IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 28, 2019, accepted January 28, 2020, date of publication February 5, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971612

Part-Based Enhanced Super Resolution Network for Low-Resolution Person Re-Identification

YAN HA[1,3], JUNFENG TIAN[2,3], QIAOWEI MIAO[2], QI YANG[2], JIAAO GUO[2], AND RUOHUI JIANG[2]

[1]School of Management, Hebei University, Baoding 071002, China
[2]School of Cyber Security and Computer, Hebei University, Baoding, China
[3]Key Laboratory on High Trusted Information System in Hebei Province, School of Cyber Security and Computer, Hebei University, Baoding 071002, China

Corresponding author: Junfeng Tian (jf_tian@126.com)

This work was supported by the Natural Science Foundation of Hebei Province-Key Program under Grant F2016201244.

ABSTRACT Person re-identification (REID) is an important task in video surveillance and forensics applications. Many previous works often build models on the assumption that they have same resolution cross different camera views, while it is divorced from reality. To increase the adaptability of person REID models, this paper focuses on the low-resolution person REID task to relax the impractical assumption when traditional low-resolution person REID models are under pixel-to-pixel supervision in low and high resolution pedestrian image pairs. In addition, they are easily influenced by the global background, illumination or pose variations across camera views. Therefore, we propose a Part-based Enhanced Super Resolution (PESR) network by employing a part division strategy and an enhanced generative adversarial network to boost the unpaired pedestrian image super resolution process. Specifically, the part-based super resolution network transforms low resolution image in probe into high resolution without any pixel-to-pixel supervision and the part-based synthetic feature extractor module can learn discriminative pedestrian feature representation for the generated high resolution images, which employ a part feature connection loss as constraint to conduct matching for person re-identification. Furthermore, evaluations on four public person REID datasets demonstrate the advantages of our method over the state-of-the-art ones.

INDEX TERMS Low-resolution person re-identification, enhanced super resolution, part based, realistic discriminator.

I. INTRODUCTION

Person re-identification (REID) aims to identify a query person by searching for the most similar instances in gallery images or video sets, where the probe and gallery images are captured from overlapping cameras (cross-view). Generally, the REID precision can be improved by acquiring more information from a larger amount of surveillance data. However, there are many difficulties in actual deployments, such as the variances in illumination, occlusion, background, and alignment in real-world applications. In order to overcome these obstacles, existing methods typically address the person REID task by designing feature representation [1]–[3] or learning distance metrics [4]–[7].

Most of the proposed methods make an assumption that all pedestrian images have sufficiently high resolution captured by a variety of cameras [8], [9]. However, the distances between pedestrian and cameras are continuously changed along with the walking of pedestrians, which makes the large resolution discrepancy enormous enough (low-resolution) to bring a negative effect on cross-view pedestrian image matching. Furthermore, different camera settings are also inconsistent from city to village. We name this challenge in REID as low-resolution problem (Figure 1). However, existing REID models often ignore this low-resolution matching problem, which is required to be solved in recent researches.

Several researches [10]–[12] have paid attention to Low Resolution Person R-identification (LRPR) task. Jing *et al.* [10] proposed a semi-coupled low-rank discriminant dictionary learning approach to solve the LRPR task by learning a pair of High Resolution (HR) and Low Resolution (LR) dictionaries and a mapping from the features of HR gallery images and LR probe images, along with a discriminant term. Jiao *et al.* [11] developed a Super Resolution(SR) and identity joint learning approach to optimize image SR and REID matching simultaneously. Wang *et al.* [12] proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca.

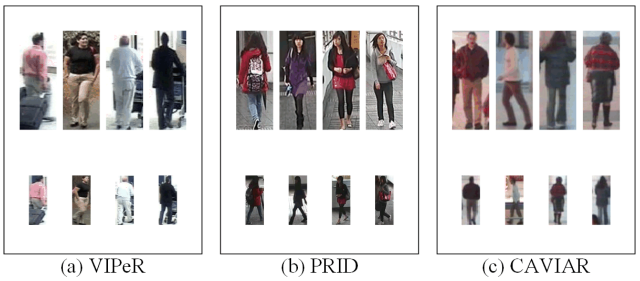


FIGURE 1. Example image pairs from three datasets. Each column shows two images of the same identity from two different cameras with different resolutions, where images in the bottom row are LR. (a) the VIPeR dataset; (b) the CAVIAR dataset; (c) the Market-1501 dataset.

a cascaded super resolution generative adversarial network to address the LRPR task. Nevertheless, these models only addressed the low-resolution person REID problems while ignored the cross-view aspect.

Therefore, some low-resolution matching methods [13], [14] are proposed recently, to tackle the cross-view aspect in Low Resolution Person Re-identification (LRPR) task. Wang *et al.* [13] combined effective embeddings based on multiple convolutional neural network layers, trained with deep-supervision. Chen *et al.* [14] proposed a novel network architecture of resolution adaptation and a REID network to solve low-resolution person REID. **The existing low-resolution matching models produce the LR images by down-sampling the original HR images, which are directly employed as pixel-to-pixel supervisions.** That limits the scalability of application in the real world, since there are no correlated HR images in low-resolution person REID to support pixel-to-pixel supervision, which is named as Directly Low Resolution Person Re-identification (DLRPR). Besides, exiting LRPR models are conducting SR on global images, instead of parts of each image, leading to much more noisy information generated from the background, pose, and illumination variations. Beyond them, some image style transfer based person REID models [15] can be employed into DLRPR task. Liu *et al.* [15] utilized style transformations to make the domain image styles portable. In terms of migration, they added GAN for resolution control to ensure that the resolution of the data collection can be unified when it is converted to other styles. But it only achieves image style transforming when is employed into DLRPR task, and can not ensure to improve the pedestrian image resolution. Therefore, existing person REID models have not worked out the DLRPR problem.

Inspired by the success of deep neural network in REID [16], [17], we propose a novel framework of Part-based Enhanced Super Resolution (PESR) network to conduct matching between low resolution images and high resolution images without any pixel-pixel supervision. The designed PESR network is presented on each cropped pedestrian part to preserve much more local information and generate discriminative part-based pedestrian features for generated and original SR images by the identical loss. Furthermore, PESR network integrates a realistic discriminator loss into a synthetic generative adversarial neural network and introduce the part feature connection (PFC) loss to enhance the discriminative ability for the extracted features.

A. MOTIVATION AND CONTRIBUTION

Though low resolution person REID is an important application, it has not been well investigated in the situation of lacking pixel-to-pixel HR supervision in practice (named Directly Low Resolution Person Re-identification, DLRPR). To better illustrate the DLRPR problem and our PESR network, we summarize the motivation and contributions below.

MOTIVATIONS

In order to achieve better matching performance for DLRPR task, the following points needs to be improved.

(1) Traditional low resolution person REID models employ down-sampling strategy to obtain LR image from its correlated HR image as a pixel-to-pixel supervision in the same view. However, they can not handle the cross-view DLRPR task without the supervision of original HR image in many realities.

