Information Fusion 100 (2023) 101901

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Cross-directional consistency network with adaptive layer normalization for multi-spectral vehicle re-identification and a high-quality benchmark

Aihua Zheng [a], Xianpeng Zhu [b,1], Zhiqi Ma [b,1], Chenglong Li [a,*], Jin Tang [b], Jixin Ma [c]

[a] *Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Artificial Intelligence, Anhui University, Hefei, 230601, China*

[b] *Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University, Hefei, 230601, China*

[c] *School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK*

ARTICLE INFO

Keywords:
Vehicle Re-ID
Multi-spectral representation
Cross-directional center consistency
Layer normalization
Benchmark dataset

ABSTRACT

To tackle the challenge of vehicle re-identification (Re-ID) in complex lighting environments and diverse scenes, multi-spectral sources like visible and infrared information are taken into consideration due to their excellent complementary advantages. However, multi-spectral vehicle Re-ID suffers cross-modality discrepancy caused by heterogeneous properties of different modalities as well as a big challenge of the diverse appearance with different views in each identity. Meanwhile, diverse environmental interference leads to heavy sample distributional discrepancy in each modality. In this work, we propose a novel cross-directional consistency network (CCNet) to simultaneously overcome the discrepancies from both modality and sample aspects. In particular, we design a new cross-directional center loss (L_{cdc}) to pull the modality centers of each identity close to mitigate cross-modality discrepancy, while the sample centers of each identity close to alleviate the sample discrepancy. Such a strategy can generate discriminative multi-spectral feature representations for vehicle Re-ID. In addition, we design an adaptive layer normalization unit (ALNU) to dynamically adjust individual feature distribution to handle distributional discrepancy of intra-modality features for robust learning. To provide a comprehensive evaluation platform, we create a high-quality RGB-NIR-TIR multi-spectral vehicle Re-ID benchmark (MSVR310), including 310 different vehicles from a broad range of viewpoints, time spans and environmental complexities. Comprehensive experiments on both created and public datasets demonstrate the effectiveness of the proposed approach comparing to the state-of-the-art methods. The dataset and code will be released for free academic usage at https://github.com/superlollipop123/Cross-directional-Center-Network-and-MSVR310.

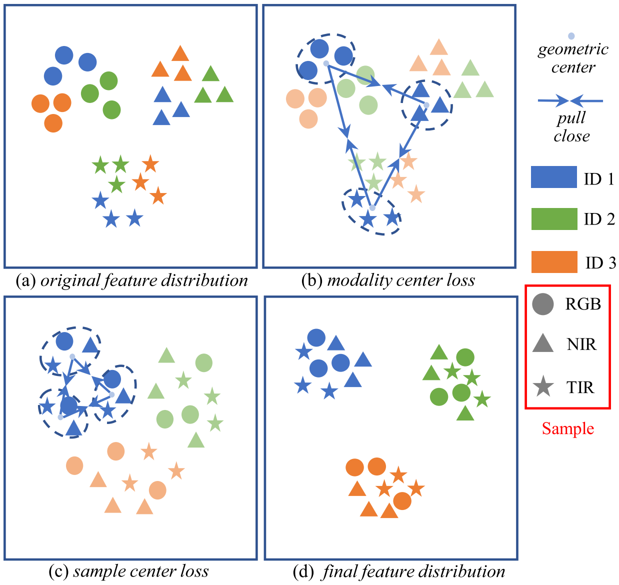
1. Introduction

Vehicle re-identification (Re-ID) aims to search the given vehicle image from the cross-camera gallery with the same identity. Due to the wide range of real-life applications such as video surveillance, smart city and intelligent transportation. Vehicle Re-ID has been attracted growing attention and experiencing rapid development with the emergence of comprehensive studies [1–4] and public large-scale datasets [5–8]. However, most existing studies only focus on visible images which suffer imaging weaknesses in complex lighting environments and extreme weather, thus cannot satisfy the demand for all-day and all-weather real-life surveillance.

Since visible (RGB), near infrared (NIR) and thermal infrared (TIR) sources have strongly complementary advantages in adverse lighting conditions and environments. RGB-NIR-TIR multi-spectral vision tasks, such as tracking [9–12], person Re-ID [13] and saliency detection [14] attract hot research interest in the both machine learning and computer vision communities. Recently, Li et al. [15] first launch the multi-spectral vehicle Re-ID task. They first propose a baseline multi-spectral vehicle Re-ID method Heterogeneity-Collaboration Aware Multi-Stream Convolutional Network (HAMNet) which utilizes multi-spectral features with class-aware weight fusion. Meanwhile, they first provide two benchmark datasets RGBN300 and RGBNT100 to multi-spectral vehicle Re-ID community. To be annotated, different from traditional vehicle Re-ID datasets which treat a single image as a sample, these two multi-spectral datasets treat an image pair (RGB-NIR in RGBN300) or an image triplet (RGB-NIR-TIR) as a sample. To avoid confusion, we use

[*] Corresponding author.
E-mail addresses: ahzheng214@foxmail.com (A. Zheng), xpzhu6325@foxmail.com (X. Zhu), doermzq0398@foxmail.com (Z. Ma), lcl1314@foxmail.com (C. Li), tangjin@ahu.edu.cn (J. Tang), j.ma@greenwich.ac.uk (J. Ma).
[1] Xianpeng Zhu and Zhiqi Ma are with equal contributions.

https://doi.org/10.1016/j.inffus.2023.101901
Received 16 January 2023; Received in revised form 6 May 2023; Accepted 25 June 2023
Available online 5 July 2023



(a) *original feature distribution* (b) *modality center loss*

(c) *sample center loss* (d) *final feature distribution*

Fig. 1. Impact demonstration of the Cross-directional Center loss $\mathcal{L}_{CdC}$ on the distribution of multi-spectral features. (a) The original distribution. (b) Impact of $\mathcal{L}_{CdC_M}$ which aims to pull modality centers of each identity close. (c) Impact of $\mathcal{L}_{CdC_S}$ which pulls sample centers of each identity close. (d) Feature distribution driven by $\mathcal{L}_{CdC}$ including both $\mathcal{L}_{CdC_M}$ and $\mathcal{L}_{CdC_S}$.

the concept *sample* in the rest of this paper to emphasize the difference from conventional single modality *image* for multi-modal Re-ID. Despite of the pioneer contribution, there are three major issues remain to be well addressed in multi-spectral vehicle Re-ID.

First, the sample discrepancy caused by the diverse imaging conditions and the modality discrepancy with the heterogeneous modality gap restrict the learning capacity of intra-class compactness. We propose a cross-directional center loss $\mathcal{L}_{CdC}$ which is composed of a sample center loss $\mathcal{L}_{CdC_S}$ and a modality center loss $\mathcal{L}_{CdC_M}$ to solve the sample and modality discrepancies from multi-modality aspect. On the one hand, it is hard to distinguish identities by only a certain spectrum data under complex environmental interference while the modality gap significantly disturbs directly utilization of different modality. To reduce the heterogeneous gap while taking the advantages of consistent information among modalities, we propose to enforce the centers of images with the same ID from different modalities in a mini-batch close by introducing a modality center loss $\mathcal{L}_{CdC_M}$, as shown in Fig. 1(b). In this way, we can intuitively enforce the modality consistency and reduce the disturbance caused by a certain modality image.

On the other hand, although sample relation has been widely concerned in RGB and cross-modality retrieval task by triplet loss [1,16], center loss [17], HC loss [18], cross-modality constrains [19,20] and metric learning [21], they are not suitable for complementary and heterogeneous multi-modality images. The heavy environmental interference caused by illumination challenge ubiquitously exists in multi-modality data. In this case, a certain image from a certain modality is possibly unreliable when it suffers from extreme environmental interference, and will easily introduce abnormal relation in pair-wise metric process. For instance, Ling et al. [21] strengthens the relational constraints between the modality center and the samples by metric learning on the intra-class, inter-class, and intra-modal and inter-modal relationships for cross-modality Re-ID. However, it relies on features from a single image when learning intra-modal relations, which is sensitive to the noise within a certain modality, similar to center loss [17], MAUM [22] and HRNet [23]. Meanwhile, it is optimized by constraining the distances between positive and negative sample pairs, resulting in the optimization of the intra-class relationship relying on the inter-class relationship. Therefore it cannot well contain the intra-class difference since the distribution of positive and negative samples in multi-modality vehicle re-identification is extremely complex.

By contrast, our cross-directional center (CdC) loss takes the intra-class consistency relationship as the goal to optimize the stronger multi-modal intra-class relationships. At the same time, CdC loss does not rely too much on single image features during optimization, which can reduce the interference of low-quality images in a certain modality during the training process. HC loss [18] solves the problem of modal differences in cross-modality Re-ID by constraining the distance between the centers of different modality features within the class. While our CdC loss constrains both the intra-class modality discrepancy and the sample discrepancy, which can better overcome the huge intra-class variance in multi-modality vehicle data.

Therefore, to learn more robust features from the complementary multi-modality images, we propose a sample center loss to pull the centers of each triplet (RGB-NIR-TIR) sample with the same identity in a mini-batch close in this paper, as shown in Fig. 1(c). By jointly optimizing sample center loss and modality center loss in a cross-directional fashion (as shown in Fig. 3) in a unified deep learning framework, it simultaneously reduces both intra-class sample discrepancy and cross-modality heterogeneity, as shown in Fig. 1.

Second, multi-modality data are usually collected in diverse and challenging environments where single modal data cannot satisfy the demand for robust recognition. In this case, the data style and quality is complex which increases the difficulty of learning relations from every single modality. Meanwhile, diverse environmental interference and large appearance gap also disturb the process of identity consistency relation learning. Therefore, due to the diverse environmental interference, features from single modality suffer from heavy distributional variation, as shown in Fig. 4. This increases the difficulty in robust feature learning for CNN and further impacts the intra-class identity consistency learning.

