Engineering Applications of Artificial Intelligence 150 (2025) 110557

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Research paper

Keypoint-guided feature enhancement and alignment for cross-resolution vehicle re-identification

Aihua Zheng [a], Longfei Zhang [a], Weijun Zhang [b], Zi Wang [c,d], Chenglong Li [a,*], Xiaofei Sheng [e]

[a] Information Materials and Intelligent Sensing Laboratory of Anhui Province, 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 Biomedical Engineering, Anhui Medical University, HeFei, 230032, China
[d] MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, HeFei, 230026, China
[e] Wuhu Simba Network Technology Co, WuHu, 241002, China

ARTICLE INFO

Keywords:
Vehicle re-identification
Cross-resolution retrieval
Keypoint guidance
Teacher–student network

ABSTRACT

Resolution mismatch between low-resolution query images and high-resolution gallery images in vehicle re-identification is rarely studied but ubiquitous in real-world applications. An intuitive approach to solving cross-resolution vehicle re-identification is to utilize super-resolution algorithms to recover detailed information from low-resolution query images. However, vehicle super-resolution algorithms not only recover the detailed information of the vehicle but also enhance the background noise, which would degrade the re-identification performance. In addition, the view mismatch problem also significantly limits the performance of vehicle re-identification. To handle these problems, we propose a novel Keypoint Guiding Network, which simultaneously addresses the problems of resolution mismatch and view mismatch from the perspective of keypoints in an end-to-end learning framework, for cross-resolution vehicle re-identification. In particular, we first generate a set of vehicle keypoints via an effective Gaussian localization method, and then adaptively construct two keypoint-based guidances using attention models. We integrate these two guidances into vehicle super-resolution and view alignment to handle the problems of resolution mismatch and view mismatch respectively. Moreover, to alleviate the heterogeneity between super-resolution query images and high-resolution gallery ones, we design a dual-path teacher–student distillation scheme to narrow their feature distributions. Comprehensive experiments on two down-sampled benchmark datasets demonstrate the effectiveness of our Keypoint Guiding Network against the state-of-the-art methods.

1. Introduction

Vehicle re-identification (Re-ID) refers to the technology of finding the specified vehicle images captured by different cameras. It has a strong application scene in the intelligent video surveillance system and can track the trail of the target to find its position. Most existing Re-ID studies (Fu et al., 2022; Li et al., 2022; Shen et al., 2023; Zheng et al., 2023; Hu et al., 2024; Zhang et al., 2024) assume that both query and gallery images possess similar resolutions. In real-world Re-ID scenarios, however, the resolution of captured vehicles may greatly vary due to the uncontrollable distances between vehicles and cameras. This can lead to misalignment challenges between low-resolution (LR) query sets and high-resolution (HR) gallery sets, which substantially degrades Re-ID performance as shown in Fig. 1(a). Lots of efforts are developed to deal with the resolution mismatch problem in cross-resolution person Re-ID. Early works (Jing et al., 2015; Wang et al., 2016; Chen et al., 2019b; Wu et al., 2023) mainly focus on learning a shared feature space between low and high resolutions. The main issue with these methods is that LR images lack the fine-grained details that are present in HR images. These details are significantly lost during the shared feature space learning. Recent efforts (Zheng et al., 2018a; Jiao et al., 2018; Cheng et al., 2020; Han et al., 2020; Zheng et al., 2022) devote to adopting a multi-task joint learning framework which cascades super-resolution (SR) and Re-ID to recover the missing details in LR images for more robust person Re-ID. Although SR methods handle resolution mismatch problems by recovering missing information of low-resolution query images, the process of super-resolution not only recovers the detailed information of low-resolution objects but also amplifies background noise, which is detrimental to the forthcoming recognition task. Meanwhile, the view mismatch problem also significantly limits the performance of vehicle Re-ID as shown in Fig. 1(a).

* Corresponding author.
E-mail address: lcl1314@foxmail.com (C. Li).

https://doi.org/10.1016/j.engappai.2025.110557
Received 26 June 2024; Received in revised form 19 December 2024; Accepted 10 March 2025
Available online 21 March 2025

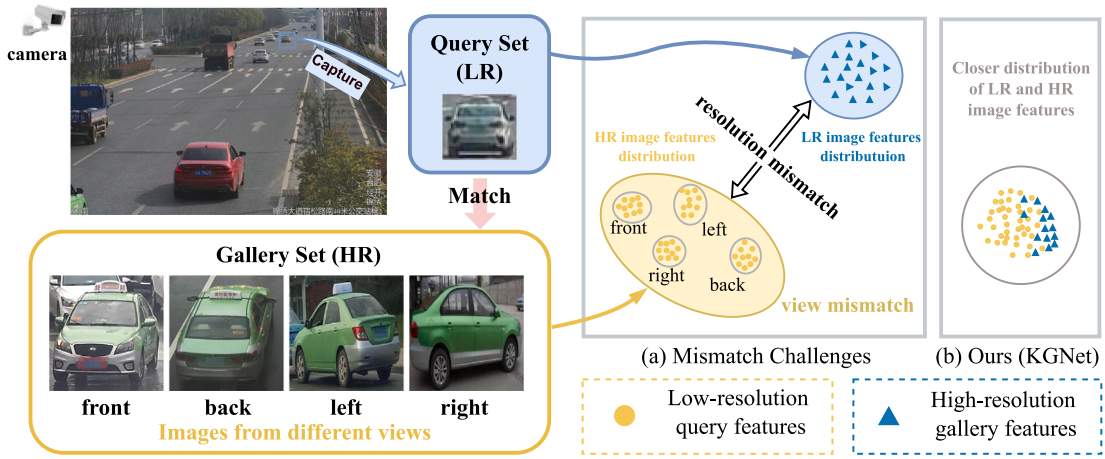

Fig. 1. Overview of the cross-resolution vehicle Re-ID task. The camera captures LR vehicle images and matches them with HR vehicle images from different viewpoints. (a) Mismatch challenges. The resolution mismatch between the LR query and the HR gallery images. Different angles of the vehicles cause the viewpoint mismatch. (b) Our KGNet brings LR and HR image feature distribution closer by dual keypoint guidance.

Keypoint, as one of the most important structural characteristics of the vehicle, can capture localized discriminative features by focusing attention on the most informative keypoint. To take advantage of keypoint information in cross-resolution Re-ID, we propose a novel network called KGNet, which simultaneously addresses the problems of resolution mismatch and view mismatch from the perspective of keypoint in an end-to-end learning framework. On the one hand, we propose to recover detailed information from LR images by keypoint guiding super-resolution to enhance recovery of the object and suppress the background noise. Specifically, we use the keypoints to locate each key part of the vehicle and divide all heatmaps into four parts according to the vehicle views to generate an attention map for each part. By incorporating the attention map into the extracted features, we can reinforce the recovery of vehicle-based views during image super-resolution. On the other hand, we propose aligning vehicle features by keypoint-guided view alignment to handle the view mismatch problem. In particular, by embedding keypoint attention maps in different views into SR images, the network can align vehicle features with the location of the keypoints. We integrate these two guidances into vehicle super-resolution and view alignment to handle the problems of resolution mismatch and view mismatch respectively as shown in Fig. 1(b).

Despite the benefit of the proposed dual keypoint guidances to enhance recovery and align features, invisible keypoints in the vehicle may cause the network to focus on the wrong locations and thus introduce wrong guidance. Existing vehicle keypoint detection methods (Wang et al., 2017; Khorramshahi et al., 2019) often use the binary map to learn heatmaps of both visible and invisible keypoints, which has two shortcomings. First, the binary map is sensitive to the noises of the keypoint location. Second, it leads to a haphazard distribution of invisible keypoints, and their wrong positions would lead to wrong guidance in both super-resolution and view alignment. To handle these problems, we propose to define keypoints in the Gaussian map and present all invisible keypoints at a fixed position. On one hand, the Gaussian map is insensitive to noises and thus robust in keypoint localization. On the other hand, by regressing all invisible keypoints to a fixed position, we can easily filter them out and alleviate their effects in the dual keypoint guidances.