(2) Existing low resolution person REID methods conduct super resolution on global pedestrian images, leading to contain more noisy information produced by global variations of background, illumination, or occlusions. That will influence the image reconstruction and discriminative ability of the feature.

(3) Traditional LRPR models often deploy generic purpose super resolution model to obtain HR images. They utilize GAN architecture to produce HR images by using the discriminator to identify each image ''whether the generated image is in HR'', which can not measure the image quality accurately.

CONTRIBUTIONS

To overcome the weaknesses mentioned in motivations, we conclude the contributions of our PESR network as three points.

(1) To achieve the cross-view DLRPR task, we propose an enhanced super resolution mechanism, by employing an enhanced generative adversarial network architecture. This GAN framework can generate images from different views to relax the pixel-to-pixel supervision in conventional LRPR methods.

(2) In order to avoid the influence of global factors and restore more local information in super resolution, we introduce the part-based feature extracting strategy both in GAN and identification module. That can produce part-based HR images and discriminative features, so as to preserve sufficient pedestrian information in the super resolution procedure.

(3) To conduct super resolution in low resolution person REID task, we utilize a realistic discriminator loss function to judge ''whether one image is more realistic than

the other images'' rather than ''whether one image is real or fake'', which can strengthen the super resolution effectiveness according to the quality of high resolution images. Furthermore, we employ a part feature connection (PFC) loss to restore beneficial detail for pedestrian discriminative matching features.

II. RELATED WORK

A. IMAGE BASED PERSON RE-IDENTIFICATION

Person REID attracts much more research interest by the development of deep learning technology in recent years. The related models are classified as global-based [8], [9], and part-based feature representations [18]–[20].

The popular solution for global-based feature representation REID models is to constitute image pair or triplet into a global convolutional neural network. In earlier studies, a quadruple loss was designed to identify people according to global features. And then, Borgia *et al.* [9] proposed a loss supervision to generate a more expanded global feature space of compact classes where the overall level of the inter-identities' interference was reduced. In the recent work [8], Zhou et al proposed a decoder network which was used to reconstruct the binary mask by using a novel local regression loss function to solve occlusion problems. However, the global-based person REID models are easily influenced by the background, illumination or occlusions. Therefore, various part-based REID models are introduced to solve this problem. In an early stage, Zhu *et al.* [18] proposed a part-based deep hashing method, in which batches of triplet samples were employed as the input of the deep hashing architecture. Each triplet samples contained two pedestrian images (or parts) with a same identity and one pedestrian image (or part) of the different identity. A triplet loss function was employed with a constraint that the Hamming distance of pedestrian images (or parts) with the same identity was smaller than ones with the different identity. In work [19], they used two CNN models to extract features from different parts of persons separately and then combine them together to construct the overall features of the input images. Finally, metric was utilized to rank the distances between the overall features of the different persons. In a recent work [20], they resorted the global and local person attributes to build global and local representation, respectively. This part-based model achieved better performance than global-based REID models, by removing the effects of global factors. Besides, the research [21] learned view-invariant subspace for person REID, and its corresponding similarity metric using an adversarial view adaption approach. As for another person REID task of video based person REID [22], [23], Wu *et al.* [24] presented the global deep video representation learning to view-based person re-identification that aggregates local 3D features across the entire video extent. The work [25] proposed a novel few-shot deep learning approahch to video-based person REID, to learn comparable representations that are discriminative and view-invariant. The research [26] designed a visual-appearance-level and spatial-temporal-level dictionary learning approach for video-based person REID task.

Inspired by the success of part-based (local) REID methods [19], [20], [24], we introduce the part division strategy into DLRPR task, which integrates pedestrian part images into GAN and feature extractor, to restore more local information in SR procedure and produce a more accurate HR images. Furthermore, we also extract part-based feature representation to conduct cross-view matching process. Different from these person REID methods above, our PESR approach can learn part-based pedestrian feature representations not only for normal pedestrian images, but also for low resolution images through the designed part-based enhanced super resolution network.

B. LOW RESOLUTION PERSON RE-IDENTIFICATION

Traditional image based person REID methods assume that all images are in high resolution, but a large variation in cross-view resolutions might occur in real applications. Various REID models have proposed to focus on this low resolution task, including [11], [27]–[29]. A common solution is to reduce the impact of different resolutions on identification task by learning the cross-mapping between LR and HR images. The approach [11] is instantiated by designing a hybrid deep convolutional neural network for improving low resolution re-id performance and utilizing an adaptive fusion algorithm for accommodating multi-resolution LR images. To reduce the influence of low resolution on the distance learning, Ma *et al.* [27] designed a clustering-based semi-coupled mapping term, which can reduce the variation between features of LR and HR videos by a semi-coupled mapping matrix. Another popular solution is to elevate LR images to a uniform HR [10], [11], [30], followed by identification tasks. In work [28], Zhuang et al cascaded multiple SRGANs in series and plug-in a REID network to get the consistent HR images. Mao *et al.* [31] proposed a similar model, which learned similar resolution-invariant image representations and was able to recover the missing details in LR input images that is beneficial to improve person re-ID performance. These models demonstrate that integrating super resolution modules into REID model can improve the performance in low resolution person re-identification problem.

However, these low resolution models do not consider the real application with lacking pixel-to-pixel high resolution supervision in cross-view pedestrian image matching task, and they can not handle the Directly Low Resolution Person Re-identification task.

C. SUPER RESOLUTION TECHNOLOGY

Image super resolution is an important research topic in computer vision. The generic super resolution models have made significant progress by using the deep learning technology [32]–[34]. Dong *et al.* [32] proposed an end-to-end way to learn the correlation between LR and HR images, and achieved excellent results. Numerous network structures have been employed, such as residual blocks [35], densely