To reduce the disturbance of intra-modality distributional variation, we design a simple but effective module called adaptive layer normalization unit (ALNU), to normalize the individual features and adaptively adjust their distributions without breaking their inner relations. Different from existing normalization operations like BN (Batch Normalization) [24], IN (Instance Normalization) [25] and GN (Group Normalization) [26], ALNU treats each input feature as an entirety and preserves original information without changing the relation across channels in feature when adjusting the distribution. Comparing with traditional layer normalization (LN) [27] which also does not change the relations across channels, our ALNU adaptively learns the gain and bias factors according to original inputs by introducing extra convolution and pooling layers and thus is more flexible. Specifically, we integrate ALNU into all branches in our network to greatly improve the discriminative ability of multi-spectral target representations and thus further boost the performance of multi-spectral vehicle Re-ID.

Third, existing multi-spectral vehicle Re-ID datasets, RGBN300 [15] and RGBNT100 [15], are limited in diversity. To provide a more comprehensive evaluation platform in multi-spectral vehicle Re-ID, we create a high-quality image benchmark dataset named MSVR310. Compared with the RGBNT100 dataset [15], our MSVR310 has following two benefits.

Longer time span. MSVR310 is collected across a relative long time span (over 40 days). Benefiting by the long time span, data collected in MSVR310 have various environmental conditions such as various illuminations, occlusions and weather. It effectively increases the diversity of our dataset. Furthermore, we annotate the time labels of samples according to their collection sequences along time. These labels would be used in improving the experimental evaluation of multi-spectral vehicle Re-ID.

More reasonable protocol. Although most advanced methods forbid to match the samples from the same camera such as Market1501 [28],

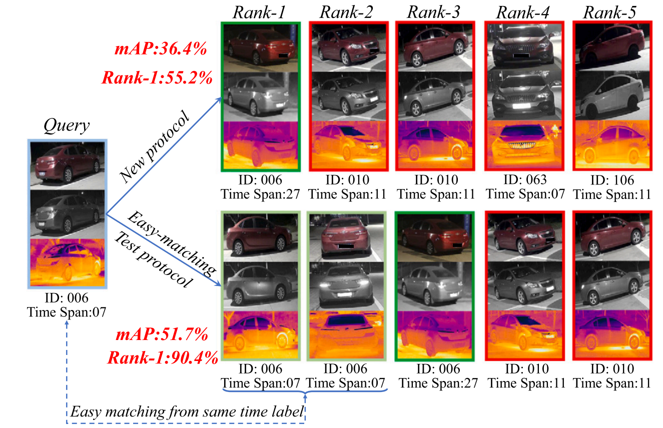


Fig. 2. Illustration of the comparison of the results between the proposed test protocol and the easy-matching protocol. Since the easy matching may easily hit the samples from the same camera or with the same viewpoint across different time spans, it results in 15.3% and 35.2% higher in mAP and Rank-1 respectively.

VeRi-776 [5], or the same viewpoint such as in RGBNT100, RGBN300 [15] to avoid the easy matching, it is not realistic enough since the same vehicle may appear in the same camera or with the same viewpoint across different time spans. Therefore, we propose to prevent the easy matching caused by similar identity-unrelated information like environments and noises by a more reasonable label, and time span, instead of the camera/viewpoint as the new protocol. Fig. 2 shows the easy matching protocol in RGBNT100 with the same time span, even though with the different viewpoints, the vehicles with the same identity and time label can be easily distinguished from others due to their high similarity on image content.

As summary, we propose a end-to-end Cross-directional Consistency Network (CCNet) to simultaneously overcome modality and sample discrepancies. And propose a new multi-spectrum vehicle Re-ID dataset MSVR310 with diverse illustration interference and rich view variation with more reasonable protocol. The contributions of this paper can be summarized as follows.

- We propose a novel cross-directional consistency network based on the cross-directional center loss to simultaneously address the problems of cross-modality discrepancy caused by heterogeneous properties of different modalities and intra-class appearance discrepancy caused by different views and adverse lighting conditions in multi-spectral vehicle Re-ID.
- We propose an adaptive layer normalization unit to dynamically adjust feature distribution within each modality. We integrate the unit into each modality branch in our network to help reducing the disturbance of intra-modality distributional variation.
- We create a high-quality benchmark dataset MSVR310, including 310 different vehicles from a broad range of viewpoints, time spans and environmental complexities. The benchmark will provide a comprehensive evaluation platform to promote the research and development of multi-spectral vehicle Re-ID.
- Comprehensive experiments on our dataset MSVR310 and the public dataset RGBNT100 validate the superior performance of our approach against several state-of-the-art multi-spectral vehicle Re-ID methods. We also conduct a random modality-missing experiment to prove the robustness of CCNet in facing the issue of missing modalities.

2. Related work

We briefly review the related works in vehicle Re-ID, cross-modality person Re-ID and multi-modality person Re-ID.

2.1. Vehicle Re-ID

In last few years, vehicle Re-ID has gained a growing attention with the rapid development of Re-ID task, which boosts the development of intelligent cities [29]. Liu et al. [5] propose a dataset called VeRi-776 with a coarse-to-fine progressive searching framework using multiple information like license plate and spatio-temporal label. Liu et al. [6] release another large-scale vehicle Re-ID dataset VehicleID and build a distance related method. Some works [30,31] introduce spatio-temporal information to provide a stricter constraint besides utilization of normal visual features. VANet [1] propose a metric loss function by treating vehicle image pairs with same or not same viewpoints differently to acquire a better distance measure. He et al. [32] design a method to enhance discriminative feature representation by introducing detection methods. Li et al. [33] propose to embed attributes and state information to enhance feature learning by reducing the intra-class feature gap. Khorramshahi et al. [34] introduce key-points information to utilize adaptive attention for vehicle Re-ID. Semantic segmentation [35] is utilized to split feature into different parts with corresponding regions in vehicles, followed by a part-aligned metric way to measure distance of image pairs more precisely. Recently, more large-scale and challenging vehicle datasets are released, like VERI-Wild [7] and CityFlow [3]. Besides real data, synthetic dataset [36] constructed via graphic engine emerges to provide arbitrary environments for learning. However, all these methods mentioned above only take a usage of single RGB modality, which is hard to satisfy the demand for all-day all weather monitoring over long period.

2.2. Cross-modality person Re-ID

To handle illumination limitations in RGB-based person [37] propose the first RGB-Infrared cross-modality benchmark SYSU-MM01 and a deep zero-padding network. RegDB [38] is also a widely used cross-modality dataset with paired visible and thermal images collected by dual camera system. Ye et al. [19] suggest a two-stream network with triplet loss to constrain the similarity in cross-modality images. An effective loss [18] is designed to supervise network learning modality invariant feature by constraining the intra-class center distance in modalities. Ye et al. [20] propose a bi-directional center-constrained loss to handle cross-modality and intra-modality variations simultaneously. Wang et al. [39] introduce a generating model to translate images to opposite modality to acquire pixel level alignment and make a feature level constraint with joint discriminator to push network produce discriminative features. Li et al. [40] introduce an auxiliary intermediate modality to reduce the gap between modalities. Lu et al. [41] propose a novel cross-modality shared-specific feature transfer algorithm to explore both modality-shared and modality-specific information. Huang et al. [42] provided a comprehensive and detailed review for cross-modality person re-id and outline the future research trends. Wei et al. [43] propose to incorporate features of heterogeneous images to generate modality-invariant representations. Ye et al. [44] propose a dynamic tri-level relation mining framework to explore intra-modality and cross-modality relations. Wei et al. [45] propose a flexible body partition model-based adversarial learning to enhance feature discriminability. Wei et al. [46] propose a reciprocal bidirectional framework for modality unification and discriminative feature learning. However, due to the lack of real aligned paired images in modalities, the heterogeneous issue in cross-modality person Re-ID still remains a key challenge.

2.3. Multi-modality person Re-ID

Similar to infrared images, depth images do not suffer the influence on lighting variation and can reflect shape and distance information of targets. Barbosa et al. [47] first propose RGB-D person Re-ID with a corresponding dataset named PAVIS. Møgelmose et al. [48] combined

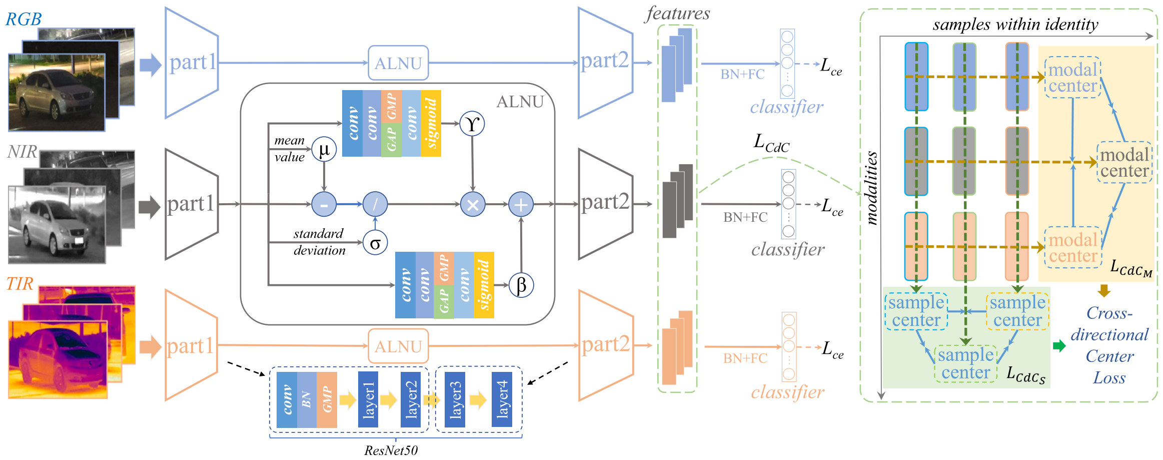


Fig. 3. Framework of the proposed Cross-directional Consistency Network (CCNet). A multi-stream network is designed to handle RGB, NIR and TIR data separately at the body part with an adaptive layer normalization unit (ALNU) embedded at the middle for each branch. Each branch is an independent ResNet50 which is split into two parts at the position between its layer2 and layer3. Then CdC loss is utilized to mine the potential intra-class relation in sample and modality level.

three different information including RGB, depth and thermal data in a joint classifier, which is the first time to utilize RGB, depth and thermal sources in person Re-ID. Munaro et al. [49] collect a RGB-D dataset named BIWI with 50 identities and multiple data sources. Wu et al. [50] utilize depth data to provide more invariant body shape and skeleton information to overcome change of illumination and color. A new cross-modality distillation network [51] has been proposed to transfer supervision between modalities like similar structural features and make a discriminative mapping to a common feature space. However, depth information is difficult to be utilized in outdoor open environments which seriously limits its application in this task.