In addition, existing cross-resolution methods (Jiao et al., 2018; Li et al., 2020a) usually use a single-path network to extract features while ignoring the heterogeneity between SR and HR images, which leads to the misalignment between the extracted features. In this paper, we propose a dual-path teacher–student distillation network to mitigate the heterogeneity between SR and HR images. Moreover, we use a distillation loss (Porrello et al., 2020) to further narrow the feature distributions between the SR result and the HR image to facilitate the progressive embedding of keypoints. The major contributions of this paper are as follows.

(1) We propose an end-to-end learning framework for cross-resolution vehicle Re-ID. To handle the resolution mismatch problem between query and gallery sets and the view mismatch, we adaptively construct two attention model-based keypoint guidances integrating into vehicle super-resolution and view alignment.

(2) We design a Gaussian keypoint localization to provide more accurate keypoint guidance. It improves the accuracy of visible keypoint localization, while simultaneously eliminating the effects of invisible keypoints by presenting at a fixed position.

(3) We propose a dual-path teacher–student distillation network to alleviate the heterogeneity between SR and HR images in query and gallery sets respectively. Moreover, we design a distillation loss to narrow the distribution of features between SR and HR images and facilitate the dual keypoint guidance.

Our approach also holds significant importance in practical applications, KGNet can be integrated with existing monitoring systems and work alongside real-time image processing technologies to achieve efficient and accurate vehicle recognition across images of varying resolutions, significantly enhancing the performance of current intelligent traffic monitoring systems. In addition, the method can help with traffic monitoring and management for urban planning by accurately identifying and tracking vehicles in real time, and generating detailed vehicle flow data. This data can be used to optimize traffic patterns, improve infrastructure planning, and reduce congestion. The paper is organized as follows. Section 2 provides an overview of the works related to cross-resolution and keypoints based Re-ID. Section 3 systematically elaborates on the proposed KGNet, including the keypoint-guided recovery mechanism and alignment mechanism, as well as teacher–student distillation. Section 4 shows the comprehensive experimental results of KGNet. Finally, Section 5 concludes the paper.

2. Related works

Currently, research on the cross-resolution vehicle Re-ID is quite limited. In this section, we briefly review the most related works of cross-resolution vehicle Re-ID and keypoint-based Re-ID and super-resolution.

2.1. Cross-resolution vehicle Re-ID

Vehicle Re-ID has become a hot topic recently due to its wide use in intelligent transportation systems. Some global-based methods

(Liu et al., 2016a; Lin et al., 2019; Jiang et al., 2023)focus on extracting comprehensive vehicle representations from the entire image, enabling robust matching across different camera views and varying conditions. Lin et al. (2019) proposed modeling the vehicle Re-ID task as vehicle matching within and across views by representing vehicle views as latent groups, using only ID annotations without additional labels. Jiang et al. (2023) proposed using a global attention mechanism to extract more useful information for vehicle Re-ID. The papers described above are all based on global features. However, global features fail to capture the subtle differences in vehicle images. Therefore, most subsequent methods (Liu et al., 2020; Lu et al., 2022; Wang et al., 2023b; Chouchane et al., 2024) focus on exploring how to better learn local features from the most discriminative regions of vehicle images. Lu et al. (2022) used a unified Vision Transformer (ViT) framework to extract global features unrelated to the background and locally variable features in perspective. Wang et al. (2023b) improved the network's attention to shallow features by combining the channel and spatial dimensions at each layer. Although vehicle Re-ID has been extensively studied, research on cross-resolution vehicle Re-ID remains limited. However, similar studies have already been explored in the context of person Re-ID. The two tasks are similar, and the methods used are also closely related. Next, we review cross-resolution person Re-ID works in this section, which can be divided into two main categories. The first is to learn the associations and shared features between the LR and HR images. Li et al. (2015) jointly utilize multi-scale distance metric learning and cross-scale image domain alignment for low-resolution person Re-ID. Wang et al. (2016) learn to distinguish the scale distance function space by changing the image scale of the LR image when matching with the HR image. Wu et al. (2023) produce resolution-adaptive representations to resolve the feature differences caused by different resolutions. However, the shared features and associations learned by these methods always lose details in LR images compared to HR images thus resulting in poor Re-ID performance. The second one is to learn a joint model for both image super-resolution and person Re-ID Jiao et al. (2018) propose the joint learning network of super-resolution and identity, and solve low-resolution person Re-ID problems by cascading multiple SR and Re-ID modules. Han et al. (2020) predict an effective scale factor based on the image content to recover missing details adaptively. Zheng et al. (2022) collaboratively learns both HR-specific and LR-specific identity features by introducing a synergistic interplay between super-resolution and discriminant re-id feature learning. Despite their great progress on cross-resolution person Re-ID, they ignore the heterogeneity between SR and HR images. In addition, SR methods not only recover the detailed information but also enhance the background noise.

2.2. Keypoints based Re-ID and super-resolution

As an important representation,keypoints provide a robust guide for Re-ID. In person Re-ID. Some methods, especially in the occluded scene, make use of the keypoints information of the human body. Liang et al. (2022) propose innovative modules in image, feature space and loss, guided by human keypoint information, to obtain coarse-grained global and fine-grained local embeddings. Miao et al. (2019) proposes a method called Pose-Guided Feature Alignment (PGFA), which utilizes pose landmarks to enhance feature learning and extract non-occluded representations. Gao et al. (2020) introduces Pose-guided Visible Part Matching (PVPM), aiming to generate discriminative embeddings with the assistance of pose-guided attention. In vehicle Re-ID, Wang et al. (2017) propose an orientation invariant feature embedding module to extract local region features of different orientations and the local features can be well aligned and combined. Khorramshahi et al. (2019) propose to learn the key-points based local feature together with the global appearance feature via a dual path model. Tang et al. (2020) propose to infer the vehicle view and shape based on the keypoints and segments in view prediction to overcome the orientation dependence. However, they either fail to eliminate the impact of invisible keypoints or utilize only limited keypoint information. Moreover, keypoints are also applied to the image recovery. Ma et al. (2020) develop a face super-resolution method with two recursive networks in an iterative fashion to simultaneously enforce the face component recovery and keypoint prediction. Yu et al. (2018) proposes to utilize a multi-task convolutional neural framework to integrate the face keypoint information into the face super-resolution process. Different from these works, we adaptively construct dual keypoint guidance to solve resolution mismatch and view mismatch for cross-resolution vehicle Re-ID.

3. Keypoint Guiding network

As shown in Fig. 2, Keypoint Guiding Network (KGNet) consists of a Dual Keypoint Guiding (K^2G) module and a Teacher–Student Distillation (TSD) module. First, to mitigate the resolution and viewpoint mismatch problem in cross-resolution vehicle Re-ID task, we propose a Dual Keypoint Guiding (K^2G) module for cross-resolution vehicle Re-ID. Second, to alleviate the heterogeneity between SR and HR images, we design a Teacher–Student Distillation (TSD) module to narrow the distribution between SR and HR features.

3.1. Network architecture

In the training stage, we first employ bilinear up-sampling on the original LR image x_{lr} to obtain the upsampled image x'_{lr}, which is then input into a pre-trained stacked hourglasses network (Newell et al., 2016) to extract the keypoint heatmaps. The influence of invisible keypoints will be eliminated through the Gaussian keypoint Localization. To recover the content information and missing details of LR image x_{lr}, and reinforce the recovery of vehicle-based orientations, we fuse the outputs after the convolution of x_{lr} and the attention map of keypoints based four orientations, and then forward to the super-resolution scheme to acquire SR results x_{sr}. Progressively, we use the keypoint attention information to fuse SR feature f_{sr} to enforce the network to better focus on the location of vehicle keypoints and achieve feature alignment with the location of keypoints. To solve the heterogeneity problem between the SR result x'_{sr} and the HR image x_{hr} and facilitate the dual guidances process of keypoints, we design a dual-path teacher–student distillation network to narrow their feature distributions. By knowledge distillation learning, the teacher–student network encourages our super-resolution process to generate more perceptually realistic outputs and align local features which is associated with the task of vehicle Re-ID.