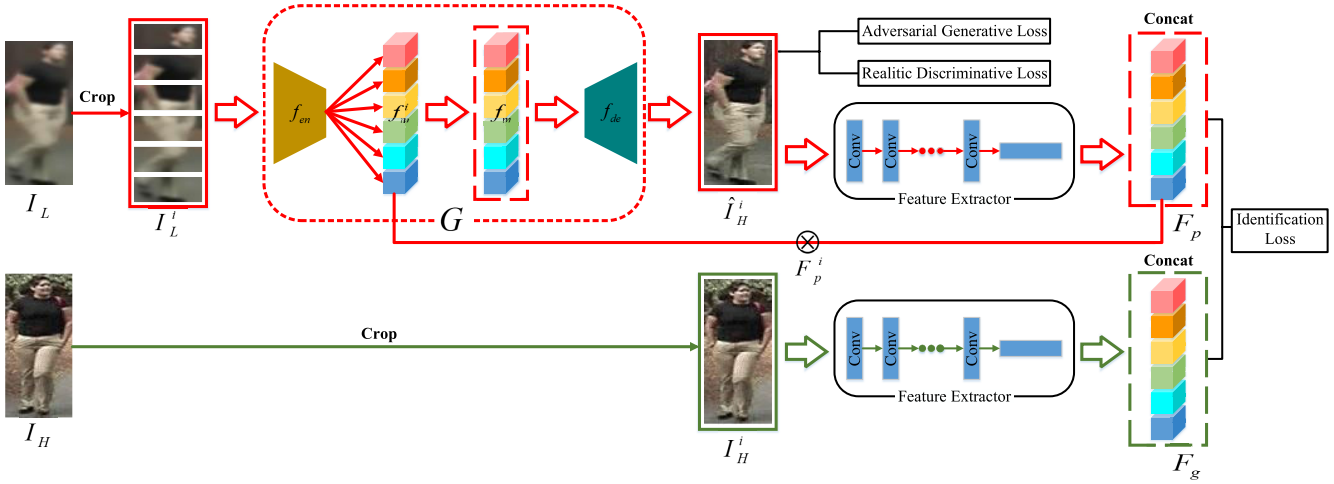


FIGURE 2. Firstly, we input the LR image I_L into clipper f_{cl} to generate image part I_L^i, which the feature map F_m^i can be obtained by the encoder f_{en}. Then, the concatenation F_m of the part feature maps are utilized to generate high resolution image $\hat{I}_H$ by the decoder f_{de}. Finally, the generated high resolution image and original high resolution image are fed into the PCB network to obtain fused synthetic features F_p and F_g by Eq.14, which are employed to calculate identification loss.

connected network [36], deep back projection [37] and residual dense network [38]. Lim *et al.* [33] removed the unnecessary batch normalization layers of the residual blocks and expanded the size of the model to improve the quality of generated images. In addition, other researchers further explored a deeper network with channel attention [39] which generates a better performance of PSNR. Beyond them, unsupervised learning or reinforcement learning methods [40], [41] also have been introduced to solve the SR problems. However, the hallucinated details of images that have been processed by the SRGAN [35] are often accompanied with unpleasant artifacts. To further enhance the visual quality, we introduce a network structure similar to ESRGAN [34], which can effectively enhance the restoration of image details by using Residual-in-Residual Dense Block.

When the generic super resolution models are utilized into a directly low resolution person re-identification task, there is still a main challenge that needs to be solved. That is, the models aiming to improve the perception of human eyes, instead of restoring discriminative detail for person REID, which will retain noisy information after super resolution. Therefore, we integrate the realistic discriminator and part feature connection loss into our PESR network, which can effectively increase the identification information of low resolution images and make the super resolution module positive for cross-view matching.

III. OUR APPROACH

A. OVERVIEW

To solve the directly low resolution person re-identification problem, we propose a Part-based Enhanced Super Resolution (PESR) Network, which consists of two main components, including the Part-based Super Resolution (PSR) network, and the Synthetic Feature Extractor (SFE) module. PSR network transforms LR images in probe view into the gallery view with high resolution, without any pixel-to-pixel supervision. PSR also conduct generation under part division for the whole pedestrian images, which can restore more local information for identification purpose. By using the obtained HR images, the SFE module can learn synthetic part-based feature representations for them to be half of the probe LR images, and it also acquires the gallery features from HR images. Therefore, PESR can bridge the resolution gap between images in low resolution pedestrian images, and transform them into a shared feature space to conduct pedestrian matching process by integrating the PSR and SFE modules. The detailed illustration is in Figure 2.

B. PROBLEM DEFINITION

We define a low resolution image without its original HR image in probe as I_L in Directly Low Resolution Person Re-identification (DLRPR) task. Its correlated LR image in Gallery is I_H, with the same identity label with I_L. In Figure 2, a pedestrian image clipper f_{cl} is employed to produce cropped pedestrian image parts,

$$f_{cl}(I_L) = \{I_L^1, \cdots, I_L^i, \cdots, I_L^n\} \quad (1)$$

$$f_{cl}(I_H) = \{I_H^1, \cdots, I_H^i, \cdots, I_H^n\} \quad (2)$$

where I_L^i denotes the i-th part of LR image, I_H^i is the i-th part in HR image, n is the number of image parts.

The image generator G is devised to generate a HR image $\hat{I}^i{}_H$ from I_L^i respectively, along with a discriminator D to conduct adversarial training. For the generated HR image $\hat{I}^i{}_H$ and gallery HR image I_H^i, we employ a backbone convolutional neural network as synthetic features extractor $f_s(\cdot;\theta)$ to learn discriminative feature for each part,

$$\hat{F}_H^i = f_s(\hat{I}_H^i;\theta) \quad (3)$$

$$F_H^i = f_s(I_H^i;\theta) \quad (4)$$

TABLE 1. Frequently-used notions.

Symbol	Definition
I_L	LR pedestrian image in Probe.
I_H	HR pedestrian image in Gallery.
$f_{cl}(\cdot)$	Image clipper.
I_L^i	i-th part of LR image I_L through $f_{cl}(I_L)$.
I_H^i	i-th part of HR image I_H through $f_{cl}(I_H)$.
G	Super resolution generator for I_L^i.
$\hat{I}_H^i$	Generated HR image for $I_{L_s}^i$.
D	Realistic discriminator for $\hat{I}_H^i$ and I_H^i.
$f_s(\cdot;\theta)$	Synthetic feature extractor for $\hat{I}_H^i$ and I_H^i.
θ	The parameter of the backbone network of $f_s(\cdot;\theta)$.

where θ is the parameter of the backbone covolutional neural network.

The annotations referred above are summarized into Table 1.

C. PART-BASED ENHANCED SUPER RESOLUTION NETWORK

1) IMAGE CLIPPER

Inspired by the success of part-based feature representation for REID, we employ the part division strategy on LR images to conduct the super resolution process. Besides, the part based representation is also deployed into feature extraction both for generated HR images and original HR images. For the part division $f_{cl}(\cdot)$ strategy, we follow the partition method in work [16].

2) PART-BASED SUPER RESOLUTION NETWORK

In the part-based approach, the resolution quality of the image has an important impact on the identification task. In order to make the segmented local images aligned in resolution, the image quality can be improved after super resolution processing, and restore some personal information that is hidden in low resolution image. The cropped images $\{I_L^1, \cdots, I_L^i, \cdots, I_L^n\} = f_{cl}(I_L)$ and $\{I_H^1, \cdots, I_H^i, \cdots, I_H^n\} = f_{cl}(I_H)$ is consequently obtained by using the image clipper $f_{cl}(\cdot)$ described in previous section. To achieve the pedestrian image super resolution, we design a generator $G = \{f_{en}, f_{de}\}$, where f_{en} is the encoder that is used to extract feature maps from the input images, and f_{de} is the decoder to reconstruct images from the feature maps (detail architecture following ESRGAN [34]). The feature map F_m^i for the i-th part image I_L^i can be calculated by using Eq.5,

$$F_m^i = f_{en}(I_L^i) \tag{5}$$

All part feature maps obtained by Eq.5 can be concatenated as the synthetic feature map F_m of LR image I_L by $F_m = [F_m^1, \cdots, F_m^i, \cdots, F_m^n]$. Therefore, the generated HR image $\hat{I}_H$ can be computed by using Eq.6,

