain the cases in the hazy weather. Thus, we carefully select the vehicle images under hazy scenarios from two datasets. The selected images are split to the training and the testing sets. The details and examples of two types of data are presented in Table 1, Table 2 and Fig. 3, respectively.

Evaluation Protocols. Followed by the protocols of the evaluation proposed in [30, 12], we randomly select one hazy image for each vehicle and put it into the probe set. The remained images form the gallery set. We adopt the cumulative matching characteristic (CMC) curve and mean average precision (mAP) to evaluate the performance.

Table 1: **Detail of the synthetic dataset.** Table 2: **Detail of the real-world dataset.**
(IDs/Images) (IDs/Images)

Set	Train	Probe	Gallery
VERI-Wild	1167/19532	389/389	389/6125
Vehicle-1M	1833/23026	611/611	611/7093
Total	3000/42558	1000/1000	1000/13218

Set	Train	Probe	Gallery
VERI-Wild	156/2472	389/389	389/5985
Vehicle-1M	247/2579	611/611	611/6242
Total	403/5051	1000/1000	1000/12227

4.2 Implementation Details

Training Stage.³ For the proposed ReID network, ResNet-50 is adopted as the backbone, whose weights are initialized from the model pre-trained on the ImageNet. In the synthetic data training stage, the dimensions of FC layers are set to 3000. The weights of the domain transformation network are initialized by Kaiming normalization [15]. The whole network is trained in an end-to-end fashion based on the training sets of synthetic and real-world datasets for learning domain transformation, vehicle ReID and ID classification simultaneously. The input image is resized to 384×384 . The training batch sizes at the synthetic data and the real-world data stages are 72 ($2M$) and 36 (M), respectively. The local patch size in the dark channel operation is 5×5 . We apply the data augmentation in the training process including the random cropping and horizontal flipping techniques. The warm-up training strategy is adopted for 120 epochs. The Adam optimizer is adopted with a decay rate of 0.6. The initial learning rate is 1.09×10^{-5} , which increases to 10^{-4} after the 10^{th} epoch. At the training stage of the synthetic data, we adopt the synthetic dataset in Table 1 and randomly select one hazy image for each vehicle and put it into the probe set. The rest of images form the gallery set. For the training stage of real-world data, we apply 5051 clear images and hazy images without ID labels, respectively. The network is trained on an Nvidia Tesla V100 GPU for 3 days and we implement it using Pytorch.

Inference Stage. At the inference stage, the encoder (\mathbf{E}_C) and two decoders of the domain transformation network (\mathbf{D}_C and \mathbf{D}_H) are not involved. The computational burden caused by them can be ignored. The Euclidean distance D is computed through embedding features to evaluate the performance.

4.3 Comparison with the Existing Methods

To evaluate the performance of the proposed method, we compare the proposed algorithm with state-of-the-art ReID methods, the VRCF [9], the VOC [45], the DMT [17], the CAL [33], the VEHICLEX [38], the TransReID [18], the PVEN [31], and the HRCN [41]. For a fair and comprehensive comparison, these methods are retrained by the following training sets: (i) The ground truth clear images in the synthetic dataset; (ii) The hazy images from the training sets of both synthetic and real-world haze datasets (denoted with the ‘-haze’); (iii) The two-stage strategy (i.e., dehazing+ReID) which is denoted with ‘-dehaze’. Specifically, this strategy is the combination of the dehazing method for pre-processing and the ReID models trained by setting (i). For the dehazing

³ More details about training each stage and results are presented in the Supplementary Material.

Table 3: **Quantitative evaluation on the hazy vehicle ReID scenario.** The words with **boldface** indicate the best results, and those with underline indicate the second-best results. The texts 'S' and 'R' indicate synthetic and real-world datasets.