In the testing stage, to enhance the recovery of the vehicle and align local features, we input the LR query image x_{lr} into K^2G module to obtain the high-quality and aligned SR result x'_{sr}. To alleviate the feature distribution discrepancy between LR and HR images and obtain more robust features, the recovered image x'_{sr} and HR gallery image x_{hr} are fed into the bottom student and top teacher networks respectively for feature extraction. At last, the learned SR and HR features are used for the final Re-ID.

3.2. Dual keypoint guiding (K^2G) module

Gaussian Keypoint Localization: GKL In the keypoint estimation stage, to obtain accurate keypoint heatmaps, we employ a stacked hourglass network (Newell et al., 2016) to estimate the location of $N(N = 20)$ heatmaps with the size of $H \times W(32 \times 32)$. Existing vehicle-based methods (Khorramshahi et al., 2019) only utilize the binary map to represent the ground-truth heatmaps of vehicles and simply set the heatmap values to one in the location of visible keypoints and zero for the rest locations. However, the binary map is sensitive to noise in locating keypoints and affects the robustness of locating keypoints, thus leading to inaccurate predictions. To obtain more effective information

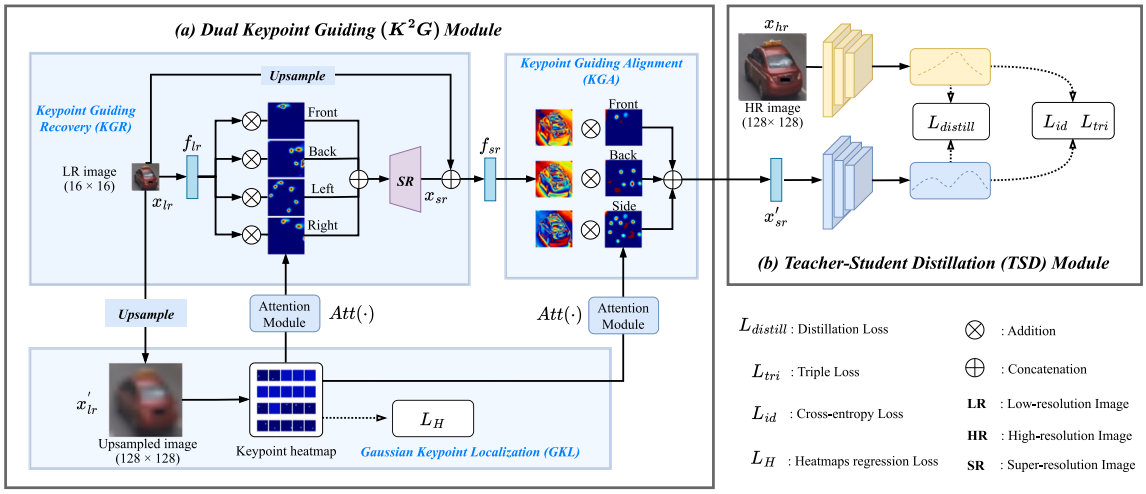

Fig. 2. Overview of the proposed Keypoint Guiding Network (KGNet) which consists of (a) the Dual Keypoint Guiding (K^2G) Module and (b) the Teacher–Student Distillation (TSD) Module. Specifically, Gaussian Keypoint Localization (GKL) at the bottom of (a) generates a set of vehicle keypoints. Then, Keypoint Guiding Recovery (KGR) and Keypoint Guiding Alignment (KGA) at the top of (a) enhance the recovery and feature alignment of the LR images x_{lr} using the generated vehicle keypoints. The SR model used in KGR is designed based on a modified Omni Self-Attention (OSA) (Wang et al., 2023a). Meanwhile, the Teacher–Student Distillation (TSD) Module is used to mitigate the heterogeneity between the SR images x'_{sr} and HR images x_{hr}.

of keypoints, we set the ground-truth keypoint heatmaps using the Gaussian-like distribution as follows,

$$h_k(i,j) = e^{-\frac{(i-x)^2+(j-y)^2}{2}}, \tag{1}$$

where h_k represent the kth ground-truth heatmap, and (i, j) and (x, y) indicate the coordinates of the kth ground-truth heatmap and the kth keypoint respectively. The coordinates of the invisible keypoints are set to $x = 0$, $y = 0$.

In this case, the ground-truth heatmaps follow the Gaussian distribution for visible keypoints and gather all invisible keypoints to a fixed position (0, 0). Compared to binary map, the Gaussian map is insensitive to noises of keypoint location and thus robust in keypoint localization. Due to the robustness of the Gaussian map to the keypoint location, we can get more robust keypoint heatmaps to assist the next tasks. To regress keypoints, we use the heatmaps regression loss $\mathcal{L}_H$ as the supervision of the training of keypoints,

$$\mathcal{L}_H = \sum_{k=1}^{N}\sum_{i=1}^{H}\sum_{j=1}^{W} || h_k(i,j) - h'_k(i,j) ||_1, \tag{2}$$

where $h'_k(i,j)$ is the predictive value of coordinate (i, j) of the kth heatmap. The surface diagram of the predicted keypoint heatmaps is shown in Fig. 3. After training the stacked hourglass network on VeRi-776 dataset, the parameters of the network are frozen. To this end, we can directly test on other datasets such as VehicleID to obtain the keypoint information without additional annotations.

However, the presence of invisible keypoints in the vehicle greatly affects both tasks, which causes the network to focus on the wrong locations and thus introduces wrong guidance. To alleviate the effects of invisible keypoints, we regress all invisible keypoints to a fixed position and filter out them as follows,

$$h_k = \begin{cases} \mathbf{0}, & \text{if } LOC(Max(h_k)) = (0,0) \\ h_k, & \text{else,} \end{cases} \tag{3}$$

where $Max(h_k)$ is the maximum of the kth heatmap, $LOC(*)$ is the coordinates of the maximum value. Herein, we can get robust keypoint heatmaps and eliminate the influence of invisible keypoints.

Keypoint Guiding Recovery: KGR Most existing SR methods (Lim et al., 2017; Li et al., 2020b) recover the images based on LR images while rarely considering the inherent structural information, moreover, they not only recover detailed information of object, but also enhance

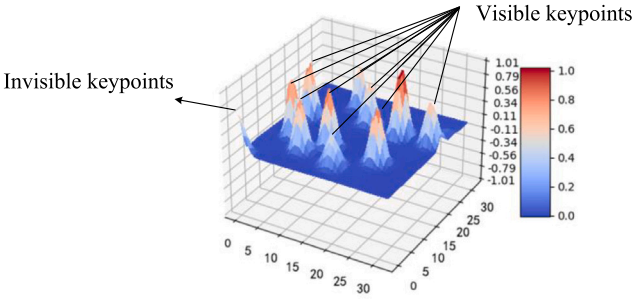

Fig. 3. Surface map of the predicted heatmaps. For the visible keypoints, the peak coordinate of the heatmaps is the coordinate position of the keypoints, while the origin (0, 0) is for the invisible keypoints.

the background noise, which would degrade the vehicle Re-ID performance. To obtain high-quality recovery images and reinforce the recovery of vehicle-based orientations, we propose guidance based keypoints to solve these problems. By incorporating keypoint information into LR images, the network can focus on the key parts of the vehicle, rather than background noise, so as to focus on vehicle recovery. Therefore, the purpose of Keypoint Guiding Recovery is to obtain the SR result x_{sr}, which enhances the recovery of the vehicle and reduces the weight of the noise component by recovering the missing details in LR vehicle object x_{lr}. To utilize the structural information of keypoints, we use the keypoint heatmap with N channels to represent the locations of N keypoints. We divide keypoints into $K(K = 4)$ orientations as (Wang et al., 2017), including $front$, $back$, $left$ and $right$, and sum up the channels in each orientation as the corresponding heatmap, which is expressed as $\{D_k\}_{k=1}^{K}$. The reason why we divide the keypoints according to the orientation is to perform different degrees of recovery for the distinct orientation of each vehicle. Then, we use the softmax function as attention mechanism to calculate K corresponding attention maps along the channel dimension of these heatmaps. The attention map and LR feature then are fused by the group convolution (Ma et al., 2020) to obtain the attention feature f_{att},

$$f_{att} = \sum_{k=1}^{K} (f_{lr} \cdot Att(D_k)), \tag{4}$$

where f_{lr} is the feature of LR image x_{lr} computed by multi-layer convolution, which contains more shallow information for LR images. $Att(*)$ is the softmax function along the channel dimension.