To provide a robust solution for overcoming environmental interference, Li et al. [15] first launch multi-spectral vehicle Re-ID datasets RGBN300 (visible and near infrared) and RGBNT100 (visible, near infrared and thermal infrared), and propose a baseline method HAMNet [15] to effectively learn better feature representation by class-aware weight fusing and consistency prediction constraining. However, HAMNet [15] mainly focuses on learning multi-modality feature relations and ignores the discrepancy in both sample and modality levels. Our CCNet mainly focuses on mitigating the widely existed discrepancies from both modality and sample aspects by introducing cross-directional center loss. Zheng et al. [13] release a new multi-spectral person Re-ID dataset RGBNT201, and a progressive fusion network for multi-modality fusion. Chen et al. [52] designed a model to inherit the advantages of CNN and Transformer for multimodal matching. Although these two works first launch RGB-NI-TI multi-spectral Re-ID task and provide two benchmark datasets and baseline methods for vehicle and person Re-ID respectively, how to effectively fuse the complementary but heterogeneous information is still a big challenge.

3. Cross-directional consistency network

To utilize the consistency and mitigate the discrepancy in multi-spectral data, we propose a robust method with cross-directional center loss and adaptive layer normalization unit for multi-spectral vehicle Re-ID, referred as Cross-directional Consistency Network (CCNet) in this paper.

As shown in Fig. 3, CCNet is a multi-branch structure with three equivalent branches aiming to extract specific features for each single spectral data. Given a sample with multiple modalities, we send the image from each spectrum into corresponding branch without sharing the parameters. In each branch, an individual ALNU (adaptive layer normalization unit) module is integrated at the middle to modify feature distribution. For input images in training mini-batches, their features are divided into different groups according to the identity. Then cross-directional center loss is introduced to mitigate the intra-class appearance discrepancy and cross-modality discrepancy simultaneously for multi-spectral vehicle Re-ID. Each branch makes a prediction supervised by the cross entropy loss to learn the identity related features.

In this work, we use $D = \{I_i \mid 1 \leq i \leq N\}$ donating the whole dataset where N is the identity size. $I_i = \{S_{i,n} \mid 1 \leq n \leq N_i\}$ donates the sample set belonging to the ith vehicle where N_i is the sample number of the vehicle I_i. $S_{i,n} = \{x^m_{i,n} \mid 1 \leq m \leq M\}$ donates the image set of nth sample from I_i and $x^m_{i,n}$ is the single image from the mth modality in the sample $S_{i,n}$. In this work, M is 3 and we can simply donate samples in a triplet form as $S_{i,n} = (x^1_{i,n}, x^2_{i,n}, x^3_{i,n})$ to represent images from RGB, NIR and TIR modality respectively. We use $Part^m_k$ to donate the kth part of the branch for the mth modality in CCNet. Then, the forward process for the image $x^m_{i,n}$ can be formulated as:

$$f^m_{i,n} = Part^m_2(ALNU^m(Part^m_1(x^m_{i,n}))), \tag{1}$$

where $f^m_{i,n}$ donates the correspond feature for the image $x^m_{i,n}$. And the final representation for $S_{i,n}$ is the concatenation of its corresponding feature triplet $(f^1_{i,n}, f^2_{i,n}, f^3_{i,n})$.

3.1. Adaptive layer normalization unit

ALNU aims to handle heavy feature distributional variation within a certain modality caused by sample differences and complex environmental interference. Specifically, it normalizes the individual features and adaptive adjusts their distributions without breaking their inner relations. As shown in Fig. 4, the mean value and standard deviation of intra-modality features are distributed in a wide range, even the images with the same identity from the same modality, which further influence the intra-class identity consistency learning. ALNU module tries to mitigate the disturbance caused by heavily distributional variation by normalizing each input feature and adjusting the distribution dynamically. On one hand, this operation reduces the discrepancy on distribution of intra-modality features and helps to extract more robust CNN features. On the other hand, it is hard to evaluate similarity accurately for intra-modality images with large distribution gap regardless of identity. And mitigating this discrepancy helps to improve the validity of final similarity comparing of intra-modality image pairs in multi-spectral vehicle Re-ID task.

Given an input image $x^m_{i,n}$, we acquire its middle feature before sending into ALNU as:

$$f^{i,m,n}_{mid} = Branch^m_1(x^m_{i,n}), \tag{2}$$

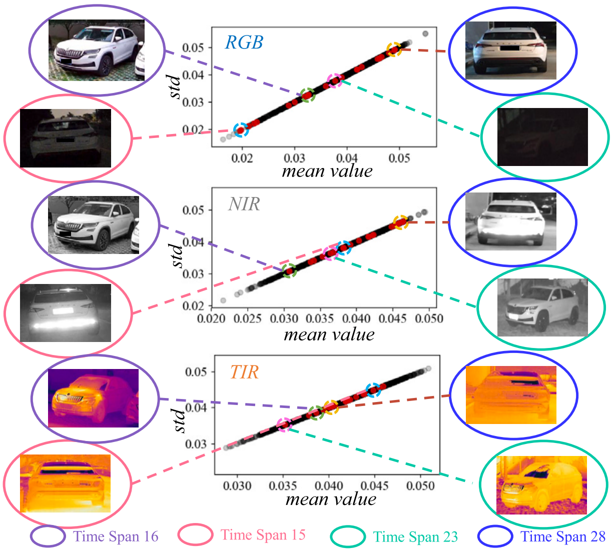


Fig. 4. The heavy feature distributional variation issue of the multi-modality images with the same ID in four different time spans. The red points together with the example images are with the same identity, which scatters with large discordance. The three images with the same color oval boxes indicate the RGB, NIR and TIR images collected in the same time span. The black points correspond to the features of the images in MSVR310 dataset.

where $f_{mid}^{i,m,n}$ is a 3-D tensor with the shape of H, W, C. We can easily obtain its mean value and standard deviation as:

$$\mu = \frac{1}{HWC}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} f_{mid}^{i,m,n}, \tag{3}$$

$$\sigma = \sqrt{\frac{1}{HWC}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C}(f_{mid}^{i,m,n} - \mu)}. \tag{4}$$

Then, we calculate a normalized feature:

$$\hat{f}_{mid} = \frac{f_{mid} - \mu}{\sqrt{\sigma^2 + \epsilon}}, \tag{5}$$

where ϵ is a small value to avoid the division over zero. BN [24] first propose to rescale and shift features in Batch Normalization, however the scale factors and shift factors are fixed during inference stage. By contrast, we propose to adaptively parameterize these two factors according to input data during both training and inference stage. Each ALNU module contains two adaptive learning blocks (ALB_γ and ALB_β), each of which is stacked by two convolutional layers, two parallel pooling layers, another convolution layer and a *Sigmoid* activation function. ALNU dynamically acquires two extra scalars by two adaptive learning blocks according to original input f_{mid} to further adjust the distribution of normalized feature $\hat{f}_{mid}$. This process can be formulated as:

$$f'_{mid} = \hat{f}_{mid} \odot \gamma + \beta, \tag{6}$$

where $\gamma = ALB_\gamma(f_{mid})$, $\beta = ALB_\beta(f_{mid})$, and f'_{mid} is the final output of ALNU.

Compared to conventional normalization operations like BN [24], IN [25], which adjust the original feature distribution in channel level, ALNU module works for individual features without breaking the relation among inner channels to avoid distinct change of the original feature distribution. Compared with LN [27] which enforces features to follow the same mean value and variance in evaluation, our ALNU learns the gain and bias factors γ and β from original input features to adaptively adjust the distribution. Different from conventional normalization operation like BN [24], LN [27], GN [26] which help models to learn easier and faster, ALNU mainly focus on intra-modality distributional variation for features, which is unrelated to their identity and increases the difficulty in robust feature learning. On one hand, ALNU adaptively modifies the distribution of features within modality and reduce the discrepancy caused by environmental interference which further mitigates the disturbance of identity related information learning. On the other hand, ALNU adaptively learns the gain and bias factors for each feature to achieve more flexible adjustment instead of enforcing all features to follow identical mean value and variance.

3.2. Cross-directional center loss

Compared with single spectral data, multi-spectral ones include more information but more challenges in vehicle Re-ID data. The challenges can be mainly summarized from two aspects, including sample discrepancy and modality discrepancy. For the sample discrepancy, a suitable representation for sample to satisfy the form of multi-modality data is necessary. Meanwhile, ubiquitous bad cases from a certain modality in multi-modality data have to be taken into consideration. For the modality discrepancy, the heterogeneous gap among modalities prevents the direct utilization for multi-modality data. We propose cross-directional center loss L_{CdC} to handle above discrepancies and mine a better identity embedding in multi-spectral vehicle Re-ID. The proposed L_{CdC} not only considers the relation between modalities like HC loss [18], but also take the relation between samples into consideration, as shown in Fig. 1.