Method	mAP		CMC@1		CMC@5		CMC@10	
	S	R	S	R	S	R	S	R
VRCF	25.90	36.60	61.70	63.70	76.50	78.80	81.30	83.20
VRCF-dehaze	61.50	50.80	85.40	78.00	95.10	92.00	97.20	95.40
VRCF-haze	69.00	58.00	88.60	81.10	97.60	93.80	98.40	96.80
VRCF-all	73.00	64.40	90.90	85.50	97.30	95.70	98.60	97.80
VOC	59.70	57.40	86.10	82.80	94.30	94.00	95.60	96.60
VOC-dehaze	63.40	49.20	87.00	74.10	94.80	89.90	96.50	94.30
VOC-haze	67.10	59.90	88.70	83.50	95.10	94.00	96.50	97.20
VOC-all	84.20	78.70	93.60	91.00	97.60	96.30	98.30	98.30
DMT	73.90	71.70	93.40	93.20	97.20	97.40	97.90	98.50
DMT-dehaze	75.10	71.60	93.40	92.40	96.90	97.50	98.30	98.40
DMT-haze	77.30	73.40	94.00	93.40	97.60	97.60	98.60	98.80
DMT-all	82.50	80.90	98.30	96.10	98.20	98.20	98.80	99.00
VehicleX	63.64	61.56	86.50	83.20	95.00	95.20	97.40	97.90
VehicleX-dehaze	73.06	64.82	89.70	83.90	96.70	95.10	98.20	97.60
VehicleX-haze	77.86	69.01	91.20	84.80	97.10	96.10	98.70	98.10
VehicleX-all	80.75	76.39	93.10	89.90	97.60	96.90	98.60	98.40
TransReID	62.90	64.00	82.40	77.70	92.30	88.80	98.40	94.00
TransReID-dehaze	66.80	65.30	83.00	76.60	94.10	89.90	98.10	94.60
TransReID-haze	73.90	72.10	84.80	82.60	95.20	90.70	98.70	95.60
TransReID-all	79.20	76.90	89.40	84.50	96.80	93.20	98.90	97.30
PVEN	72.83	75.36	63.73	66.48	84.39	86.53	89.65	91.20
PVEN-dehaze	81.70	78.13	73.29	69.47	92.50	89.16	96.04	93.43
PVEN-haze	84.55	81.92	76.60	74.09	95.02	92.15	97.84	95.66
PVEN-all	88.63	84.08	83.55	78.31	98.45	95.40	99.20	97.76
HRCN	81.22	71.77	92.00	85.30	97.60	95.40	99.10	97.50
HRCN-dehaze	83.44	72.78	92.20	84.60	98.00	96.10	99.00	97.80
HRCN-haze	85.40	78.64	92.80	89.40	98.50	96.70	99.10	98.40
HRCN-all	87.91	81.41	94.60	91.80	98.20	97.30	99.30	99.00
CAL	75.52	75.94	92.50	91.70	96.50	97.60	97.90	98.40
CAL-dehaze	83.21	77.49	94.80	94.00	98.30	98.00	98.90	98.80
CAL-haze	86.00	80.31	95.00	94.20	97.90	98.30	98.90	99.10
CAL-all	88.20	83.84	96.30	96.00	98.40	98.20	98.90	99.00
Ours	88.66	84.12	96.70	<u>95.60</u>	98.60	98.60	99.30	99.30
Ours-F	89.14	87.72	96.50	96.90	98.60	98.40	99.40	99.60

Table 4: **Ablation study for each module in the real-world test set.**

Method	Metric			
	mAP	CMC@1	CMC@5	CMC@10
Baseline-haze	76.17	93.40	97.50	98.50
Baseline-all	77.34	93.40	97.60	98.60
Ours w/o \mathcal{L}_{CR} & \mathcal{L}_{MIDC}	76.02	93.50	97.20	98.60
Ours w/o \mathcal{L}_{MIDC}	81.19	94.20	98.20	99.00
Ours w/o \mathcal{L}_{CR}	82.27	95.30	98.00	99.20
Ours w/o \mathcal{L}_{DC} & \mathcal{L}_{TV}	77.31	93.63	97.38	98.80
Ours w/o \mathcal{L}_{DC}	81.20	94.30	98.30	98.85
Ours w/o \mathcal{L}_{TV}	82.50	94.60	98.50	98.90
Ours	84.12	95.60	98.60	99.30

Table 5: **Comparison of performance for using different blocks as encoders E_C and E_H in the real-world test set.**

	mAP	CMC@1	CMC@5	CMC@10
Conv_2	84.12	95.60	98.60	99.30
Conv_3	83.84	94.90	98.20	98.90
Conv_4	82.56	94.40	98.20	99.00
Conv_5	80.82	94.90	98.10	99.00