We feed the feature f_{att} to SR module to better recover the missing details of the LR images x_{lr} according to the different orientations of the vehicle, and the input f_{att} of SR module can locate the key parts of the vehicle under the guidance of keypoints. Our SR module is a modification of the Omni Self-Attention (OSA) block (Wang et al., 2023a), which is based on the dense interaction principle. This block can simultaneously model pixel interaction from both spatial and channel dimensions, effectively mining potential correlations across all axes, including both spatial and channel aspects. To preserve the useful information of the original LR image, we add the un-sampled LR image x_{lr} to the result of the SR module. The former contains more content information of the LR image, while the latter contains the recovered detailed information. The pixel-wise MSE loss between SR image x_{sr} and the ground-truth HR image x_{hr} is calculated as,

$$\mathcal{L}_{rec} = \| x_{hr} - x_{sr} \|_2^2 . \tag{5}$$

Keypoint Guiding Alignment: KGA Although we achieve enhanced recovery of vehicle objects by keypoint guidance, the diverse orientations of the vehicle will lead to feature misalignment. To capture and align localized discriminative features, we focus on the most informative keypoints by embedding keypoint attention maps based on orientation to the recovered image.

As shown in Fig. 2, similar to Keypoint Guiding Recovery, Keypoint Guiding Alignment first performs a multi-layer residual convolution operation on the SR result x_{sr} to obtain SR features f_{sr} which contains more contextual information. We divide the heatmaps of keypoints, which have eliminated the effect of invisible keypoints, into three orientations: $front$, $back$ and $side$. Then we generate orientation attention maps by our attention mechanism that utilizes softmax function on three channels along the heatmaps. To construct feature alignment at the channel level, the attention maps generated in each orientation are multiplied two-by-two with the three channels of f_{sr}. To this end, the network well focuses on the location of keypoints, we can achieve the alignment of vehicle features with the location of the keypoints and alleviate the problem of view mismatch that exists in vehicle Re-ID. In addition, the dual keypoint guidances are jointly optimized. Therefore, KGR and KGA promote each other in a unified framework. Finally, their summed results are fed into the Re-ID network for further feature extraction.

3.3. Teacher–student distillation (TSD) module

Although the SR process can recover realistic high-quality images, our ultimate goal is to contribute to the Re-ID task. Therefore, we not only expect to reduce the difference between the recovered SR and HR images at the image level, but also anticipate narrowing their feature distributions. To obtain a more robust SR feature associated with the Re-ID task, we propose a teacher–student distillation network to narrow the distance between SR and HR features, and the feature learning process of the student network is guided by the teacher network.

As shown in Fig. 6(a), the teacher–student distillation network takes the keypoints embedded SR result x'_{sr} and the corresponding HR ground-truth image x_{hr} as input, and extracting their feature representations through student branch E_{stu} and teacher branch E_{tea} respectively for final vehicle Re-ID. To further alleviate the difference between HR images and SR images, we propose to introduce a feature distillation loss $\mathcal{L}_{distill}$ to suppress their feature distributions,

$$\mathcal{L}_{distill} = \| E_{tea}(x_{hr}) - E_{stu}(x'_{sr}) \|_2^2 . \tag{6}$$

It needs to emphasize that the main purpose of the SR module is to make the recovered SR images better serve our vehicle Re-ID, rather than simply recovering the missing details of the low-resolution vehicle images. In this case, we encourage the SR module to perform Re-ID oriented recovery through feature distillation loss $\mathcal{L}_{distill}$.

Due to the heterogeneity between HR ground-truth images and recovered SR results, we use the same architecture without sharing parameters for teacher and student networks. During the testing phase, we feed the LR images in query and the HR images in the gallery into student and teacher branches respectively for cross-resolution vehicle Re-ID. In both teacher and student branches, the Re-ID loss is the integration of cross-entropy loss $\mathcal{L}_{id}$ and triplet loss $\mathcal{L}_{tri}$:

$$\mathcal{L}_{tea} = \mathcal{L}_{id}^{hr} + \mathcal{L}_{tri}^{hr}. \tag{7}$$

$$\mathcal{L}_{stu} = \mathcal{L}_{id}^{sr} + \mathcal{L}_{tri}^{sr}. \tag{8}$$

The final Re-ID loss $\mathcal{L}_{reid}$ is the sum of $\mathcal{L}_{tea}$ and $\mathcal{L}_{stu}$:

$$\mathcal{L}_{reid} = \mathcal{L}_{tea} + \mathcal{L}_{stu}. \tag{9}$$

At last, the final loss of KGNet is formulated as,

$$\mathcal{L}_{final} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{d}\mathcal{L}_{distill} + \mathcal{L}_{reid}, \tag{10}$$

where λ_{rec} and $\lambda_{distill}$ are the hyper-parameters setting as 50.0 and 10.0 respectively. Based on the proposed Keypoint Guiding network, we achieve enhanced recovery of vehicle objects and local feature alignment by the guidance of keypoints, and distillation loss further narrows the feature distribution between SR results and HR images.

4. Experiments

To fairly evaluate the proposed method, we reconstruct two benchmarks, VeRi-776 (Liu et al., 2017) and VehicleID (Liu et al., 2016b) to the cross-resolution scenario. Specifically, during training, we first resize the resolution to $128 \times 128 \times 3$ as the HR images, which is then down-sampled into $16 \times 16 \times 3$ as the LR images in both model training and testing. Note that HR images are no longer required while processing LR query images in the testing stage. The datasets used in our experiments are publicly available and have been anonymized, with all license plate information removed. To protect user privacy in practical applications, obvious vehicle identifiers, such as license plates and driver facial features, can be obscured during data processing. The system can be deployed in an internal network to ensure data security and used in conjunction with a data de-sensitization system for user privacy protection.

4.1. Datasets and evaluation

VeRi-776 (Liu et al., 2017) is the first large scale vehicle Re-ID dataset. It contains 51032 images of 776 vehicles captured from 20 cameras with different orientations, occlusions and illuminations. 576 vehicle identities are constructed for the training set while the remaining 200 vehicles for the testing set.

VehicleID (Liu et al., 2016b) is an extensive benchmark for vehicle Re-ID. It covers 26,267 vehicle identities with 221,763 images captured under the front or back viewpoint. It contains three different testing sets with small, medium and large sizes respectively. In the inference phase, one image of each vehicle is randomly selected to form the gallery set, while the rest images are used to form the query set.

Following Zheng *et al.* (Zheng et al., 2018b), we use the Cumulative Matching Characteristic (CMC) curve and mean average precision (mAP) for evaluation. CMC scores reflect the retrieval precision, where rank-1, rank-5 and rank-10 scores are reported in our experiments. mAP measures the mean of all queries of average precision (the area under the Precision–Recall curve) which reflects the recall.

Table 1
Compared to the conventional state-of-the-art vehicle Re-ID methods on reconstructed VeRi-776 and VehicleID datasets (in %).