In training process, we randomly select P identities with K samples in each mini-batch, which forms totally $M \times K \times P$ images. Then, let $F_i = \{f_{i,k}^m \mid 1 \le m \le M, 1 \le k \le K\}$ donate the final features belonging to the ith identity in a training mini-batch. The geometric sample center for the kth sample in F_i can be formulated as:

$$C_{S\,i,k} = \frac{1}{M}\sum_{m=1}^{M} f_{i,k}^m. \tag{7}$$

To overcome the sample discrepancy in multi-modality case, we propose a Sample Center Loss to pull intra-class sample centers as close as possible. This process can be formulated as:

$$\mathcal{L}_{CdC_S} = \frac{1}{2K(K-1)}\sum_{i=1}^{P}\sum_{1\le k_1<k_2\le K}\left\| C_{S\,i,k_1} - C_{S\,i,k_2}\right\|_2^2. \tag{8}$$

Similar, the geometric modality center for the mth modality in F_i can be formulated as:

$$C_{M\,i}^m = \frac{1}{K}\sum_{k=1}^{K} f_{i,k}^m. \tag{9}$$

In the same manner, to overcome the modality discrepancy, we propose a Modality Center Loss to pull intra-class modality centers as close as possible. This process can be formulated as:

$$\mathcal{L}_{CdC_M} = \frac{1}{2M(M-1)}\sum_{i=i}^{P}\sum_{1\le m_1<m_2\le M}\left\| C_{M\,i}^{m_1} - C_{M\,i}^{m_2}\right\|_2^2. \tag{10}$$

Then, the cross-directional center loss L_{CdC} is defined as:

$$\mathcal{L}_{CdC} = \mathcal{L}_{CdC_S} + \mathcal{L}_{CdC_M}. \tag{11}$$

More intuitive demonstration is shown in Fig. 3. The gradients of L_{CdC} with respect to $f_{i,k}^m$ can be solved as (since L_{CdC} only concerns intra-class relation, we simply ignore i below):

$$\begin{aligned}
&\frac{\partial L_{CdC}}{\partial f_k^m} = \frac{\partial L_{CdC_S}}{\partial f_k^m} + \frac{\partial L_{CdC_M}}{\partial f_k^m}\\
=&\frac{1}{K-1}(C_{S\,k} - \bar{C_S})\frac{\partial C_{S\,k}}{\partial f_k^m} + \frac{1}{M-1}(C_M^m - \bar{C_M})\frac{\partial C_M^m}{\partial f_k^m}\\
=&\frac{1}{M(K-1)}(C_{S\,k} - \bar{f}) + \frac{1}{K(M-1)}(C_M^m - \bar{f}),
\end{aligned} \tag{12}$$

where $\bar{C_S}$, $\bar{C_M}$, $\bar{f}$ can be formulated as:

$$\bar{C_S} = \frac{1}{K}\sum_{k=1}^{K} C_{Sk} = \frac{1}{MK}\sum_{k=1}^{K}\sum_{m=1}^{M} f_k^m = \bar{f}, \tag{13}$$

$$\bar{C_M} = \frac{1}{M}\sum_{m=1}^{M} C_M^m = \frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K} f_k^m = \bar{f}. \tag{14}$$

Thus, the final optimizing strength of L_{CdC} with respect to f_k^m is linearly dependent on its corresponding sample center C_{Sk}, modality center C_M^m and global identity center $\bar{f}$. Intra-class features within sample (modality) are in same gradient along sample (modality) direction. Besides, the gradient of L_{CdC} with respect to f_k^m is not directly related with f_k^m itself, which is not such sensitive when f_k^m corresponds to the bad cases in a certain modality.

In this work, K and P is set to 4 and 8 respectively. As shown in Eq. (12), the final factors of gradient along sample and modality directions are different ($\frac{1}{M(K-1)}$ and $\frac{1}{K(M-1)}$ respectively). Thus, we introduce a hyper-parameter α to balance their strengths. The final formulation of L_{CdC} is defined as:

$$\mathcal{L}_{CdC} = \mathcal{L}_{CdC_S} + \alpha\mathcal{L}_{CdC_M}. \tag{15}$$

Cross-directional center loss L_{CdC} focuses on optimizing intra-class relation along sample and modality directions. To enhance the ability of discriminative inter-class learning, we further introduce the cross entropy loss L_{ce}. The total loss is defined as:

$$\mathcal{L}_{total} = L_{ce} + \lambda L_{CdC}, \tag{16}$$

where the factor λ is a hyper-parameter used to balance the importance of components. In our experiments, λ and α are set to 0.3 and 0.6 respectively according to the experiments on hyper-parameter analysis, as shown in 5.9.

4. MSVR310 benchmark

In this work, we release a new dataset called MSVR310 for multi-spectral vehicle Re-ID.

4.1. Imaging platform

In MSVR310, three different spectral modalities, RGB, NIR and TIR are captured for each sample. The RGB images are captured by two devices, a 360 D866 camera for day time and a Mi8 mobile phone camera for night time. All the NIR images are captured by the 360 D866 camera, which can be switched to the near infrared mode. The TIR image capture device is FLIR SC620 which contains a thermal infrared camera with the resolution of 640 × 480.

For each sample in our dataset, it is formed as a triplet constructed by three images from RGB, NIR and TIR respectively. We manually select bounding boxes for the targets in original captured images.

4.2. Data setting and statistics

Our dataset contains 2087 samples from 310 vehicles and each sample is a triplet, which results in total 6261 images in our dataset. The number of image samples of each vehicle varies from 2 to 20. We randomly select 155 vehicles with 1032 samples as the training set, while the rest 155 vehicles with 1055 samples as the gallery set. We randomly select 52 vehicles with 591 samples from gallery set as query set. Each query identity has been captured at least twice with different time labels to support cross time matching. The data distribution is shown in Fig. 5.

We annotate data with time labels according to their collection order along time. Fig. 6 demonstrates the distribution of the captured time. Fig. 7 demonstrates some example images of four vehicles in MSVR310 along time labels. And each vehicle appears in various conditions with complex interference like strong illustration, reflection, shadow, color distortion and so on. Thus, bad cases in a certain modality exist ubiquitously and intra-class appearance discrepancy is very significant in MSVR310. The illumination disturbance in such degree is quite rare in existing works [5–7,15]. However, these disturbances represent differently in different modalities, and data across modalities are complementary in content against interference which requires for better utilization of multi-spectral data.

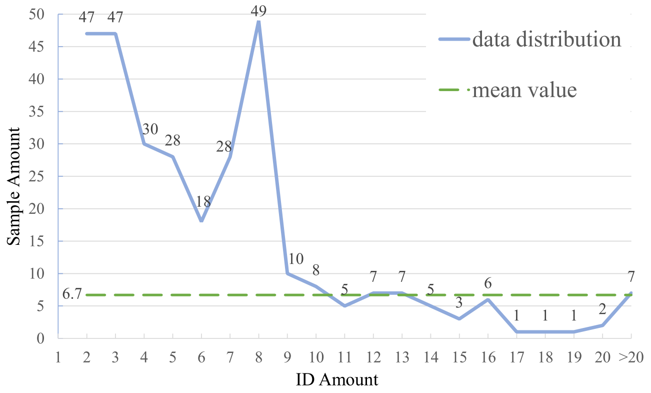


Fig. 5. Distribution for number of identities across sample sizes.

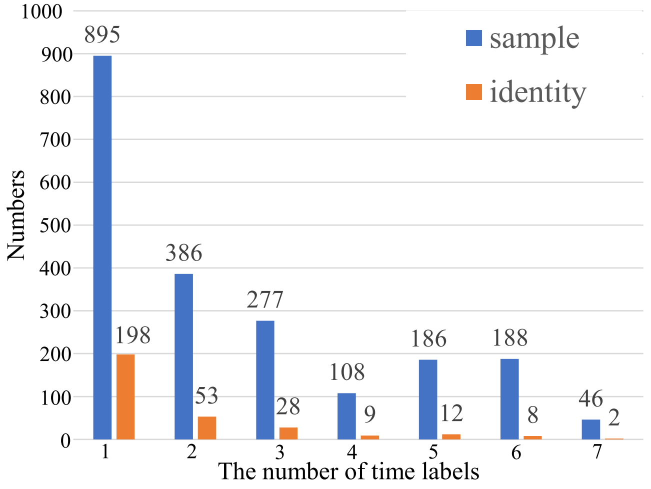


Fig. 6. Distribution of samples and identities across the number of time labels in MSVR310.

Table 1
Comparison of RGBN300, RGBNT100 and MSVR310, where ’–’ denotes ’not available’.

Benchmark	IDs	Videos	Modality	Views	Time labels
RGBN300	300	4100	R+N	8	–
RGBNT100	100	2070	R+N+T	8	–
MSVR310	310	6261	R+N+T	8	28

4.3. Difference from previous work

Li et al. [15] first propose two benchmarks multi-spectral vehicle Re-ID datasets RGBN300 and RGBNT100, as shown in Table 1. First, although RGBN300 and RGBNT100 contain much more images than MSVR310, it is actually collected from 2070 short videos (690 videos for each modality) which leads to a bunch of similar frames. We construct MSVR310 in various environments such as large changes of illuminations, occlusions and weather by capturing high-quality images instead of videos. Second, MSVR310 is collected across long time spans which leads to rich collections of various environments and vehicles. These significantly increase the diversity and difficulty of our

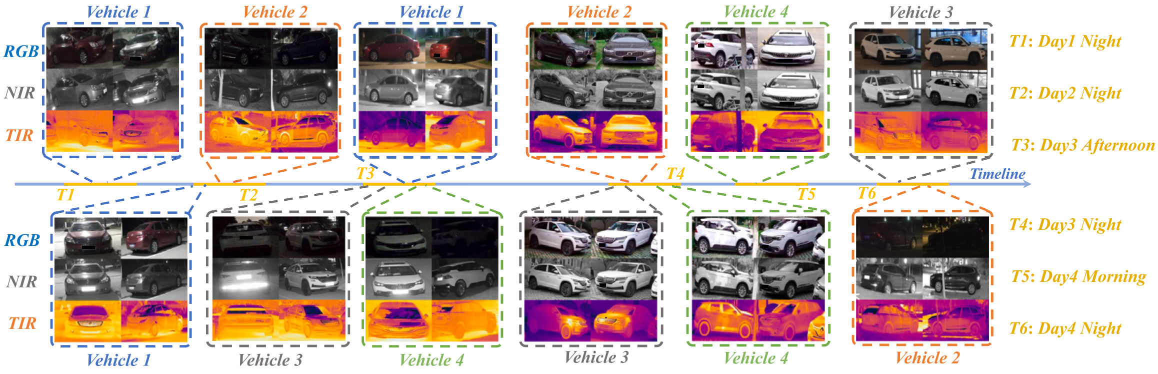


Fig. 7. Illustration of four sample data in MSVR310. Images in box with the same color indicate the multi-modality samples of the same identity with different time labels.

dataset. Third, although matching between samples in same identity and same viewpoint is not allowed in RGBN300 and RGBNT100 [15], environmental similarity among samples tends to raise easy matchings. Instead, MSVR310 introduces time labels to avoid easy matching. Matching between samples with the same identity and the same time label is forbidden in MSVR310, as shown in Fig. 2. This protocol effectively handles the easy matching problem and provides a more reliable evaluation.