$$\hat{I}_H = f_{de}(F_m) \tag{6}$$

In order to train the PSR module effectively, we attach a discriminator D to the generated and the original gallery HR images, following the VGG-19 network pretrained on ImageNet [42] (removing the probability prediction layer in adversarial learning). Different from the traditional generative adversarial networks, we employ the realistic discriminative loss in our PSR generative adversarial network to ensure the generated high resolution images are more accurate than existing realistic HR images in gallery. The realistic discriminator is utilized to achieve the directly low resolution pedestrian image generation. Specifically, the standard discriminator D_s in GAN tries to identify the image by using Eq.7,

$$D_s\left(I_H\right) = \sigma\left(f_d\left(I_H\right)\right) \to 1 \tag{7}$$

$$D_s\left(\hat{I}_H\right) = \sigma\left(f_d\left(\hat{I}_H\right)\right) \to 0 \tag{8}$$

where $D_s(\cdot)$ is the output of discriminator to identify whether the input is a generated images, $f_d(\cdot)$ is the discriminative feature extractor, and the σ is the sigmoid function. This adversarial loss only judges "whether the input image is the real or fake", which can not handle the cross-view and low-resolution problems in DLRPR task. Therefore, we replace D_s by our realistic discriminator D, which tries to predict the probability that the real gallery HR image I_H is more realistic than the generated HR image $\hat{I}_H$ from I_L,

$$D\left(I_H, \hat{I}_H\right) = \sigma\left(f_d\left(I_H\right) - \mathbb{E}_{\hat{I}_H}\left[f_d\left(\hat{I}_H\right)\right]\right) \to 1 \tag{9}$$

$$D\left(\hat{I}_H, I_H\right) = \sigma\left(f_d\left(\hat{I}_H\right) - \mathbb{E}_{I_H}\left[f_d\left(I_H\right)\right]\right) \to 0 \tag{10}$$

where $\mathbb{E}_{\hat{I}_H}[\cdot]$ denotes the average operator of all generated HR images in the mini-batch, and $\mathbb{E}_{I_H}[\cdot]$ is the average operator of all real HR images in the mini-batch.

Based on the analysis above, the realistic discriminator loss function can be concluded as,

$$\begin{aligned} L_D = &-\mathbb{E}_{I_H}\left[\log\left(D\left(I_H, \hat{I}_H\right)\right)\right] \\ &-\mathbb{E}_{\hat{I}_H}\left[\log\left(1 - D\left(\hat{I}_H, I_H\right)\right)\right] \end{aligned} \tag{11}$$

On the contrary, the adversarial loss for the HR image generator is,

$$\begin{aligned} L_G = &-\mathbb{E}_{I_H}\left[\log\left(1 - D\left(I_H, \hat{I}_H\right)\right)\right] \\ &-\mathbb{E}_{\hat{I}_H}\left[\log\left(D\left(I_H, \hat{I}_H\right)\right)\right] \end{aligned} \tag{12}$$

Therefore, the model is jointly trained on the samples from I_H and $\hat{I}_H$, by minimizing the loss functions of discriminator and generator simultaneously. Hence, the optimization of our SR module can be boosted by the gradients from the generated HR images and real HR images, along with the adversarial learning procedure.

Since the real HR images in gallery is not pixel-to-pixel connected, the reconstructed images may lose important information, such as discriminative pedestrian features. The reconstruction loss in conventional SR do not employed in this network. However, the ultimate purpose of our PESR network is to match correct pedestrian HR image by giving a LR probe image, which requires maintaining the discriminative information in the HR image generation. Therefore,

we propose a Part Feature Connection (PFC) loss to achieve this goal, rather than the pixel-to-pixel supervision.

$$L_{pfc} = \sum_{i=1}^{n} \|f_p(I_H^i) - f_p(\hat{I}_H^i)\|^2 \tag{13}$$

where $f_p(\cdot)$ is the part feature extractor by a fine-tuned VGG network which modifies the last fully connected layer as 512-dimensions, and $\hat{I}_H^i$ is i-th part for $\hat{I}_H$ after the image clipper $f_{cl}(\hat{I}_H)$.

By utilizing the PFC loss, the PESR network not only resolves the low resolution image generation task, but also retain the discriminative information by the part-based accurate representation.

3) PEDESTRIAN FEATURE REPRESENTATION

After obtained the generated HR images for LR image in probe, the main task is to enhance the discriminative feature representative ability in super re-solution network. Therefore, we devise an effective feature representation approach to learn the similarity metric between $\hat{I}_H$ and I_H. For feature learning, we adopt a Part-based Convolutional Baseline (PCB) architecture [16] as the feature learning method, named as $f_{pcb}(\cdot)$, and an identical loss function on the learned PCB features to train the whole network.

PCB network $f_{pcb}(\cdot)$ outputs a convolutional descriptor consisting of several part-level features for a given image input. Detailed architecture can be found in work [16]. We remove its last FC layer to predict the ID predictions, and modify the output as a fixed dimensional feature vector for each part, as same as the part feature extractor $f_p(\cdot)$ in super resolution network. For the final feature representation of the input images, we can calculate part feature vector for the image I_L in probe by using Eq.14,

$$F_p^i = \{f_{pcb}(\hat{I_H})\}^i \otimes f_p(I_L^i) \tag{14}$$

where $\{f_{pcb}(\hat{I_H})\}^i$ is the PCB-based feature vector for i-th part of the input images. Unlike existing LRPR models, we fusion the original feature vector of LR image with the PCB features to avoid missing discriminative information in the SR process. The synthetic feature representation of the probe image is $F_p = [F_p^1, \cdot, F_p^i, \cdot, F_p^n]$. As for the feature of image in gallery $F_g = [F_g^1, \cdot, F_g^i, \cdot, F_g^n]$, we utilize the original PCB feature $F_g^i = f_{pcb}(I_H)$ to conduct the matching process.

Then, we constraint the feature representations F_g and F_p by the identical loss,

$$\begin{aligned} L_{ide}(F_p, F_g) = (1-l)\{\max(0, \alpha - \|F_p - F_g\|)\}^2 \\ + l\|F_p - F_g\|^2 \end{aligned} \tag{15}$$

where α is a margin parameter which constrain the dispersion in the feature space, and $l = 1$ when F_p and F_g are from the same pedestrian identity, otherwise $l = 0$.

Thus, the final loss function for the part-based enhanced SR network can be rewritten as,

$$L_{pesr} = L_{ide} + \lambda L_G + \gamma L_{pfc} \tag{16}$$

Algorithm 1 Part-based enhanced super resolution network.

Require:
LR image I_L, and HR image I_H.
Parameters $\lambda = 0.4$, $\gamma = 0.6$, $\alpha = 1.5$.
Learning rate=$1e-4$.
Down-sampling scale=$[\frac{1}{2}, \frac{1}{3}, \frac{1}{4}]$.
Network Initialization.

Ensure:
1: Divide I_L into six equal parts, and represent each part as I_L^i; Divide I_H into six equal parts, and represent each part as I_H^i.
2: Pass each part I_L^i to generator G to obtain image $\hat{I}_H$.
3: Pass $\hat{I}_H$ as input into the PCB net f_{pcb} and obtain part feature vector F_p^i by Eq. 14. Then calculate the final feature F_p; Pass I_H^i as input into PCB net to obtain F_g.
4: Optimize the network by the gradients of L_{pesr}
5: **return** Network parameters;

where λ and γ are balance weights of the identification loss, generator loss, and PFC loss. When we train the network by the L_{pesr} loss, the adversarial loss of the discriminator is embedded into the optimization. Detailed algorithm is illustrated in Algorithm 1.