Datasets	Reconstructed VeRi-776				Reconstructed VehicleID								
Settings	Query = 1678, Test = 11579				Test Size = 800			Test Size = 1600			Test Size = 2400		
Methods	mAP	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10
ViT (Dosovitskiy et al., 2020)	24.3	45.5	66.6	75.9	20.4	41.3	51.5	20.0	40.2	50.4	19.6	39.5	49.5
ResNet50 (He et al., 2016)	30.2	59.5	71.3	76.6	45.2	60.3	68.4	44.4	58.9	64.2	43.5	61.4	67.6
SEResNet50 (Jie et al., 2020)	33.1	63.8	78.4	84.6	41.7	58.2	66.5	34.4	56.6	62.1	28.6	52.4	60.8
SEResNext50 (Xie et al., 2017)	31.3	61.6	75.9	83.2	35.9	53.8	60.6	35.5	52.6	59.1	34.8	54.1	58.8
ABDNet (Chen et al., 2019a)	31.9	59.8	70.2	76.8	49.2	65.1	70.3	46.2	59.3	65.7	43.1	62.3	68.2
OSNet (Zhou et al., 2020)	32.3	62.4	76.3	82.5	50.5	64.5	71.2	48.3	61.2	66.3	44.9	62.9	67.1
BagTricks (Hao, 2019)	34.7	64.7	77.8	84.1	52.8	68.9	74.9	49.5	65.2	71.2	45.6	63.6	69.3
VOC-ReID (Zhu et al., 2020)	32.3	60.9	77.9	85.3	51.1	63.5	72.1	50.9	60.3	70.4	44.3	58.4	67.6
VehicleNet (Zheng et al., 2020)	32.8	65.7	80.2	85.1	52.3	64.1	75.6	51.4	61.5	73.2	46.7	56.9	68.4
CLIP-ReID (Li et al., 2023)	33.4	47.9	69.4	79.4	51.0	78.1	86.0	47.6	72.0	80.7	40.4	65.4	75.6
TransReID (He et al., 2021)	34.2	68.9	79.5	86.0	52.9	71.4	76.9	50.6	68.2	72.8	47.4	64.1	70.5
AGW (Ye et al., 2021)	37.1	68.8	80.4	85.3	53.5	70.6	77.7	51.4	66.8	73.9	48.6	65.2	71.9
KGNet (Ours)	**57.0**	**83.4**	**92.6**	**95.3**	**69.6**	**87.2**	**91.7**	**65.9**	**80.9**	**87.1**	**62.0**	**77.6**	**83.5**

Table 2
Comparison of the results of different super-resolution methods combined with BagTricks (Hao, 2019) on super-resolution and ReID tasks on reconstructed VeRi-776 and VehicleID datasets (in %). The best, second and third results are in red, blue and green colors, respectively.

Datasets	Reconstructed VeRi-776						Reconstructed VehicleID								
Settings	Query = 1678, Test = 11579						Test Size = 800			Test Size = 1600			Test Size = 2400		
Methods	PSNR	SSIM	mAP	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10
BagTricks (Hao, 2019)	–	-	34.7	64.7	77.8	84.1	52.8	68.9	74.9	49.5	65.2	71.2	45.6	63.6	69.3
+ MemNet (Tai et al., 2017)	23.4	62.8	38.7	67.5	77.3	81.2	47.9	64.8	70.9	43.1	60.4	66.7	38.7	57.6	63.8
+ SRFBN (Li et al., 2019)	24.0	67.3	40.4	70.9	80.6	84.5	61.2	72.8	76.5	59.3	68.2	74.9	56.1	65.2	71.6
+ EDSR (Lim et al., 2017)	24.5	69.7	42.0	71.2	81.8	86.3	59.9	71.2	78.4	54.1	68.4	75.1	51.2	65.4	71.5
+ RCAN (Zhang et al., 2018)	24.8	71.4	42.3	74.1	87.7	88.0	64.5	76.7	80.2	58.7	69.2	74.5	55.9	66.2	72.1
+ DBPN (Haris et al., 2018)	24.7	71.0	43.0	74.7	83.6	86.6	67.1	77.2	82.1	61.5	71.6	76.9	58.2	64.1	70.6
+ LAPAR (Li et al., 2020b)	24.7	71.1	45.6	68.9	85.0	90.3	64.7	78.1	82.7	60.9	72.3	77.5	57.3	68.6	73.4
+ ECBSR (Zhang et al., 2021)	26.5	82.4	54.0	83.0	91.7	95.0	68.4	84.4	90.7	66.5	80.7	86.2	61.2	77.2	83.4
+ SAFMN (Sun et al., 2023)	27.7	85.1	54.2	83.2	92.5	95.8	67.4	83.7	89.6	65.8	81.5	86.3	61.4	77.1	83.6
+ SRFORMER (Zhou et al., 2023)	29.5	90.3	54.6	83.1	92.6	95.1	68.1	84.2	90.6	66.8	80.5	86.0	61.8	77.9	83.0
+ Omni (Wang et al., 2023a)	28.9	88.5	55.6	82.4	90.9	93.6	68.7	85.1	91.0	66.5	80.6	86.5	61.5	77.4	83.1
+ KGR + KGA + TSD (Ours)	29.1	88.9	57.0	83.4	92.6	95.3	69.6	87.2	91.7	65.9	80.9	87.1	62.0	77.6	83.5

4.2. Implementation details

The implementation platform is Pytorch with an NVIDIA GTX 3090 GPU. We use the network pre-trained on ImageNet (Jia et al., 2009) as the backbone. We do not share the parameters for the backbone that is commonly used in both teacher and student networks. We randomly initialize the weights of the classifier. We employ Adam (Kingma and Ba, 2014) optimizer with a batch size of 8. When using the warm-up (Fan et al., 2019) to bootstrap the network, we first increase the learning rate to 2×10^{-4}, then decay it to 2×10^{-5} at the 40-th epoch and 2×10^{-6} at the 70-th epoch. The total epochs of our model during the training is 120.

4.3. Comparison with Re-ID methods

We evaluate our method on the reconstructed cross-resolution datasets compared with a wide range of state-of-the-art Re-ID methods, where we train the corresponding Re-ID model on upsampled LR training images (resize the images to 16×16 and then to 128×128) and test on upsampled LR query images and HR gallery images. As compared in Table 1, although these Re-ID methods achieve superior performance when both query and gallery images are high resolution, they do not address the more common cross-resolution scenarios in real-world applications. Our method introduces SR technology for cross-resolution settings, while simultaneously using vehicle keypoints to reduce the impact of background noise during the SR process. Keypoint-guided vehicle feature alignment is then applied to alleviate the problem of viewpoint mismatch. Finally, distillation is used to further reduce the heterogeneity between SR and HR images. Through these techniques, our KGNet achieves significantly better performance in cross-resolution settings compared to these Re-ID methods.

4.4. Comparison with the super-resolution methods

To further demonstrate the effectiveness of our dual keypoint guidances, we evaluate our KGNet on the reconstructed cross-resolution datasets compared with super-resolution models, as shown in Table 2. Specifically, we use the SR query images and the HR gallery images to individually train the representative Re-ID method BagTricks (Hao, 2019). Clearly, super-resolution-based methods can recover missing details information of LR images to some content. Therefore, by introducing the SR module into the vehicle Re-ID method BagTricks (Hao, 2019), they achieve significant improvement in both mAP and ranking scores. By enhancing the recovery of vehicle objects, and considering the local feature alignment, and alleviating the heterogeneity between SR results and HR images, our KGNet has achieved competitive results, which promises the effectiveness of the proposed method while handling cross-resolution vehicle Re-ID task.

In addition, in order to verify whether the effect of cross-resolution vehicle Re-ID is only related to the performance of image super-resolution, we conduct the experiment to examine the association image SR and vehicle Re-ID. We measure the pixel-wise SR quality of the recovered images by K^2G module by PSNR and SSIM metrics on the reconstructed VeRi-776 test set. As shown in Table 2, although SRFormer (Zhou et al., 2023) slightly outperforms our method in PSNR and SSIM with better super-resolution quality, it works overshadowed in cross-resolution Re-ID task by 2.4% and 0.3% worse in mAP and rank-1 scores than ours. This indicates that the super-resolution quality is not always in direct proportion to the future Re-ID. By the dual keypoint guiding and distillation loss $\mathcal{L}_{distill}$), the recovery of LR images and the alignment of vehicle features are further enhanced.

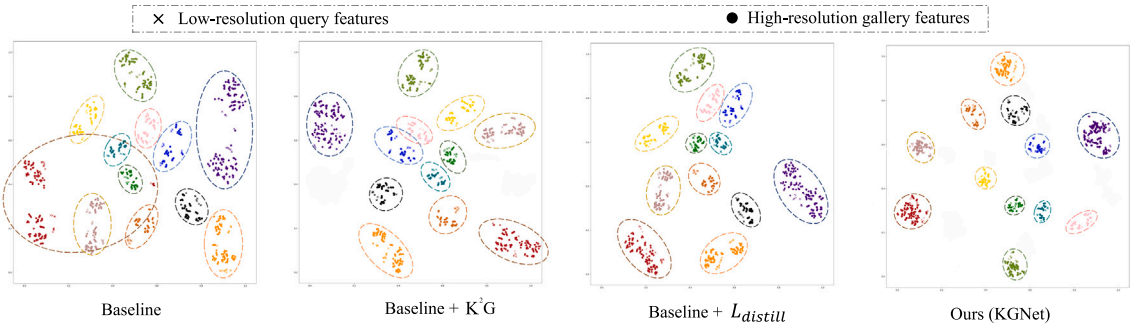

Fig. 4. T-SNE visualization of feature distribution of 12 identities on the test set of reconstructed VeRi-776 dataset. The same color and shape represent the same identity.