Table 6: **Ablation study for using different training data stages in real-world test set.**

Stage	Metric			
	mAP	CMC@1	CMC@5	CMC@10
Syn	78.17	92.20	96.70	97.20
Syn+RC	80.03	94.10	98.20	98.90
Syn+RH	81.26	94.60	98.10	98.90
Ours	84.12	95.60	98.60	99.30

method, we adopt one of the state-of-the-art dehazing methods called MPR-Net [39] which was retrained on hazy vehicle images. (iv) The same training images used in our method including synthetic haze, real-world clear and real-world haze datasets (denoted with '-all'). The aforementioned settings are all with complete ID labels.

The results are reported in Table 3. We can observe the following results. First, compared with other strategies, the proposed method can achieve the competitive performance on vehicle ReID in hazy weather on both synthetic and real-world datasets in terms of mAP and CMC. Second, existing methods trained on all data can obtain better performance compared to the methods trained on other training settings. Third, other methods may have limited performance in real-world scenarios, especially when they are only trained on synthetic images. Last, *surprisingly, though our method is trained without ID labels in real-world data, it can outperform most supervised methods trained with complete ID labels.*

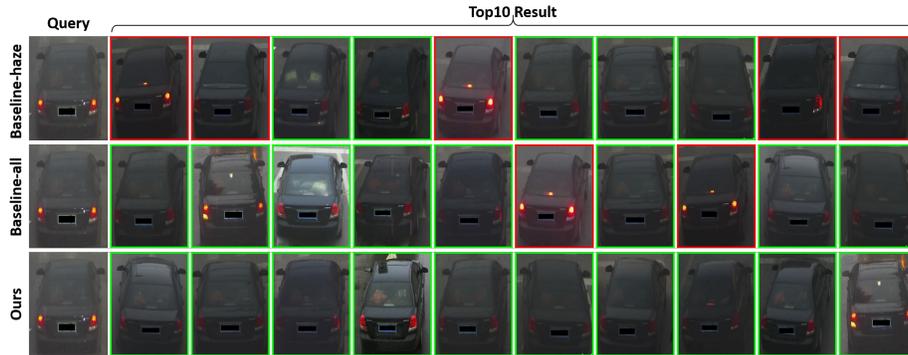


Fig. 4: **Visual results of the ranking list on the real-world dataset.** The query images are in the first column and the retrieved top-10 ranking results are in the rest columns. We denote the correct retrieved images with a green border while the false instances are with a red border.



Fig. 5: **Visual comparison of using the unsupervised losses \mathcal{L}_{CR} and \mathcal{L}_{MIDC} in the domain transformation network.** These loss functions can benefit the rendering process to generate more desirable results.

We also adopt our method trained with ID labels in the real-world data training stage which is denoted with the suffix 'F'. Specifically, we introduce the triplet loss and the ID loss defined in (11) and (10) to train the real-world data stage. The result indicates that the performance can be improved if we use complete ID labels in the training stage. Our method can be also adopted in the fully supervised scenarios and obtain the decent performance.

4.4 Ablation Studies

Effectiveness of the Semi-supervised Strategy. In this paper, we proposed the semi-supervised training technique to solve hazy vehicle ReID problem. It uses the domain transformation mechanism which enables us to train the real-world data without the ID labels. We present the effectiveness of this strategy in Table 4. We adopt the our ReID network as the baseline and train with two settings for comparison, that -is, the settings (ii) and (iv) reported in subsection 4.3 with complete ID labels. One can see that our method is against the first setting favorably. Moreover, even without the ID

labels of real-world data, our method can outperform the baseline trained with complete ID labels since our methods integrate the domain transformation technique, which can improve the ReID network to learn better representation under the hazy scenes. We also show the visual results of the ranking list on the real-world dataset in Fig. 4. One can see that, the baseline retrieves wrong instances because the important features such as the window and light become ambiguous due to the degradation of haze, which may deteriorate the performance of ReID.

Moreover, in Table 5, we show the results of assigning different convolution blocks as the encoders. One can see that adopting the first two convolution blocks as the encoders can obtain the best performance. However, the proposed domain transformation architecture can assist the network to learn accurate ReID in the haze scenario.