IV. EXPERIMENTS

A. DATASETS AND CONFIGURATIONS

Datasets: As shown in research [43], VIPeR [44], CAVIAR [45], Market1501 [46], DukeTMTC-ReID [47], are four commonly used datasets for the LRPR REID task. Therefore, we use these four datasets to train and evaluate our proposed PESR network.

VIPeR [44]: The VIPeR dataset package is widely used and contains 1,264 outdoor images taken from two different views of 632 people. We normalize all individual images to a size of 128 × 48 pixels. Following SING [11], we randomly divide this dataset into two non-overlapping halves based on the identity labels. Namely, images of a subject belong to either the training set or the test set.

CAVIAR [45]: The CAVIAR dataset consists of 72 individual images captured from two cameras at the center of shopping. Among 72 identities, 50 of them have images from two camera views and the rest 22 only from one camera images. Each identity is carefully selected to maximize the resolution variance. The images in these two cameras have a large variance in resolution. In our experiment, the 50 identities captured from two views are utilized to train the network and we split them into 2 sets with equal number for training and testing.

Market-1501 [46]: The Market-1501 dataset is collected in front of a supermarket in Tsinghua University. A total of six cameras are used, including 5 HR cameras, and one LR camera. Overlap exists among different cameras. The overall dataset contains 32,668 annotated bounding boxes

TABLE 2. Comparing with state-of-the-art low resolution person re-identification methods on VIPeR, CAVIAR, Market-1501 and DukeMTMC-reID. Bold and underlined numbers indicate the top results.

Dataset	VIPeR			CAVIAR			Market-1501			DukeMTMC-reID		
Model	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10	r-1	r-5	r-10
JUDEA [30]	26.0	55.1	69.2	22.0	60.1	80.8	-	-	-	-	-	-
SLD^2L [10]	20.3	44.0	62.0	18.4	44.8	61.2	-	-	-	-	-	-
SDF [43]	9.3	38.1	52.4	14.3	37.5	62.5	-	-	-	-	-	-
SING [11]	33.5	57.0	66.5	33.5	72.7	89.0	74.4	87.8	91.6	65.2	80.1	84.8
CSR-GAN [12]	37.2	62.3	71.6	34.7	72.5	87.4	76.4	88.5	91.9	67.6	81.4	85.1
CRGAN [29]	43.1	68.2	77.5	42.8	76.2	91.5	83.7	92.7	95.8	75.6	86.7	89.6
PESR	**47.3**	**74.1**	**81.6**	**45.1**	**83.2**	**93.7**	**85.6**	**94.8**	**97.5**	**79.4**	**91.3**	**92.1**

of 1,501 identities. In this open system, images of each identity are captured by at most six cameras.

DukeMTMC-ReID [47]: The DukeMTMC dataset is a large-scale tagged multi-objective multi-camera pedestrian tracking dataset. It provides a new large video dataset recorded by 8 synchronous cameras, with more than 7,000 single-camera tracks and more than 2,700 independent characters. DukeMTMC-ReID is the pedestrian re-recognition subset of the DukeMTMC dataset. The DukeMTMC-reID dataset contains 36, 411 images of 1, 404 identities captured by 8 cameras. We adopt the benchmarking 702/702 training/testing identity split in our experiments.

Measurement: To evaluate the PESR network, we adopt rank-n accuracy and Cumulative Match Characteristic (CMC) curves to measure the performance. Rank-n accuracy denotes whether the samples in top-n matching similarities contains the correct matching identity. CMC curves are often employed to measure REID models, which shows the overall identification performance and the value of CMC indicates the percentage of the real matching ranked in the top-n. Moreover, we also report the mean Average Precision (mAP) of the propose PESR network, which is often employed to evaluate the performance of person re-identification model. To evaluate the quality of generated high resolution images, we introduce the Mean Opinion Score (MOS) to measure this points. MOS is a numerical measure of the human-judged overall quality of an event or experience, which most often judge on a scale of 1(bad) to 5 (excellent). It is the average of a number of other human-scored individual parameters.

B. IMPLEMENTATION DETAIL

Our network is implemented by PyTorch framework, and employs two NVIDIA GEFORCE 2080TI GPUs to train PESR network. First of all, we unify all the images into 256×128 dimensions before training, and set batch size $= 16$, $\alpha = 1.5$, the learning rate $= 1e-4$, and the Leaky Relu is selected as the activation function. Moreover, Adam optimizer is employed to optimize the network. For these four datasets, we select one camera view as the probe with three down-sampling scales $\left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}\right\}$ to produce the LR images, and test them separately when we train and test the network. Then, we choose the original images in another camera view as the gallery to provide HR pedestrian images. In the test stage, an evaluated pedestrian LR image is input into the network for searching the correct pedestrian in the gallery, without fixed input resolution constraints. Therefore, we test our network for three LR level pedestrian images to validate the effectiveness of our PESR network and report the average results of them.