Table 3
Comparison results between the proposed method and the cross-resolution person Re-ID methods on reconstructed VeRi-776 dataset (in %).

Methods	rank-1	rank-5	mAP
FSRCNN-reID (Chao et al., 2016)	56.3	78.0	45.8
SING (Jiao et al., 2018)	55.2	77.3	45.1
CSR-GAN (Zheng et al., 2018a)	58.4	80.1	48.5
CADNet (Li et al., 2020a)	68.7	85.3	53.6
Ours (KGNet)	**83.4**	**92.6**	**57.0**

Table 4
Ablation study of KGNet on reconstructed VeRi-776 dataset (in %).

Components		mAP	rank-1	rank-5	rank-10
K^2G	TSD				
–	–	53.3	81.4	90.4	93.5
✓	–	55.8	82.6	90.9	94.3
–	✓	55.3	82.8	91.2	95.0
✓	✓	**57.0**	**83.4**	**92.6**	**95.3**

Table 5
Evaluation on dual keypoint guiding (K^2G) on reconstructed VeRi-776 dataset (in %).

	mAP	rank-1	rank-5	rank-10
w/o KGR	55.4	82.1	92.1	94.0
w/o KGA	55.0	82.0	91.8	94.1
w/o GKL	56.1	82.9	92.0	95.0
Full K^2G	**57.0**	**83.4**	**92.6**	**95.3**

Table 6
Comparative results using single or dual encoders in the teacher–student network on reconstructed VeRi-776 dataset (in %).

Number of encoder	mAP	rank-1	rank-5	rank-10
single encoder	52.8	80.6	91.2	94.6
dual encoders	**57.0**	**83.4**	**92.6**	**95.3**

4.5. Comparison with cross-resolution person Re-ID methods

We compare the proposed KGNet with four existing cross-resolution person Re-ID methods FSRCNN-reID (Chao et al., 2016), SING (Jiao et al., 2018), CSR-GAN (Zheng et al., 2018a) and CADNet (Li et al., 2020a) on the reconstructed VeRi-776 dataset. It is evident from Table 3 that our KGNet outperforms all the competitors in most of the cases. This indicates the advantages of the proposed dual keypoint guiding to capture the structure information and the dual-path teacher–student network to mitigate the heterogeneity between SR and HR images. KGNet works overshadowed CADNet in mAP since CADNet uses multiple pre-trained backbones to iteratively approximate the LR and HR feature distributions. However, we outperform CADNet in rank-1 by a large margin, since our method better solves the resolution mismatch and view mismatch by dual keypoint guidances.

4.6. Ablation studies

To evaluate the effectiveness of the components in our model, we conduct the ablation study on the dual keypoint guiding module, including keypoint guiding recovery (KGR) and keypoint guiding alignment (KGA), and the distillation loss in KGNet on VeRi-776, as shown in Table 4. We employ OSA (Wang et al., 2023a) as SR scheme and Bagtricks (Hao, 2019) as the Re-ID scheme without the dual keypoint guiding (K^2G) and teacher–student distillation(TSD) as the baseline. Note that both dual keypoint guiding module and teacher–student distillation enhance the performance, which verifies the significance of each component. By jointly enforcing both modules, our KGNet achieves the best performance. Fig. 4 further illustrates the example feature distribution of 12 identities on the test set of VeRi-776 dataset. It visually demonstrates the contribution of each component by reducing the intra-class differences and increasing inter-class variation.

Evaluation on dual keypoint guiding module. We further evaluate the validity of keypoint guiding recovery (KGR), keypoint guiding alignment (KGA) and the Gaussian keypoint localization (GKL) in the proposed dual keypoint guiding module as shown in Table 5. Clearly, by removing each component individually, the performance significantly drops, which indicates the important role of each component in our method. First, we can recover detailed information from LR images by KGR to enhance the recovery of the object and suppress the background noise. Second, we further align vehicle features by KGA to handle the view mismatch problem. However, dual keypoint guidance is based on obtaining a robust set of keypoints. Therefore, we utilize Gaussian keypoint localization (GKL) to extract keypoint heatmaps. As shown in Fig. 5, compared to the binary map, our GKL module is insensitive to noises and thus robust in keypoint localization. In addition, GKL can regress all invisible keypoints to a fixed position and we can easily filter them out and alleviate their effects in the dual keypoint guidances.

Evaluation on the teacher–student structure. To validate the effectiveness of the dual-path teacher–student structure, we compare it with the single-encoder structure as shown in Fig. 6(b), which uses only one backbone to extract SR and HR features, but the dual-encoder schematic uses two different backbone which do not share parameters to extract features. As shown in Table 6, the dual-encoder structure outperforms the single-encoder structure in both mAP and rank scores. The main reason is the single-encoder structure cannot handle the heterogeneity between the SR results recovered from the LR image and the HR images, and the dual-encoder structure can better extract robust features according to the characteristics of the two images.

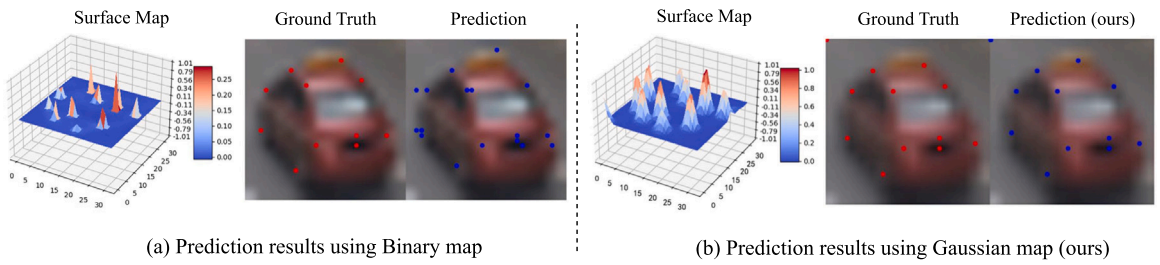

Fig. 5. (a) The surface of the predicted keypoint heatmaps, the ground-truth (GT) and the predicted coordinates by binary map. (b) The corresponding results by our Gaussian map, where the peak coordinates of the heatmaps indicate the position of the visible keypoints while the origin (0, 0) for the invisible keypoints.

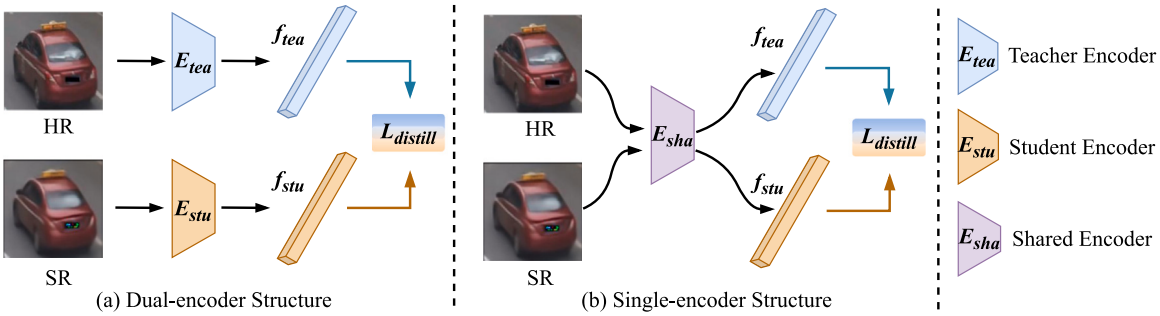

Fig. 6. Schematics of (a) dual-encoder and (b) single-encoder structure designs.