Effectiveness of the Loss Functions. In Table 4, we verify the effectiveness of the adopted loss functions: the monotonously increasing dark channel loss \mathcal{L}_{MICD} and the colinear relation constraint \mathcal{L}_{CR} . One can see that, with two loss functions, the performance of ReID can be improved in both mAP and CMC metrics since the DT-Net can benefit the encoders to learn more robust features with appropriate constraints which can further benefit the performance of ReID. Furthermore, using both the dark channel loss \mathcal{L}_{DC} and the total variation loss \mathcal{L}_{TV} can improve the performance of the network. Fig. 5 presents that, with the proposed loss functions, the rendered results can be more realistic compared with other modules. The rendered results may have the color distortion problems without using \mathcal{L}_{CR} .

Effectiveness of Each Training Stage. We verify the effectiveness of using real clear data or real hazy data in the training process. We construct three settings for the comparison. Specifically, we adopt: (i) only the synthetic data stage (**Syn**), (ii) **Syn** with real clear data stage (**Syn+RC**), and (iii) **Syn** with real haze data stage (**Syn+RH**). The results are reported in Table 6. We can see that only adopting the synthetic data may cause limited performance in real-world scenarios due to the domain gap problem.

5 Conclusion

In this paper, to address the vehicle ReID problem under hazy scenarios, a semi-supervised training framework that integrates the domain transformation network and the ReID network is proposed. Moreover, to constrain the unsupervised training stage, several loss functions to bound the two networks are proposed. With these techniques, the proposed method can learn haze-invariant features for robust vehicle ReID. Experimental results show that, compared to existing methods trained on complete ID labels, the proposed methods can achieve decent performance even without using the ID labels in real-world data.

6 Acknowledgement

We thank to National Center for High-performance Computing (NCHC) for providing computational and storage resources. This research was supported by the Ministry of Science and Technology, Taiwan under Grants MOST 108-2221-E-002-072-MY3, MOST 108-2638-E-002-002-MY2, and MOST 111-2221-E-002-136-MY3.

References

1. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016) [5](#), [8](#)
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* **25**(11), 5187–5198 (2016) [2](#)
3. Chen, W.T., Chen, I.H., Yeh, C.Y., Yang, H.H., Ding, J.J., Kuo, S.Y.: Sjd1-vehicle: Semi-supervised joint defogging learning for foggy vehicle re-identification (2022) [2](#)
4. Chen, W.T., Ding, J.J., Kuo, S.Y.: Pms-net: Robust haze removal based on patch map for single images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11681–11689 (2019) [2](#)
5. Chen, W.T., Fang, H.Y., Hsieh, C.L., Tsai, C.C., Chen, I., Ding, J.J., Kuo, S.Y., et al.: All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4196–4205 (2021) [2](#)
6. Chen, W.T., Huang, Z.K., Tsai, C.C., Yang, H.H., Ding, J.J., Kuo, S.Y.: Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17653–17662 (2022) [5](#)
7. Chen, W.T., Lou, H.L., Fang, H.Y., Chen, I.S., Chen, Y.W., Ding, J.J., Kuo, S.Y.: Desmokenet: A two-stage smoke removal pipeline based on self-attentive feature consensus and multi-level contrastive regularization. In: *IEEE Transactions on Circuits and Systems for Video Technology* (2021) [6](#)
8. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2157–2167 (2020) [5](#)
9. Gao, C., Hu, Y., Zhang, Y., Yao, R., Zhou, Y., Zhao, J.: Vehicle re-identification based on complementary features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 590–591 (2020) [4](#), [11](#)
10. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014) [8](#)
11. Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J.: A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering* (2021) [8](#)
12. Guo, H., Zhao, C., Liu, Z., Wang, J., Lu, H.: Learning coarse-to-fine structured feature embedding for vehicle re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018) [4](#), [10](#)
13. He, B., Li, J., Zhao, Y., Tian, Y.: Part-regularized near-duplicate vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3997–4005 (2019) [2](#), [4](#)
14. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* **33**(12), 2341–2353 (2010) [5](#), [7](#), [8](#)
15. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015) [11](#)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2016) [9](#)
17. He, S., Luo, H., Chen, W., Zhang, M., Zhang, Y., Wang, F., Li, H., Jiang, W.: Multi-domain learning and identity mining for vehicle re-identification. In: Proceedings of the IEEE/CVF