C. EVALUATION ON REID PERFORMANCE

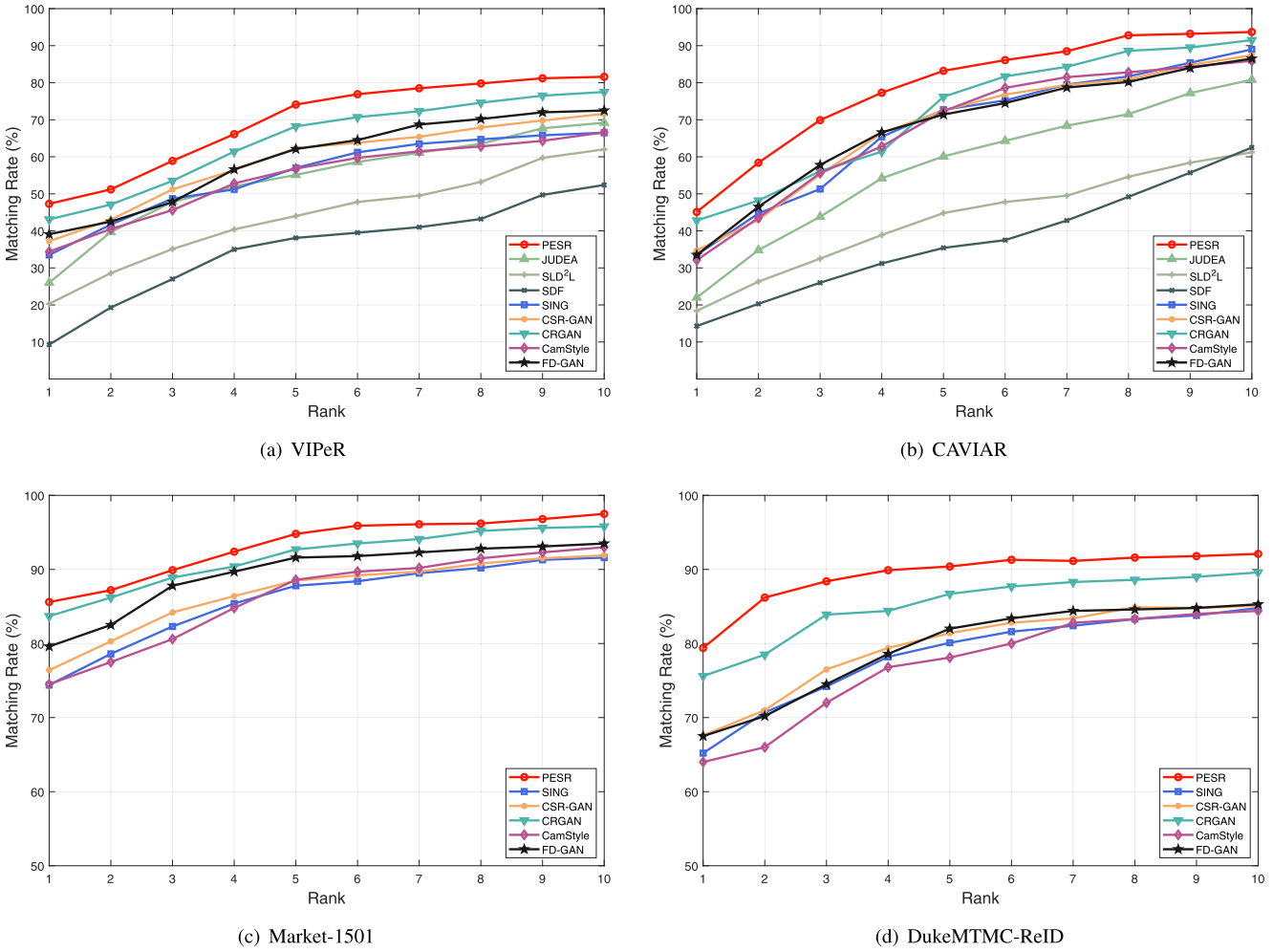
In this subsection, we measure the REID performance of our PESR network on the DLRPR scenario, and compare our method with several advanced LRPR models, then conclude the competitiveness of our network.

1) COMPARED METHODS

We employ six compared LR person REID models approaches [10]–[12], [29], [30], [43]. In detail, JUDEA [30]: This work proposed a heterogeneous mean deviation criterion for cross-scale image domain alignment and applied it to finish REID task. SLD^2L [10]:This method constructed a dictionary with a feature mapping relationship between HR gallery images and LR probe images. SDF [43]:They proposed a method that use a discriminant surface which can distinguish feasible and infeasible function in scale distance function space to finish person REID work. SING [11]: This approach was instantiated by designing a hybrid deep Convolutional Neural Network for improving low resolution REID performance and combined an adaptive fusion algorithm for accommodating multi-resolution LR images. CSR-GAN [12]:This work cascaded multiple SRGANs in series and plug-in a REID network. CRGAN [29]: They proposed a model that could restore image details while maintaining the original resolution. All the compared methods follow their setting in their papers.

2) RESULTS

The results of our PESR network are summarized in Table 2, and it achieves achieves rank-1 accuracies of 47.3% on



(a) VIPeR (b) CAVIAR

(c) Market-1501 (d) DukeMTMC-ReID

FIGURE 3. Experimental results on three datasets for REID problem. For each dataset, we compare the proposed method (PESR) with JUDEA [30], SLD^{2}L [10], SDF [43], SING [11], CSR-GAN [12], CRGAN [29].

TABLE 3. mAP performance of our proposed PESR network.

Dataset	VIPeR	CAVIAR	Market-1501	DukeMTMC-reID
Rank-1	47.3	45.1	85.6	79.4
mAP	27.8	26.6	79.2	55.2

VIPeR, 45.1% on CAVIARA, 85.6% on Market-1501, and 79.4% on on DukeMTMC-reID dataset. Furthermore, the CMC curves of PESR are also illustrated in Figure 3. From Table 2 and Figure 3, the rank-n accuracies of PESR network also has competitiveness in realistic applications. In addition, we calculate the mAP of our PESR network and summarize the rank-1 and mAP in Table 3, which achieves 27.8% on VIPeR, 26.6% on CAVIAR, 79.2% on Market-1501, and 55.2% on DukeMTMC-reID.

From the results of compared methods, our PESR network obtains the advantage matching performance than them, and outperform them at least 1.9%(85.6%-83.7%). Moreover, both of rank-n results and CMC curves also demonstrate that the performance of PESR in different ranks are higher than all the existing LRPR approaches. It shows the competitiveness of PESR over existing LRPR models.

D. EVALUATION ON SUPER RESOLUTION PERFORMANCE

To illustrate the improvement on super resolution in REID task, we show the generated high resolution image samples in Figure 4, together with a few traditional methods.

1) BASELINES

In this subsection, we aims to prove the PESR is applicable for low-resolution image generation. In Figure 4, we show some low resolution images that come from different datasets. It is important to note that LR images are not sampled at an uniform scale, and they are constructed by down sampling the original images at scales from $\left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}\right\}$ to simulate the low-resolution situation in reality. The super resolution results are generated respectively by bicubic, PESR without PFC loss, and PESR with PFC loss. Among them, bicubic interpolation is a basic method to increase the image resolution. In image processing, bicubic interpolation is often