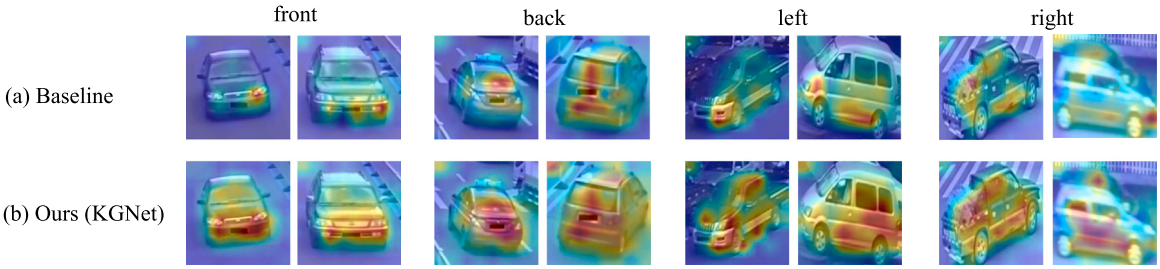

Fig. 7. Feature map visualization of baseline and our KGNet in different orientations.

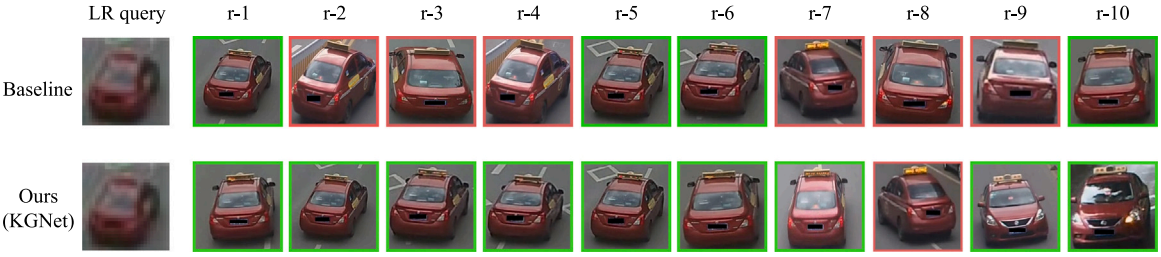

Fig. 8. Visualization of cross-resolution Re-ID results using baseline and the proposed method.

4.7. Qualitative results

To understand how the keypoints and the distillation network help our cross-resolution vehicle Re-ID task, we visualize the last layer feature maps as shown in Fig. 7. The progressive keypoint guidance to recovery and alignment, in conjunction with the teacher–student distillation module, serves to reduce heterogeneity between the super-resolution images and high-resolution images, as illustrated in Fig. 7(b). This enables the network to focus more effectively on the key components of the vehicle, in comparison to Fig. 7(a). Fig. 8 further shows the corresponding ranking results of a particular query, demonstrating the promising performance of our network in dealing with the challenging inter-class similarity.

4.8. Evaluation of computational cost and model parameters

To demonstrate the efficiency of our method, we evaluate the computational and parameter costs of KGNet in comparison with several state-of-the-art super-resolution methods combined with BagTricks (Hao, 2019). As shown in Table 7, Our model's FLOPs are comparable to SAFMN (Sun et al., 2023) and ECBSR (Zhang et al., 2021), and significantly lower than SRFormer (Zhou et al., 2023) (reduced by approximately 94%), while the parameters are on par with them. This indicates that our method offers an advantage in computational efficiency. Moreover, our approach achieves the highest mAP and Rank-1 metrics across two datasets, outperforming these methods. Generally, super-resolution-based methods tend to incur higher computational and parameter costs. Considering the need for end-to-end training

Table 7
Computational and parameter costs of different models. The methods marked with * are based on BagTricks (Hao, 2019) with the corresponding super-resolution models applied.

Model	Flops (Billion)	Parameters (Million)	Reconstructed VeRi-776		Reconstructed VehicleID
			mAP	r-1	r-1 (Test Size = 800)
SRFormer * (Zhou et al., 2023)	146.27	64.49	54.6	83.1	68.1
SAFMN * (Sun et al., 2023)	7.93	59.97	54.2	83.2	67.4
ECBSR * (Zhang et al., 2021)	6.96	59.07	54.0	83.0	68.4
Ours (KGNet)	8.22	64.04	57.0	83.4	69.6

and deployment in real-time systems, our proposed method maintains low model complexity, minimizing computational overhead while preserving performance.

5. Conclusion

In this paper, to handle the resolution mismatch problem between query and gallery sets, as well as the view mismatch problem in cross-resolution vehicle re-identification, we propose a well-designed end-to-end keypoint guiding framework (KGNet). It first guides the super-resolution to recover the detailed information of the vehicle while suppressing the background noise. Then it simultaneously guides the feature alignment among diverse views. Meanwhile, to alleviate heterogeneity between SR query images and HR gallery images, we design a dual-path teacher–student distillation network to extract features of LR and HR images separately. Moreover, the feature distribution between SR results and HR images is narrowed by feature distillation losses. Extensive results evidence the effectiveness of KGNet for cross-resolution vehicle Re-ID task. In some special real-world scenarios, the resolution of images may vary due to the uncertainty in camera positioning. Our SR-based approach is designed to recover LR images at fixed scaling factors. In the future, we will research exploring adaptive SR modules, as well as optimization strategies of multi-resolution prior knowledge, and improve the ability of approaches in complex scenarios, such as security surveillance, and autonomous driving.

CRediT authorship contribution statement

Aihua Zheng: Writing – original draft, Methodology, Conceptualization. **Longfei Zhang:** Validation, Investigation. **Weijun Zhang:** Visualization, Formal analysis. **Zi Wang:** Resources, Data curation. **Chenglong Li:** Supervision. **Xiaofei Sheng:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62372003), the Natural Science Foundation of Anhui Province, China (No. 2308085Y40 and No. 2208085J18), the University Synergy Innovation Program of Anhui Province (No. GXXT-2022-036), Anhui Provincial Key Research and Development Program (No. 202304a05020056), and MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China (Grant No. 2421004).

Data availability

The data that has been used is confidential.

References

Chao, D., Chen, C.L., Tang, X., 2016. Accelerating the super-resolution convolutional neural network. In: Proceedings of the European Conference on Computer Vision.. pp. 391–407.

Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., Ren, Z., Wang, Z., 2019a. ABD-net: Attentive but diverse person re-identification. In: Proceedings of the IEEE International Conference on Computer.. pp. 8351–8361.

Chen, Y.C., Li, Y.J., Du, X., Wang, Y.C.F., 2019b. Learning resolution-invariant deep representations for person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01. pp. 8215–8222.

Cheng, Z., Dong, Q., Gong, S., Zhu, X., 2020. Inter-task association critic for cross-resolution person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2602–2612.

Chouchane, A., Bessaoudi, M., Kheddar, H., Ouamane, A., Vieira, T., Hassaballah, M., 2024. Multilinear subspace learning for person re-identification based fusion of high order tensor features. Eng. Appl. Artif. Intell. 128, 107521.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations.

Fan, X., Jiang, W., Luo, H., Fei, M., 2019. SphereReID: Deep hypersphere manifold embedding for person re-identification. J. Vis. Commun. Image Represent. 60, 51–58.

Fu, X., Peng, J., Jiang, G., Wang, H., 2022. Learning latent features with local channel drop network for vehicle re-identification. Eng. Appl. Artif. Intell. 107, 104540.

Gao, S., Wang, J., Lu, H., Liu, Z., 2020. Pose-guided visible part matching for occluded person reid. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. pp. 11744–11752.

Han, K., Huang, Y., Chen, Z., Wang, L., Tan, T., 2020. Prediction and recovery for adaptive low-resolution person re-identification. In: Proceedings of the European Conference on Computer Vision. pp. 193–209.

Hao, L., 2019. Bags of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. pp. 1487–1495.

Haris, M., Shakhnarovich, G., Ukita, N., 2018. Deep back-projection networks for super-resolution. pp. 1664–1673, ArXiv.

He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W., 2021. TransReID: Transformer-based object re-identification. In: Proceedings of the IEEE International Conference on Computer. pp. 15013–15022.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 770–778.

Hu, W., Zhan, H., Shivakumara, P., Pal, U., Lu, Y., 2024. TANet: Text region attention learning for vehicle re-identification. Eng. Appl. Artif. Intell. 133, 108448.

Jia, D., Wei, D., Socher, R., Li, L.J., Kai, L., Li, F.F., 2009. ImageNet: A large-scale hierarchical image database. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 248–255.