5. Experiments

5.1. Datasets and evaluation metrics

To evaluate the effectiveness of the proposed CCNet on our proposed multi-spectral vehicle Re-ID dataset and public dataset, we provide comprehensive experimental results in this section. Due to there are only one public RGB-N-T image dataset RGBNT100 for the evaluation of multi-spectral vehicle Re-ID methods. We finally implement the experiments on MSVR310 and RGBNT100 following their own evaluation protocols.

To ensure the fairness of experimental evaluation, we follow the commonly used Cumulative Matching Characteristic (CMC) curve and the mean Average Precision (mAP) for evaluation. CMC score reflects the retrieval precision, where $Rank-1$, $Rank-5$, $Rank-10$ scores are reported in our experiments. mAP measure the mean of all queries of average precision (the area under the Precision Recall curve), which reflects the recall and precision comprehensively.

5.2. Implementation details

We use a strong baseline BoT [53] which is modified from ResNet50 [54] pretrained on ImageNet [55] as our backbone and the implementation platform is Pytorch 1.0.1 with one NVIDIA GTX 1080Ti GPU. We use the Adam [56] optimizer to optimize our network with the initial learning rate as 3.5×10^{-4} which will be decayed to 3.5×10^{-5} and 3.5×10^{-6} at 300-th epoch and 550-th epoch respectively of total 1200 epochs. In training process, the input images are resized to 128 × 256 and some data augmentation methods like random cropping, horizontal flipping and random erasing are used. We randomly select 8 identities which will provide 4 samples (12 images) by each one respectively as our training samples in each training mini-batch. In evaluation, we concatenate the features extracted after BNNeck [53] from three parallel branches as final representation for a sample in the absence of additional instructions.

Table 2
Experimental comparison of the effectiveness of modalities between ResNet50 and CCNet on MSVR310 (in %). In the column of test feature, R, N and T represents features from corresponding spectrum (branch) while '+' denotes feature concatenating operation.

Network	Test feature	mAP	Rank-1	Rank-5	Rank-10
ResNet50	R	20.0	29.9	49.9	61.6
	N	17.8	28.9	51.3	62.8
	T	11.9	23.2	37.4	46.4
	$R + N$	23.6	36.7	57.0	66.2
	$R + T$	22.6	35.4	54.7	63.5
	$N + T$	21.4	37.2	56.3	64.3
	$R + N + T$	25.6	39.4	58.5	67.9
CCNet	R	30.7	49.4	65.5	73.3
	N	26.3	45.5	67.3	73.1
	T	19.6	35.7	53.5	61.9
	$R + N$	34.0	53.6	70.2	76.3
	$R + T$	34.6	52.8	68.7	75.5
	$N + T$	31.4	51.6	68.9	76.6
	$R + N + T$	**36.4**	**55.2**	**72.4**	**79.7**

5.3. Evaluation on MSVR310 dataset

We first evaluate our CCNet compared with the ResNet50 on MSVR310 dataset, as reported in Table 2. For fairness, we use the same implementation of ResNet50 from BoT [53] for comparison, which is the same as the backbone of CCNet. Specifically, the results of ResNet50 are achieved by a multi-branch network constructed by three separated ResNet50 in which each branch handles data from a certain modality. The multi-modality branches are independent with no interaction with other branches, while CCNet simultaneously utilizes multi-spectral data in the training phase and thus achieves much better performance. The R, N and T in Table 2 represent the features used in test phase for distance computing from corresponding spectrum. Note that, we use all three modality data during the training phase.

From Table 2 we can see, (i) First of all, none of the single spectrum achieves satisfactory performance due to the complex lighting environments on MSVR310 dataset. In general, both RGB and NIR provide comparable reliable appearances thus lead to much better performance comparing to TIR. (ii) Two spectrum scenarios significantly improve all the metrics than the single ones while the three spectrum scenarios further boost both performances of ResNet50 and CCNet. This strongly proves the effectiveness of the introduced multi-spectral data. (iii) Our CCNet is superior to ResNet50 by a large margin while there are limited differences on network structure between CCNet and ResNet50. This strongly indicates the rightness of our discrepancy mitigating design and effectiveness of the proposed CdC loss and ALNU module.

5.4. Evaluation on different backbones and baselines

To validate the generality of our method, we integrate our CCNet into six backbones and four baselines including MobileNetV2 [57],

Table 3
Comparison to state-of-the-art Re-ID methods on MSVR310 and RGBNT100 (in %). The best three scores are highlighted in Red, Green, and Blue respectively.

Models	Reference	MSVR310				RGBNT100			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
DMML	ICCV 2019	19.1	31.1	48.7	57.2	58.5	82.0	85.1	86.2
Circle Loss	CVPR 2020	22.7	34.2	52.1	57.2	59.4	81.7	83.7	85.2
PCB	ECCV 2018	23.2	42.9	58.0	64.6	57.2	83.5	85.6	87.9
MGN	ACM MM 2018	26.2	44.3	59.0	66.8	58.1	83.1	85.6	88.0
Strong Baseline	CVPRW 2021	23.5	38.4	56.8	64.8	78.0	95.1	95.8	96.4
HRCN	ICCV 2021	23.4	44.2	66.0	73.9	67.1	91.8	93.1	93.8
OSNet	ICCV2019	28.7	44.8	66.2	73.1	75.0	95.6	97.0	97.4
AGW	T-PAMI 2021	28.9	46.9	64.3	72.3	73.1	92.7	94.3	94.9
TransReID	ICCV 2021	26.9	43.5	62.4	70.7	75.6	92.9	93.9	94.6
PFNet	AAAI 2021	23.5	37.4	57.0	67.3	68.1	94.1	95.3	96.0
HAMNet	AAAI 2020	27.1	42.3	61.6	69.5	74.5	93.3	94.3	95.2
PFD	AAAI 2022	23.0	39.9	56.3	64.0	67.5	92.6	94.3	96.5
FED	CVPR 2022	21.7	37.4	58.9	67.3	65.8	91.7	94.6	96.3
IEEE	AAAI 2022	21.0	41.0	57.7	65.0	61.3	87.8	90.2	92.1
CCNet	OURS	36.4	55.2	72.4	79.7	77.2	96.3	97.2	97.7

Table 4
Plugin our key components ALNU and CdC loss into different baselines and backbones on MSVR310.

Methods	MSVR310			
	mAP	Rank-1	Rank-5	Rank-10
SENet	22.7	40.9	60.6	69.9
+OURS	**29.5**	**47.0**	**67.7**	**73.8**
InceptionV3	23.1	43.7	59.7	68.5
+ OURS	**28.0**	**49.4**	**64.0**	**72.3**
Desenet-121	35.8	54.1	71.2	81.2
+ OURS	**38.7**	**58.5**	**76.5**	**82.2**
ResNet-101	25.2	38.9	58.5	68.2
+ OURS	**30.5**	**47.9**	**64.5**	**72.6**
ViT	30.8	**49.9**	66.2	72.1
$+L_{CdC}$	**34.4**	**49.9**	**69.4**	**78.7**
MobileNetV2	22.5	37.6	53.6	64.1
+ OURS	**24.0**	**43.5**	**59.2**	**70.1**
Strong Baseline	23.5	38.4	56.8	64.8
+ OURS	**25.6**	**47.0**	**68.9**	**74.6**
OSNet	28.7	44.8	66.2	73.1
+OURS	**30.3**	**50.3**	**67.7**	**75.3**
AGW	28.9	46.9	64.3	72.3
+OURS	**33.0**	**52.6**	**69.5**	**75.6**
TransReID	26.9	43.5	62.4	70.7
$+L_{CdC}$	**28.2**	**44.5**	**62.3**	**73.1**

SENet [58], InceptionV3 [59], Desenet-121 [60], ResNet-101 [54], ViT [61], Strong Baseline [62], OSNet [63], AGW [64] and TransReID [65] as shown in Table 4. Note that due to the conflict between convolution layer in ALNU and transformer structure, we only integrate the CdC loss into ViT [61] and TransReID [65]. Generally speaking, after integrating our ALNU and CdC loss into the different baselines and backbones, all the metrics significantly improve, which indicates the generality of our method.

5.5. Comparison with state-of-the-art methods

To validate the effectiveness of our method, we extend nine state-of-the-art single modality Re-ID methods including DMML [66], Circle loss [67], PCB [68], MGN [69], Strong Baseline [62], HRCN [70], OSNet [63], AGW [64] and TransReID [65] to multi-modality version for comparison. At last, we compare our CCNet with the multi-spectrum vehicle Re-ID method HAMNet [15] and the multi-spectrum person Re-ID method PFNet [13]. Specifically, we train the single-modality methods on multiple spectral data respectively and then concatenate the final features from modalities of the same sample as the final representation. The experimental comparison of these methods is shown in Table 3.

First, all the methods perform much worse on MSVR310 than RGBNT100 which is caused by the huge challenge of the proposed MSVR310 dataset and our evaluation protocol which filters easy matchings caused by easy positive samples with same time label. The purposed CCNet beats all the comparison methods by a large margin on MSVR310, which strongly proves the effectiveness of our method. And on RGBNT100 which is much easier with fewer challenges and limited diversity, our method also achieves very competitive performance. Transformer-based methods achieve superior performance in both person and vehicle Re-ID, due to their self-attention mechanism to simultaneously consider comprehensive local information for better global information learning. However, in multi-modality Re-ID, the key challenge is to effectively explore the complementarity while suppressing the heterogeneity among different modalities. Therefore, simply extending the transReID into the multi-modality task works overshadowed.

Second, as a first baseline multi-spectral vehicle Re-ID method, HAMNet [15] presents a simple network structure with considerable performance on three benchmark datasets, which proves its effectiveness on multi-spectral feature learning. However, HAMNet mainly focuses on learning multi-modality feature relations and ignores the discrepancy in both sample and modality levels. While our CCNet mainly focuses on mitigating heavily intra-class and intra-modality discrepancies by introducing CdC loss and ALNU. PFNet [13] is the first work for multi-spectral person Re-ID, while the local feature separation seems to be more suitable for person data than vehicle data.

5.6. Ablation study and visualization

To verify the contributions of proposed components in our model, we implement the ablation study of several variants of CCNet on MSVR310 dataset, as reported in Table 5. Note that the sample center loss L_{CdC_S}, the modality center loss L_{CdC_M} and the adaptive layer normalization unit (ALNU) all make positive improvements on our baseline, which demonstrates the contributions of the corresponding modules.