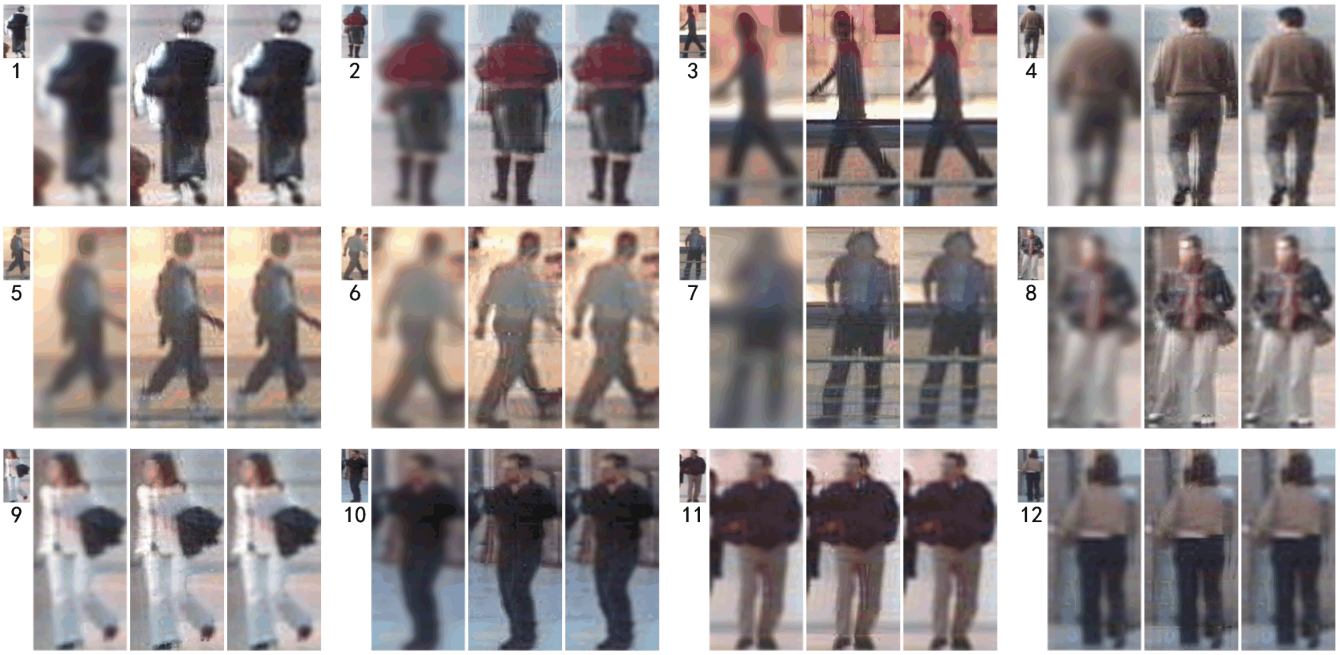


FIGURE 4. Some subjective results by super resolution. We show 12 groups of results for low resolution images. Image 1-3 are selected from the VIPeR dataset, image 4-6 are selected from the CAVIAR dataset, image 7-9 are selected from the Market-1501 dataset, and image 9-12 are selected from the Duke-MTMC dataset. For each group, four images are LR image, a SR image by bicubic, PESR without PFC loss, and original PESR method.

chosen over bilinear or nearest-neighbor interpolation in image re-sampling. In contrast to bilinear interpolation, which only takes 4 pixels (2×2) into account, bicubic interpolation considers 16 pixels (4×4). Therefore, images re-sampled by bicubic interpolation are smoother and have fewer interpolation artifacts.

Meanwhile, we perform a Mean Opinion Score (MOS) test to quantify the ability of low-resolution image generation. Specifically, we ask 5 raters to assign an integral score from 1 (bad quality) to 5 (excellent quality) to the super resolution images. Each rater rates all the testing images of four datasets. We find no significant differences between the ratings of the identical images. The experimental results of the conducted MOS tests are summarized in Table 4. It can be seen that the proposed method outperforms the general SR methods (nearest and bicubic) in all the resolution groups. It should be mentioned that SR module and REID module are integrated in PESR, which are not independent. Isolated super resolution networks do not guarantee that it only restore the detail in images but would be benefit REID task, and most of the other low resolution REID methods aim to separate the SR module from the REID module. In other words, LR images are generated into HR images by SR model alone, and then REID network is trained by generated HR images independently in conventional super resolution person re-identification models.

TABLE 4. The MOS test results on the testing images from four different datasets, and compare the proposed PESR method with the nearest and the bicubic methods.

Dataset	nearest	bicubic	PESR
VIPeR	1.05	1.12	**1.98**
CAVIAR	3.10	3.25	**4.20**
Market-1501	2.11	2.14	**2.98**
DukeMTMC-ReID	2.54	2.72	**3.15**

2) PERFORMANCE

To verify that PESR works well for different ranges of datasets, we select images from four datasets to show the super resolution performance in Figure 4, by bicubic, PESR without PFC loss,and original PESR network. It is obvious that PESR greatly improves the image resolution, and effectively strengthens the details of the pedestrian, which can effectively enhance the accuracy of the network for REID. In terms of image restoration alone, the images after PESR have better human eye perception. Through the comparison in figure 4, we can find that PESR without the use of PFC loss can effectively process low-resolution image generation, but there is still a certain degree of incoordination in details. These incongruous details not only affect the smoothness of the image, but also may cause difficulties in image recognition. With the combination of PFC loss, the resolution of the image is well improved, while the edges of the human image become smoother and the generated images are more natural. At the same time, we can notice a problem with bicubic method, which can only improve the image's fixed effect, and cannot deal with low-resolution re-identification problem flexibly. Thus, PESR network can unify the generated images in resolution variations as far as possible, and effectively

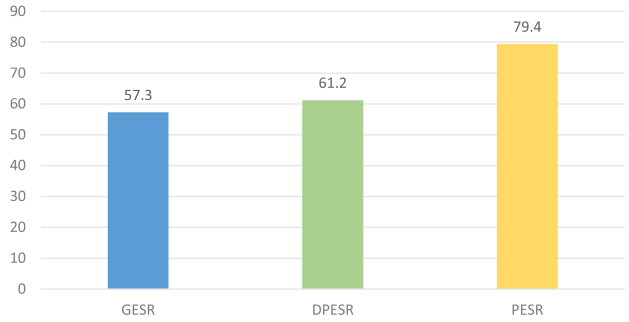


FIGURE 5. Rank-1 accuracy (%) between GESR, DPESR and PESR on DukeMTMC-reID dataset.

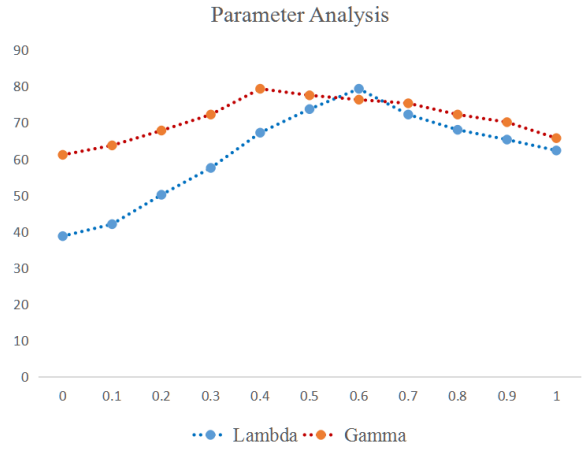


FIGURE 6. Rank-1 accuracy (%) along with different value of parameter λ and γ on DukeMTMC-reID dataset.

ensure the consistency with the input pedestrian for person re-identification task.

In the MOS test, the results are shown in Table 4. In small datasets, PESR outperforms bicubic 0.86(VIPeR) and 0.95(CAVIAR) and outperforms nearest method of 0.93(VIPeR) and 1.1(CAVIAR). It is proved that PESR can be trained and processed effectively in small datasets. In large datasets, PESR outperforms bicubic of 0.84%(Market-1501), 0.43(DukeMTMC-reID) and outperforms nearest method 0.87%(Market-1501) and 0.61(DukeMTMC-reID). That demonstrates the PESR network can also steadily improve the image resolution in DLRPR task.

E. DISCUSSION

1) EFFECTIVENESS OF PART-BASED STRATEGY

Part-based method is a common technique to improve REID performance, and deep learning is expert in extracting discriminative features than hand crafted methods. By extracting local features from each part of pedestrian image and combining them together, we can obtain more accurate pedestrian information than extracting global features directly. In order to verify the role of the part-based mechanism in PESR network, we conduct comparative tests, which we remove the part-based mechanism from the SR module and the REID module, and directly use the global image as the input. This is called Global Enhanced Super Resolution Network(GESR).