- Conference on Computer Vision and Pattern Recognition Workshops. pp. 582–583 (2020) [4](#), [11](#)
18. He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W.: Transreid: Transformer-based object re-identification. arXiv preprint arXiv:2102.04378 (2021) [2](#), [4](#), [11](#)
 19. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017) [4](#), [9](#)
 20. Hui, Z., Li, J., Wang, X., Gao, X.: Image fine-grained inpainting. arXiv preprint arXiv:2002.02609 (2020) [6](#)
 21. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017) [3](#)
 22. Khorramshahi, P., Peri, N., Chen, J.c., Chellappa, R.: The devil is in the details: Self-supervised attention for vehicle re-identification. In: European Conference on Computer Vision. pp. 369–386. Springer (2020) [4](#)
 23. Koschmieder, H.: Theorie der horizontalen sichtweite. Beitrage zur Physik der freien Atmosphere pp. 33–53 (1924) [2](#), [5](#)
 24. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2018) [10](#)
 25. Li, L., Dong, Y., Ren, W., Pan, J., Gao, C., Sang, N., Yang, M.H.: Semi-supervised image dehazing. IEEE Transactions on Image Processing **29**, 2766–2779 (2019) [3](#), [6](#)
 26. Li, M., Huang, X., Zhang, Z.: Self-supervised geometric features discovery via interpretable attention for vehicle re-identification and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 194–204 (2021) [4](#)
 27. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE transactions on pattern analysis and machine intelligence **38**(10), 2024–2039 (2015) [10](#)
 28. Liu, H., Tian, Y., Wang, Y., Pang, L., Huang, T.: Deep relative distance learning: Tell the difference between similar vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2167–2175 (2016) [4](#)
 29. Liu, X., Liu, W., Mei, T., Ma, H.: Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. IEEE Transactions on Multimedia **20**(3), 645–658 (2017) [4](#)
 30. Lou, Y., Bai, Y., Liu, J., Wang, S., Duan, L.: Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3235–3243 (2019) [4](#), [10](#)
 31. Meng, D., Li, L., Liu, X., Li, Y., Yang, S., Zha, Z.J., Gao, X., Wang, S., Huang, Q.: Parsing-based view-aware embedding network for vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7103–7112 (2020) [2](#), [4](#), [11](#)
 32. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019) [5](#)
 33. Rao, Y., Chen, G., Lu, J., Zhou, J.: Counterfactual attention learning for fine-grained visual categorization and re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1025–1034 (2021) [2](#), [4](#), [11](#)
 34. Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X.: Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1900–1909 (2017) [4](#)
 35. Wang, Z., Tang, L., Liu, X., Yao, Z., Yi, S., Shao, J., Yan, J., Wang, S., Li, H., Wang, X.: Orientation invariant feature embedding and spatial temporal regularization for vehicle re-

- identification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 379–387 (2017) [4](#)
36. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10551–10560 (2021) [5](#)
 37. Yan, W., Sharma, A., Tan, R.T.: Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13259–13268 (2020) [3](#), [8](#)
 38. Yao, Y., Zheng, L., Yang, X., Naphade, M., Gedeon, T.: Simulating content consistent vehicle datasets with attribute descent. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. pp. 775–791. Springer (2020) [4](#), [11](#)
 39. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021) [2](#), [5](#), [12](#)
 40. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018) [2](#)
 41. Zhao, J., Zhao, Y., Li, J., Yan, K., Tian, Y.: Heterogeneous relational complement for vehicle re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 205–214 (2021) [2](#), [4](#), [11](#)
 42. Zheng, A., Lin, X., Li, C., He, R., Tang, J.: Attributes guided feature learning for vehicle re-identification. arXiv preprint arXiv:1905.08997 (2019) [4](#)
 43. Zhou, Y., Shao, L.: Aware attentive multi-view inference for vehicle re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6489–6498 (2018) [4](#)
 44. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing **24**(11), 3522–3533 (2015) [5](#)
 45. Zhu, X., Luo, Z., Fu, P., Ji, X.: Voc-reid: Vehicle re-identification based on vehicle-orientation-camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 602–603 (2020) [11](#)