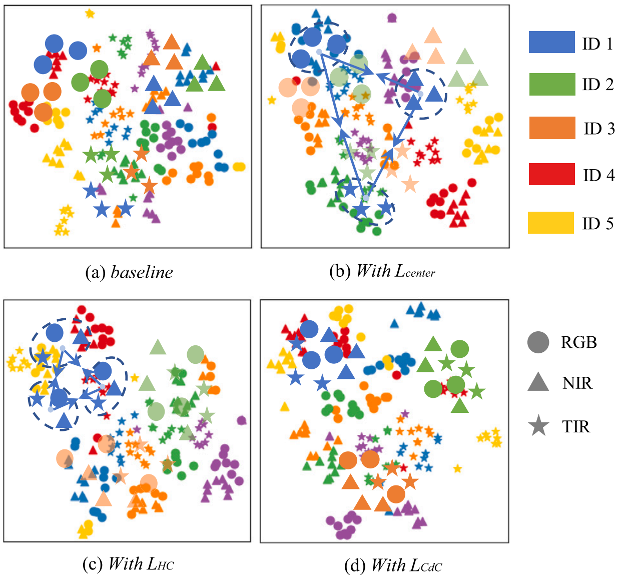
We verify the contribution of our ALNU module by comparing two conventional normalization operations, instance normalization (IN) [25] and layer normalization (LN) [27] as shown in Table 6. IN [25] is widely used in image style transfer by normalizing instance features in channel level directly. LN [27] and ALNU both treat each feature as an entirety for normalization, however LN [27] strictly enforces all features to follow the same mean value and variance while our ALNU dynamically learns the gain and bias factors which are more reasonable for complex data. We also verify the contribution of our CdC loss by comparing two widely used center-type losses, Center loss [17] and HC loss [18] as shown in Table 6. Both HC loss and Center loss are implemented based on ResNet50 with same setting as our baseline.

Table 5
Ablation study on MSVR310 (in %). Note that $L_{CdC} = L_{CdC_S} + L_{CdC_M}$.

Models	MSVR310			
	mAP	Rank-1	Rank-5	Rank-10
baseline	25.6	39.4	58.5	67.9
$+ALNU$	29.4	47.2	66.0	74.3
$+L_{CdC_S}$	27.4	41.6	61.8	69.0
$+L_{CdC_M}$	31.4	48.6	65.1	73.6
$+L_{CdC_S} + L_{CdC_M}$	33.7	51.8	68.2	76.0
$+L_{CdC} + ALNU$	**36.4**	**55.2**	**72.4**	**79.7**

Table 6
Experimental comparison with different normalizations and losses on MSVR310 (in %).

Methods	mAP	Rank-1	Rank-5	Rank-10
baseline	25.6	39.4	58.5	67.9
+ IN	26.8	42.3	61.9	70.6
+ LN	28.8	45.9	**66.3**	72.3
+ ALNU	**29.4**	**47.2**	66.0	**74.3**
$+L_{center}$	25.8	42.0	60.1	66.8
$+L_{hc}$	30.5	48.7	64.8	72.1
$+L_{CdC}$	**33.7**	**51.8**	**68.2**	**76.0**



(a) *baseline* (b) *With* L_{center}

(c) *With* L_{HC} (d) *With* L_{CdC}

Fig. 8. T-SNE [71] illustration of the feature distributions extracted by CCNet trained (a) baseline, (b) baseline with L_{Center}, (c) baseline with L_{HC} and (d) baseline with L_{CdC}.

We implement Center loss [17] to pull features within identity close regardless of modality. And HC loss [18] is implemented to reduce the modality gap within identity. However, Center loss [17] is not good at handling the ubiquitous bad cases from a certain modality while HC loss [18] ignores the discrepancy among intra-class samples in multi-modality situations. Both Center loss [17] and HC loss [18] work overshadowed by our CdC loss which simultaneously constrains intra-class relations from both modality and sample aspects. This proves the validity and robustness of our CdC loss in multi-spectral vehicle Re-ID task.

Fig. 8 demonstrates the feature distribution comparison of the network trained with different losses. When training with the baseline as shown in Fig. 8(a), features from different modalities are mixed and hard to be separated by identity labels. After introducing the center loss L_{Center}, as shown in Fig. 8(b), features with same identity in a certain modality tend to be clustered together. However, the inter-modality gap of the same identity is still very large. As shown in Fig. 8(c), introducing the HC loss L_{HC} can better eliminate the modality gap comparing with the center loss. However, some hard identities are still blended together such as the ID4, ID5 and ID6. After introducing the proposed CdC loss L_{CdC}, as shown in Fig. 8(d), features from different modalities with same identity are constrained to follow stronger consistency in both sample and modality levels.

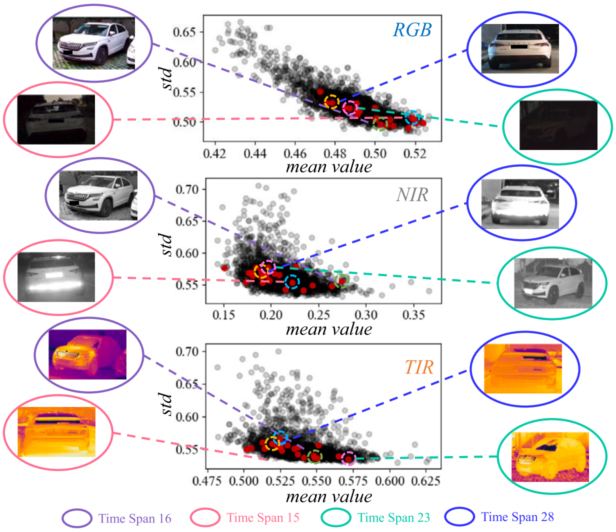


Fig. 9. Illustration of the feature distribution as shown in Fig. 4 after introducing ALNU. Comparing to Fig. 4 we can obverse the distributional discrepancy is significantly mitigated.

Fig. 9 demonstrates the distribution for multi-modality features after introducing ALNU. Compared with Fig. 4, the ALNU pushes features to distribute with similar mean values and standard deviations to reduce the distributional variation.

5.7. Comparison to cross-modality Re-ID

To better evaluate the necessity of the multi-modality vehicle Re-ID, we compare our method with three state-of-the-art cross-modality Re-ID methods including LBA [72], DDAG [73], HC Loss [18], MPANet [74] and MMN [75]. Specifically, we reconstruct data in MSVR310 into cross-modality setting followed by the data splitting protocol in RegDB [38] for the cross-modality evaluation. As shown in Table 7, due to the huge heterogeneity across modalities, all the five cross-modality Re-ID methods present inferior results. CCNet achieves the distinct superior performance by utilizing both (RGB and TI/NI) modalities, which evidences that CCNet can simultaneously utilize the complementary information among the modalities and overcome the cross-modality heterogeneity.

5.8. Evaluation on random modality missing

To verify the generality of the proposed method and dataset in diverse real scenarios, we further evaluate CCNet in handling the missing modality issue.

Specifically, we adjust the samples with a certain ratio of missing modalities in the test set for evaluation. The ratio indicates the probability of the samples with random partial (one/ two modalities in equal proportion) modality missing. For example, ratio r% indicates r% of the samples suffer from random modality missing, which means half of them ($r/2$%) randomly miss one modality while the rest half randomly missing two modalities. To overcome the sample feature misalignment caused by modality missing, we use geometric center of the existing modality/modalities as the final representation of the sample.

Table 7
Comparison of state-of-the-art cross-modality Re-ID methods on reconstructed MSVR310.

Method		RGB to TI		TI to RGB		RGB to NI		NI to RGB	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Cross-Modality	LBA	10.4	18.3	11.2	19.0	22.5	39.4	21.7	39.6
	DDAG	11.9	17.6	13.5	21.3	23.0	37.9	22.5	39.1
	HCLoss	11.3	20.0	12.4	20.0	20.0	37.9	18.9	36.9
	MPANet	12.6	17.9	11.52	15.7	21.0	35.5	20.1	37.9
	MMN	6.48	12.35	5.2	6.6	20.4	43.2	20.1	42.6
Ours (Multi-Modality)	CCNet	mAP: **18.7** Rank-1: **29.1**				mAP: **29.4** Rank-1: **43.5**			

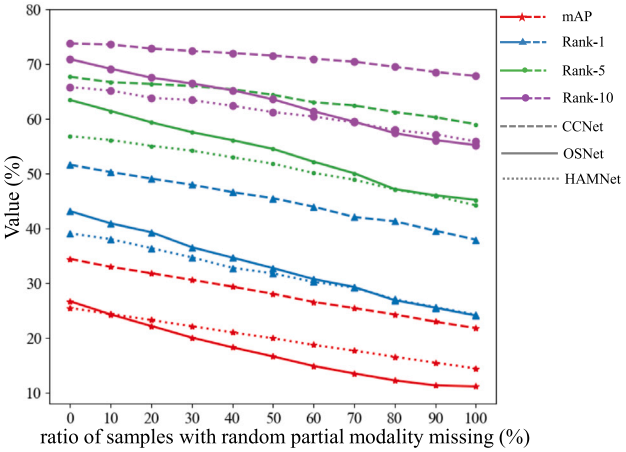


Fig. 10. Performance changing of different methods in different ratio of samples with random partial modality missing on MSVR310. When a certain ratio of samples with random partial modality missing, the probability of missing one or two modalities is equal.

In normal case without modality missing, CCNet extracts a final representation $f_{i,k}$ for sample $S_{i,k}$ (the kth sample for ith identity) where $f_{i,k}$ is a triplet of corresponding modality features. To handle the modality missing case, we generate a binary triplet mask $T_{i,k}$ for $f_{i,k}$, to indicate whether the corresponding modality is missing or not. Then, the geometric center of sample can be formulated as:

$$C'_{S_{i,k}} = \frac{1}{\sum T_{i,k}} \sum_{m=1}^{M} T_{i,k}^{m} f_{i,k}^{m}, \tag{17}$$

where $C'_{S_{i,k}}$ is the final representation of $f_{i,k}$.