Jiang, G., Pang, X., Tian, X., Zheng, Y., Meng, Q., 2023. Global reference attention network for vehicle re-identification. Appl. Intell. 53 (9), 11328–11343.

Jiao, J., Zheng, W.-S., Wu, A., Zhu, X., Gong, S., 2018. Deep low-resolution person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1.

Jie, H., Li, S., Gang, S., Albanie, S., 2020. Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (99), 2011–2023.

Jing, X.-Y., Zhu, X., Wu, F., You, X., Liu, Q., Yue, D., Hu, R., Xu, B., 2015. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. pp. 695–704.

Khorramshahi, P., Kumar, A., Peri, N., Rambhatla, S.S., Chen, J.C., Chellappa, R., 2019. A dual-path model with adaptive attention for vehicle re-identification. In: Proceedings of the IEEE International Conference on Computer.. pp. 6131–6140.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. Comput. Sci. 2602–2612.

Li, Y.-J., Chen, Y.-C., Lin, Y.-Y., Wang, Y.-C.F., 2020a. Cross-resolution adversarial dual network for person re-identification and beyond. arXiv preprint arXiv:2002.09274.

Li, K., Ding, Z., Li, K., Zhang, Y., Fu, Y., 2022. Vehicle and person re-identification with support neighbor loss. IEEE Trans. Neural Networks Learn. Syst. 33 (2), 826–838.

Li, S., Sun, L., Li, Q., 2023. Clip-reid: Exploiting vision-language model for image re-identification without concrete text labels. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, No. 1. pp. 1405–1413.

Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., Wu, W., 2019. Feedback network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. pp. 3862–3871.

Li, X., Zheng, W.-S., Wang, X., Xiang, T., Gong, S., 2015. Multi-scale learning for low-resolution person re-identification. In: Proceedings of the IEEE International Conference on Computer.. pp. 3765–3773.

Li, W., Zhou, K., Qi, L., Jiang, N., Lu, J., Jia, J., 2020b. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. Adv. Neural Inf. Process. Syst. 33, 20343–20355.

Liang, T., Jin, Y., Liu, W., Feng, S., Wang, T., Li, Y., 2022. Keypoint-guided modality-invariant discriminative learning for visible-infrared person re-identification. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 3965–3973.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144.

Lin, W., Li, Y., Yang, X., Peng, P., Xing, J., 2019. Multi-view learning for vehicle re-identification. In: 2019 IEEE International Conference on Multimedia and Expo. ICME, IEEE, pp. 832–837.

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

Liu, X., Liu, W., Mei, T., Ma, H., 2017. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. IEEE Trans. Multimed. 20 (3), 645–658.

Liu, X., Liu, W., Zheng, J., Yan, C., Mei, T., 2020. Beyond the parts: Learning multi-view cross-part correlation for vehicle re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 907–915.

Liu, H., Tian, Y., Wang, Y., Lu, P., Huang, T., 2016b. 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.

Lu, Z., Lin, R., Hu, H., 2022. MART: Mask-aware reasoning transformer for vehicle re-identification. IEEE Trans. Intell. Transp. Syst. 24 (2), 1994–2009.

Ma, C., Jiang, Z., Rao, Y., Lu, J., Zhou, J., 2020. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 5568–5577.

Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y., 2019. Pose-guided feature alignment for occluded person re-identification. In: Proceedings of the IEEE International Conference on Computer.. pp. 542–551.

Newell, A., Yang, K., Jia, D., 2016. Stacked hourglass networks for human pose estimation. In: Proceedings of the European Conference on Computer Vision.. pp. 483–499.

Porrello, A., Bergamini, L., Calderara, S., 2020. Robust re-identification by multiple views knowledge distillation. In: Proceedings of the European Conference on Computer Vision.. pp. 93–110.

Shen, F., Xie, Y., Zhu, J., Zhu, X., Zeng, H., 2023. Git: Graph interactive transformer for vehicle re-identification. IEEE Trans. Image Process. 32, 1039–1051.

Sun, L., Dong, J., Tang, J., Pan, J., 2023. Spatially-adaptive feature modulation for efficient image super-resolution. In: Proceedings of the IEEE International Conference on Computer.

Tai, Y., Yang, J., Liu, X., Xu, C., 2017. MemNet: A persistent memory network for image restoration. In: Proceedings of the IEEE International Conference on Computer. pp. 4539–4547.

Tang, Z., Naphade, M., Birchfield, S., Tremblay, J., Hodge, W., Kumar, R., Wang, S., Yang, X., 2020. PAMTRI: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. In: Proceedings of the IEEE International Conference on Computer.. pp. 211–220.

Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J., 2023a. Omni aggregation networks for lightweight image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. pp. 22378–22387.

Wang, Z., Hu, R., Yu, Y., Jiang, J., Liang, C., Wang, J., 2016. Scale-adaptive low-resolution person re-identification via learning a discriminating surface.. In: International Joint Conferences on Artificial Intelligence, Vol. 2. p. 6.

Wang, M., Ma, H., Huang, Y., 2023b. Information complementary attention-based multidimension feature learning for person re-identification. Eng. Appl. Artif. Intell. 123, 106348.

Wang, Z., Tang, L., Liu, X., Yao, Z., Wang, X., 2017. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In: Proceedings of the IEEE International Conference on Computer. pp. 379–387.

Wu, L.Y., Liu, L., Wang, Y., Zhang, Z., Boussaid, F., Bennamoun, M., Xie, X., 2023. Learning resolution-adaptive representations for cross-resolution person re-identification. IEEE Trans. Image Process..

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5987–5995.

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C., 2021. Deep learning for person re-identification: A survey and outlook. IEEE Trans. Pattern Anal. Mach. Intell. 44, 2872–2893.

Yu, X., Fernando, B., Ghanem, B., Porikli, F., Hartley, R., 2018. Face super-resolution guided by facial component heatmaps. In: Proceedings of the European Conference on Computer Vision.. pp. 93–110.

Zhang, W., Li, Z., Du, H., Tong, J., Liu, Z., 2024. Dual-stream feature fusion network for person re-identification. Eng. Appl. Artif. Intell. 131, 107888.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision. pp. 286–301.

Zhang, X., Zeng, H., Zhang, L., 2021. Edge-oriented convolution block for real-time super resolution on mobile devices. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 4034–4043.

Zheng, W.-S., Hong, J., Jiao, J., Wu, A., Zhu, X., Gong, S., Qin, J., Lai, J., 2022. Joint bilateral-resolution identity modeling for cross-resolution person re-identification. Int. J. Comput. Vis. 1–21.

Zheng, Z., Ruan, T., Wei, Y., Yang, Y., Mei, T., 2020. VehicleNet: Learning robust visual representation for vehicle re-identification. IEEE Trans. Multimed. 23, 2683–2693.

Zheng, W., Ye, M., Fan, Y., Xiang, B., Satoh, S., 2018a. Cascaded SR-GAN for scale-adaptive low resolution person re-identification. In: International Joint Conferences on Artificial Intelligence. pp. 3891–3897.

Zheng, Z., Zheng, L., Yang, Y., 2018b. A discriminatively learned CNN embedding for person re-identification. ACM Trans. Multimed. Comput. Commun. Appl. 14 (1), 176–193.

Zheng, A., Zhu, X., Ma, Z., Li, C., Tang, J., Ma, J., 2023. Cross-directional consistency network with adaptive layer normalization for multi-spectral vehicle re-identification and a high-quality benchmark. Inf. Fusion 100, 101901.

Zhou, Y., Li, Z., Guo, C.-L., Bai, S., Cheng, M.-M., Hou, Q., 2023. Srformer: Permuted self-attention for single image super-resolution. In: Proceedings of the IEEE International Conference on Computer. pp. 12780–12791.

Zhou, K., Yang, Y., Cavallaro, A., Xiang, T., 2020. Omni-scale feature learning for person re-identification. In: Proceedings of the IEEE International Conference on Computer.. pp. 3701–3711.

Zhu, X., Luo, Z., Fu, P., Ji, X., 2020. VOC-RelD: Vehicle re-identification based on vehicle-orientation-camera. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 2566–2573.