We evaluate the stability of our method in handling modality missing comparing with the representative multi-modality Re-ID method HAMNet [15] and the state-of-the-art single modality Re-ID method OSNet [63]. All the experiments are evaluated based on the mean value of 10 random trials. Fig. 10 demonstrates the comparison performance against the ratio of samples with partial modality missing. Generally speaking, CCNet consistently outperforms both HAMNet [15] and OSNet [63] by a large margin. Even all the samples occur modality missing (when the ratio is 100%), CCNet still achieves competitive performance which is comparable with the results at low missing ratio of HAMNet [15] and OSNet [63]. This verifies the stability of our method in handling the modality missing. Meanwhile, all the metrics drop as the missing ratio increases, especially for mAP and $Rank-1$, which indicates the importance of complementary information of the multi-modality resources. As a state-of-the-art single modality Re-ID method, OSNet [63] drops much faster than two multi-modality methods HAMNet [15] and CCNet, which indicates the advantage of fusing multi-modality information in the two multi-modality methods in handling modality missing issue.

Table 8
Hyper-parameter analysis on MSVR310. (in %).

Hyper-parameters	MSVR310			
λ ($\alpha = 0.6$)	mA	Rank-1	Rank-5	Rank-10
0.1	31.3	47.9	66.5	72.6
0.2	33.3	50.6	67.2	73.8
0.3	**33.7**	**51.8**	**68.2**	**76.0**
0.4	33.4	50.9	67.0	74.6
0.5	33.0	50.6	67.7	74.3
0.6	32.8	50.4	66.8	73.3
0.7	32.7	50.1	66.2	73.8
0.8	32.3	49.7	66.5	72.9
0.9	31.9	49.2	65.7	72.6
1.0	31.3	48.4	65.8	72.8
α ($\lambda = 0.3$)	mAP	Rank-1	Rank-5	Rank-10
0.1	32.6	48.6	66.2	74.0
0.2	33.5	51.6	67.2	76.0
0.3	33.1	50.4	66.9	75.5
0.4	33.4	50.1	67.1	76.4
0.5	33.1	50.1	**68.9**	**77.0**
0.6	33.7	**51.8**	68.2	76.0
0.7	**33.8**	51.6	67.9	75.8
0.8	33.3	51.0	67.2	74.7
0.9	33.1	51.0	68.2	75.0
1.0	33.1	50.4	67.7	76.5

5.9. Hyper-parameter analysis

There are two hyper-parameters in our method, *e.g.*, λ in Eq. (16) which controls the importance of CdC loss in total loss and α in Eq. (15) which balances the strength of gradient along sample and modality directions in CdC loss. Large λ may affect the inter-class discrimination ability provided by L_{ce} and large α may break the balance between L_{CdC-M} and L_{CdC-S}. Therefore, we vary λ and α between 0.1 and 1.0 for the analysis. The analysis on diverse values of these two hyper-parameters is reported in Table 8. It is clear that, our method achieves the top when λ is set to 0.3 while it is not sensitive to α. We fix λ and α as 0.3 and 0.6 for the best performance in our method.

6. Conclusion

In this work, we propose a novel end-to-end trained convolutional network named CCNet for robust multi-spectral vehicle Re-ID. CCNet contains a novel cross-directional center (CdC) loss to simultaneously overcome the problems of cross-modality discrepancy and intra-class individual discrepancy. Meanwhile, a simple yet effective module named adaptive layer normalization unit is designed to embed in CCNet to mitigate the distributional variation of intra-class features for robust feature learning. Furthermore, we create a high-quality benchmark dataset MSVR310 with diverse conditions and reasonable evaluation protocol. Comprehensive experiments on our benchmark dataset MSVR310 and the public dataset RGBNT100 validate the superior performance of our CCNet and the research value of the proposed benchmark dataset.

CRediT authorship contribution statement

Aihua Zheng: Conceptualization, Methodology. **Xianpeng Zhu:** Investigation, Writing – original draft. **Zhiqi Ma:** Validation, Visualization. **Chenglong Li:** Formal analysis, Data curation. **Jin Tang:** Resources, Interpretation of data. **Jixin Ma:** Writing – review & editing.

Declaration of competing interest

All authors disclosed no relevant relationships.

Data availability

I have shared the link to the data and code in the manuscript.

Acknowledgments

This research is partly supported by the National Natural Science Foundation of China (Grant No. 61976002) and the University Synergy Innovation Program of Anhui Province (Grant Nos. GXXT-2022-036, Natural Science Foundation of Anhui Province (Grant Nos. 2208085J18) and Natural Science Foundation of Anhui Higher Education Institution (Grant Nos. 2022AH040014).

References

[1] R. Chu, Y. Sun, Y. Li, Z. Liu, C. Zhang, Y. Wei, Vehicle re-identification with viewpoint-aware metric learning, in: Proc. IEEE/CVF International Conference on Computer Vision, 2019, pp. 8281–8290.

[2] Y. Lou, Y. Bai, J. Liu, S. Wang, L. yu Duan, Embedding adversarial learning for vehicle re-identification, IEEE Trans. Image Process. 28 (2019) 3794–3807.

[3] Z. Tang, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D.C. Anastasiu, J. Hwang, CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification, in: Proc. IEEE/CVF Internaltional Conference on Computer Vision and Pattern Recognition, 2019, pp. 8789–8798.

[4] F.-P. An, J. e Liu, Pedestrian re-identification algorithm based on visual attention-positive sample generation network deep learning model, Inf. Fusion 86–87 (2022) 136–145.

[5] X. Liu, W. Liu, H. Ma, H. Fu, Large-scale vehicle re-identification in urban surveillance videos, in: Proc. IEEE International Conference on Multimedia and Expo, 2016, pp. 1–6.

[6] H. Liu, Y. Tian, Y. Wang, L. Pang, T. Huang, Deep relative distance learning: Tell the difference between similar vehicles, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[7] Y. Lou, Y. Bai, J. Liu, S. Wang, L. yu Duan, VERI-wild: A large dataset and a new method for vehicle re-identification in the wild, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3230–3238.

[8] G. Haiyun, Z. Chaoyang, L. Zhiwei, W. Jinqiao, L. Hanqing, Learning coarse-to-fine structured feature embedding for vehicle re-identification, in: Proc. AAAI Conference on Artificial Intelligence, Vol. 32, (1) 2018.

[9] A. Lu, C. Li, Y. Yan, J. Tang, B. Luo, RGBT tracking via multi-adapter network with hierarchical divergence loss, IEEE Trans. Image Process. 30 (2021) 5613–5625.

[10] C. Li, H. Cheng, S. Hu, X. Liu, J. Tang, L. Lin, Learning collaborative sparse representation for grayscale-thermal tracking, IEEE Trans. Image Process. 25 (2016) 5743–5756.

[11] Z. Tu, C. Lin, W. Zhao, C. Li, J. Tang, M5L: Multi-modal multi-margin metric learning for RGBT tracking, IEEE Trans. Image Process. 31 (2022) 85–98.

[12] I. Afyouni, Z.A. Aghbari, R.A. Razack, Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey, Inf. Fusion 79 (2022) 279–308.

[13] A. Zheng, Z. Wang, Z.-H. Chen, C. Li, J. Tang, Robust multi-modality person re-identification, in: Proc. AAAI Conference on Artificial Intelligence, Vol. 35, (4) 2021, pp. 3529–3537.

[14] Z. Tu, Z. Li, C. Li, Y. Lang, J. Tang, Multi-interactive dual-decoder for RGB-thermal salient object detection, IEEE Trans. Image Process. 30 (2021) 5678–5691.

[15] H. Li, C. Li, X. Zhu, A. Zheng, B. Luo, Multi-spectral vehicle re-identification: A challenge, in: Proc. AAAI Conference on Artificial Intelligence, 2020, pp. 11345–11353.

[16] A. Hermans, L. Beyer, B. Leibe, In defense of the triplet loss for person re-identification, 2017, arXiv preprint arXiv:1703.07737.

[17] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: Proc. European Conference on Computer Vision, 2016.

[18] Y. Zhu, Z. Yang, L.-C. Wang, S. Zhao, X. Hu, D. Tao, Hetero-center loss for cross-modality person re-identification, Neurocomputing 386 (2020) 97–109.

[19] M. Ye, Z. Wang, X. Lan, P. Yuen, Visible thermal person re-identification via dual-constrained top-ranking, in: Proc. International Joint Conference on Artificial Intelligence, 2018.

[20] M. Ye, X. Lan, Z. Wang, P.C. Yuen, Bi-directional center-constrained top-ranking for visible thermal person re-identification, IEEE Trans. Inf. Forensics Secur. 15 (2020) 407–419.

[21] Y. Ling, Z. Zhong, Z. Luo, P. Rota, S. Li, N. Sebe, Class-aware modality mix and center-guided metric learning for visible-thermal person re-identification, in: Proc. ACM International Conference on Multimedia, 2020, pp. 889–897.

[22] J. Liu, Y. Sun, F. Zhu, H. Pei, Y. Yang, W. Li, Learning memory-augmented unidirectional metrics for cross-modality person re-identification, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 19344–19353, http://dx.doi.org/10.1109/CVPR52688.2022.01876.

[23] J. Wu, J. Jiang, M. Qi, C. Chen, J. Zhang, An end-to-end heterogeneous restraint network for RGB-D cross-modal person re-identification, ACM Trans. Multimed. Comput., Commun. Appl. (TOMM) 18 (4) (2022) 1–22.

[24] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, arXiv preprint arXiv:1502.03167.

[25] D. Ulyanov, A. Vedaldi, V.S. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2016, arXiv preprint arXiv:1607.08022.

[26] Y. Wu, K. He, Group normalization, in: Proc. European Conference on Computer Vision, 2018.

[27] J. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, 2016, arXiv preprint arXiv:1607.06450.

[28] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proc. IEEE/CVF International Conference on Computer Vision, 2015, pp. 1116–1124.

[29] Y. Wen, D. Bein, S. Phoha, Dynamic clustering of multi-modal sensor networks in urban scenarios, Inf. Fusion 15 (2014) 130–140.

[30] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, J. Yan, S. Wang, H. Li, X. Wang, Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 379–387.

[31] Y. Shen, T. Xiao, H. Li, S. Yi, X. Wang, Learning deep neural networks for vehicle re-ID with visual-spatio-temporal path proposals, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 1918–1927.

[32] B. He, J. Li, Y. Zhao, Y. Tian, Part-regularized near-duplicate vehicle re-identification, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3992–4000.

[33] H. Li, C. Li, A. Zheng, J. Tang, B. Luo, Attribute and state guided structural embedding network for vehicle re-identification, IEEE Trans. Image Process. 31 (2022) 5949–5962.

[34] P. Khorramshahi, A. Kumar, N. Peri, S.S. Rambhatla, J. Chen, R. Chellappa, A dual-path model with adaptive attention for vehicle re-identification, in: Proc. IEEE/CVF International Conference on Computer Vision, 2019, pp. 6131–6140.

[35] D. Meng, L. Li, X. Liu, Y. Li, S. Yang, Z. Zha, X. Gao, S. Wang, Q. Huang, Parsing-based view-aware embedding network for vehicle re-identification, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7101–7110.

[36] Y. Yao, L. Zheng, X. Yang, M.R. Naphade, T. Gedeon, Simulating content consistent vehicle datasets with attribute descent, in: Proc. European Conference on Computer Vision, 2020.

[37] A. Wu, W. Zheng, H.-X. Yu, S. Gong, J. Lai, RGB-infrared cross-modality person re-identification, in: Proc. IEEE International Conference on Computer Vision, 2017, pp. 5390–5399.

[38] T.D. Nguyen, H. Hong, K. Kim, K. Park, Person recognition system based on a combination of body images from visible light and thermal cameras, Sensors (Basel, Switzerland) 17 (2017).

[39] G. Wang, T. Zhang, J. Cheng, S. Liu, Y. Yang, Z.-H. Hou, RGB-infrared cross-modality person re-identification via joint pixel and feature alignment, in: Proc. IEEE/CVF International Conference on Computer Vision, 2019, pp. 3622–3631.

[40] D. Li, X. Wei, X. Hong, Y. Gong, Infrared-visible cross-modal person re-identification with an X modality, in: Proc. AAAI Conference on Artificial Intelligence, 2020, pp. 4610–4617.

[41] Y. Lu, Y. Wu, B. Liu, T. Zhang, B. Li, Q. Chu, N. Yu, Cross-modality person re-identification with shared-specific feature transfer, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13376–13386.

[42] N. Huang, J. Liu, Y. Miao, Q. Zhang, J. Han, Deep learning for visible-infrared cross-modality person re-identification: A comprehensive review, Inf. Fusion 91 (2023) 396–411.

[43] Z. Wei, X. Yang, N. Wang, X. Gao, Syncretic modality collaborative learning for visible infrared person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 225–234.

[44] M. Ye, C. Chen, J. Shen, L. Shao, Dynamic tri-level relation mining with attentive graph for visible infrared re-identification, IEEE Trans. Inf. Forensics Secur. 17 (2021) 386–398.

[45] Z. Wei, X. Yang, N. Wang, X. Gao, Flexible body partition-based adversarial learning for visible infrared person re-identification, IEEE Trans. Neural Netw. Learn. Syst. 33 (9) (2021) 4676–4687.

[46] Z. Wei, X. Yang, N. Wang, X. Gao, Rbdf: Reciprocal bidirectional framework for visible infrared person reidentification, IEEE Trans. Cybern. 52 (10) (2022) 10988–10998.

[47] I.B. Barbosa, M. Cristani, A.D. Bue, L. Bazzani, V. Murino, Re-identification with RGB-D sensors, in: Proc. European Conference on Computer Vision Workshops, 2012.

[48] A. Møgelmose, C. Bahnsen, T. Moeslund, A. Clapés, S. Escalera, Tri-modal person re-identification with RGB, depth and thermal features, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 301–307.

[49] M. Munaro, A. Basso, A. Fossati, L.V. Gool, E. Menegatti, 3D reconstruction of freely moving persons for re-identification with a depth sensor, in: Proc. IEEE International Conference on Robotics and Automation, 2014, pp. 4512–4519.

[50] A. Wu, W. Zheng, J. Lai, Robust depth-based person re-identification, IEEE Trans. Image Process. 26 (2017) 2588–2603.

[51] F. Hafner, A. Bhuiyan, J.F.P. Kooij, E. Granger, A cross-modal distillation network for person re-identification in RGB-depth, 2018, arXiv preprint arXiv:1810.11641.

[52] J. Chen, X. Chen, S. Chen, Y. Liu, Y. Rao, Y. Yang, H. Wang, D. Wu, Shape-former: Bridging CNN and transformer via ShapeConv for multimodal image matching, Inf. Fusion 91 (2023) 445–457.

[53] H. Luo, W. Jiang, Y. Gu, F. Liu, X. Liao, S. Lai, J. Gu, A strong baseline and batch normalization neck for deep person re-identification, IEEE Trans. Multimed. 22 (2020) 2597–2609.

[54] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[55] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: Proc. IEEE/CVF International Conference on Computer Vision, 2009.

[56] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2015, CoRR arXiv:1412.6980.

[57] M. Sandler, A.G. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[58] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 2011–2023.

[59] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[60] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[61] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[62] S.V. Huynh, N.-H. Nguyen, N.-T. Nguyen, V. Nguyen, C. Huynh, C.H. Nguyen, A strong baseline for vehicle re-identification, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021, pp. 4142–4149.

[63] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Omni-scale feature learning for person re-identification, in: Proc. IEEE/CVF International Conference on Computer Vision, 2019, pp. 3701–3711.

[64] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S.C.H. Hoi, Deep learning for person re-identification: A survey and outlook, IEEE Trans. Pattern Anal. Mach. Intell. PP (2021).

[65] S. He, H. Luo, P. Wang, F. Wang, H. Li, W. Jiang, TransReID: Transformer-based object re-identification, in: Proc. IEEE/CVF International Conference on Computer Vision, 2021, pp. 15013–15022.

[66] G. Chen, T. Zhang, J. Lu, J. Zhou, Deep meta metric learning, in: Proc. IEEE/CVF International Conference on Computer Vision, 2019.

[67] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zheng, Z. Wang, Y. Wei, Circle loss: A unified perspective of pair similarity optimization, in: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6397–6406.

[68] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: Person retrieval with refined part pooling, in: Proc. European Conference on Computer Vision, 2018.

[69] G. Wang, Y. Yuan, X. Chen, J. Li, X. Zhou, Learning discriminative features with multiple granularities for person re-identification, in: Proc. ACM International Conference on Multimedia, 2018.

[70] J. Zhao, Y. Zhao, J. Li, K. Yan, Y. Tian, Heterogeneous relational complement for vehicle re-identification, in: Proc. IEEE/CVF International Conference on Computer Vision, 2021, pp. 205–214.

[71] L. van der Maaten, G.E. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

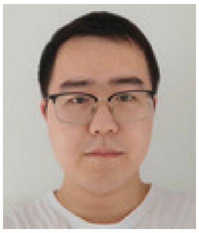
[72] H. Park, S. Lee, J. Lee, B. Ham, Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences, in: Proc. IEEE/CVF International Conference on Computer Vision, 2021, pp. 12046–12055.

[73] M. Ye, J. Shen, D. J Crandall, L. Shao, J. Luo, Dynamic dual-attentive aggregation learning for visible-infrared person re-identification, in: Proc. European Conference on Computer Vision, Springer, 2020, pp. 229–247.

[74] Q. Wu, P. Dai, J. Chen, C.-W. Lin, Y. Wu, F. Huang, B. Zhong, R. Ji, Discover cross-modality nuances for visible-infrared person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4330–4339.

[75] Y. Zhang, Y. Yan, Y. Lu, H. Wang, Towards a unified middle modality learning for visible-infrared person re-identification, in: Proceedings of the 29th ACM International Conference on Multimedia, MM '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 788–796, http://dx.doi.org/10.1145/3474085.3475250.

Aihua Zheng received B.Eng. degrees and finished Master-Docter combined program in Computer Science and Technology from Anhui University of China in 2006 and 2008, respectively. And received Ph.D. degree in computer science from University of Greenwich of UK in 2012. She visited University of Stirling and Texas State University during June to September in 2013 and during September 2019 to August 2020 respectively. She is currently an Associate Professor and PhD supervisor at the School of Artificial Intelligence, Anhui University. Her main research interests include vision based artificial intelligence and pattern recognition. Especially on person/vehicle re-identification, audio visual computing, and multi-modal intelligence.



Xianpeng Zhu received his B.Eng. degree in 2018 and is currently pursuing the M.Eng degree in the School of Computer Science and Technology, Anhui University, Hefei, China. His research interests include Computer Vision, Vehicle Re-identification, Multi-modal Intelligence and Deep Learning.

Zhiqi Ma received his B.Eng. degree in 2021 and is currently pursuing the M.Eng degree in the School of Computer Science and Technology, Anhui University, Hefei, China. His research interests include Computer Vision, Multi-modal Intelligence and Vehicle Re-identification.

Chenglong Li received the M.S. and Ph.D. degrees from the School of Computer Science and Technology, Anhui University, Hefei, China, in 2013 and 2016, respectively. From 2014 to 2015, he worked as a Visiting Student with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. He was a postdoctoral research fellow at the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He is currently an Associate Professor and PhD supervisor at the School of Artificial Intelligence, Anhui University. His research interests include computer vision and deep learning. He was a recipient of the ACM Hefei Doctoral Dissertation Award in 2016.

Jin Tang received the B.Eng. degree in automation and the Ph.D. degree in Computer Science from Anhui University, Hefei, China, in 1999 and 2007, respectively. He is currently a Professor and PhD supervisor with the School of Computer Science and Technology, Anhui University. His research interests include computer vision, pattern recognition, and machine learning.

Jixin Ma is a Full Professor and the Director of PhD/MPhil programme in the School of Computing and Mathematical Sciences, at University of Greenwich, U.K. Prof. Ma obtained his BSc and MSc of Mathematics in 1982 and 1988, respectively, and PhD of Computer Sciences in 1994. His research interests include Temporal Logic, Temporal Databases, Reasoning about Action and Change, Case-Based Reasoning, Pattern Recognition, Machine Learning and Information Security.