We conduct the GESR on DukeMTMC-reID dataset, and record their rank-1 matching accuracies of GESR and PESR, as shown in Fig 5. From the results, we can see that PESR is more effective (at least 22.1%) than the modified GESR. This strongly proves that the part-based mechanism has a positive effect on PESR.

2) EFFECTIVENESS OF PFC LOSS

In our PESR network, PFC loss is an important connection between part images. Through the PFC loss, PESR network can not only resolve the low resolution image generation task, but also retain the discriminative information by the part-based accurate representation. In order to verify the effect of PFC loss on PESR, we conducted a comparative experiment, which we set $\gamma = 0$ to remove the PFC loss. This model is called Disconnected Part-based Enhanced Super Resolution Network(DPESR).

The rank-1 result of DPESR is also reported in Figure 5 and it achieves 61.2% on DukeMTMC-reID dataset. From the comparison, we can see that PESR is more effective (at least 18.2%) than the modified DPESR. The comparative experiments on other datasets also produce the same conclusion. These results demonstrate that PFC loss plays an important role in PESR because of the connection between each part images.

3) PARAMETER ANALYSIS

In order to evaluate the impacts of generator loss and PFC loss on PESR network, we independently change one balance weight of them when fix another at optimal value. We change λ and γ to achieve rank-1 accuracies on DukeMTMC-reID from 0 to 1, and we report their results are shown in Figure 6.

From Figure 6, it can be seen that PESR model achieves rank-1 accuracy of 38.8% when $\lambda = 0$ (generative loss lose efficacy). That demonstrates the poor performance when directly employ REID model on low resolution person REID task without any super resolution operations. Besides, when we set $\gamma = 0$, the PESR model obtain rank-1 accuracy of 61.2% on DukeMTMC-reID. It also illustrates the importance of PFC loss in our model. As a conclusion, our PESR network achieves optimal result of 79.4% when $\lambda = 0.4$ and $\gamma = 0.6$. Similar performance on other datasets are as well as on DukeMTMC-reID in parameter analysis, which $\lambda \in [0.4, 0.6]$ and $\gamma \in [0.3, 0.5]$ can achieve their best performances.

V. CONCLUSION

This paper mainly studies Directly Low Resolution Person Re-identification (DLRPR) task and come up with a Part-based Enhanced Super Resolution (PESR) network to relax the pixel-to-pixel supervision in DLRPR task. Our PESR network introduces a part-based mechanism to extract robust pedestrian feature and an enhanced generative

adversarial network to boost the unpaired pedestrian image super resolution process. Furthermore, in order to reconstruct higher resolution image features over low resolution images, PESR network introduces not only relativistic discriminator, but also part feature connection loss to connect each part of images. Finally, the features of each part are fused for person re-identification. The experimental results and discussion demonstrate that PESR network is superior to the most advanced low-resolution person REID methods and solve the DLRPR task effectively.

REFERENCES

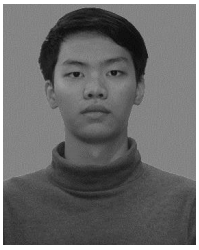
[1] L. Chen, H. Yang, and Z. Gao, "Joint attentive spatial-temporal feature aggregation for video-based person re-identification," *IEEE Access*, vol. 7, pp. 41230–41240, 2019.
[2] L. He, J. Liang, H. Li, and Z. Sun, "Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7073–7082.
[3] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang, "Dual attention matching network for context-aware feature sequence based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5363–5372.
[4] Y.-J. Cho and K.-J. Yoon, "Distance-based camera network topology inference for person re-identification," *Pattern Recognit. Lett.*, vol. 125, pp. 220–227, Jul. 2019.
[5] X. Zhu, X.-Y. Jing, F. Zhang, X. Zhang, X. You, and X. Cui, "Distance learning by mining hard and easy negative samples for person re-identification," *Pattern Recognit.*, vol. 95, pp. 211–222, Nov. 2019.
[6] X. Xu and Y. Chen, "Video-based person re-identification based on regularised hull distance learning," *IET Comput. Vis.*, vol. 13, no. 4, pp. 385–394, Jun. 2019.
[7] X. Zhu, X.-Y. Jing, X. You, W. Zuo, S. Shan, and W.-S. Zheng, "Image to video person re-identification by learning heterogeneous dictionary pair with feature projection matrix," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 717–732, Mar. 2018.
[8] S. Zhou, J. Wang, D. Meng, Y. Liang, Y. Gong, and N. Zheng, "Discriminative feature learning with foreground attention for person re-identification," *IEEE Trans. Image Process.*, to be published.
[9] A. Borgia, Y. Hua, E. Kodirov, and N. M. Robertson, "Cross-view discriminative feature learning for person re-identification," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5338–5349, Nov. 2018.
[10] X.-Y. Jing, X. Zhu, F. Wu, R. Hu, X. You, Y. Wang, H. Feng, and J.-Y. Yang, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1363–1378, Mar. 2017.
[11] J. Jiao, W.-S. Zheng, A. Wu, X. Zhu, and S. Gong, "Deep low-resolution person re-identification," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 6967–6974.
[12] Z. Wang, M. Ye, F. Yang, X. Bai, and S. Satoh, "Cascaded SR-GAN for scale-adaptive low resolution person re-identification," in *Proc. IJCAI*, 2018, pp. 3891–3897.
[13] Y. Wang, L. Wang, Y. You, X. Zou, V. Chen, S. Li, G. Huang, B. Hariharan, and K. Q. Weinberger, "Resource aware person re-identification across multiple resolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8042–8051.
[14] Y.-C. Chen, Y.-J. Li, X. Du, and Y.-C. F. Wang, "Learning resolution-invariant deep representations for person re-identification," in *Proc. AAAI*, vol. 33, Aug. 2019, pp. 8215–8222.
[15] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 7202–7211.
[16] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 480–496.
[17] Q. Yang, H.-X. Yu, A. Wu, and W.-S. Zheng, "Patch-based discriminative feature learning for unsupervised person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 3633–3642.
[18] F. Zhu, X. Kong, L. Zheng, H. Fu, and Q. Tian, "Part-based deep hashing for large-scale person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4806–4817, Oct. 2017.
[19] C. Liu, T. Bao, and M. Zhu, "Part-based feature extraction for person re-identification," in *Proc. 10th Int. Conf. Mach. Learn. Comput.*, 2018, pp. 172–177.
[20] Y. Zhang, X. Gu, J. Tang, K. Cheng, and S. Tan, "Part-based attribute-aware network for person re-identification," *IEEE Access*, vol. 7, pp. 53585–53595, 2019.
[21] L. Wu, R. Hong, Y. Wang, and M. Wang, "Cross-entropy adversarial view adaptation for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[22] X. Zhu, X.-Y. Jing, L. Yang, X. You, D. Chen, G. Gao, and Y. Wang, "Semi-supervised cross-view projection-based dictionary learning for video-based person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2599–2611, Oct. 2018.
[23] X. Zhu, X.-Y. Jing, X. You, X. Zhang, and T. Zhang, "Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5683–5695, Nov. 2018.
[24] L. Wu, Y. Wang, L. Shao, and M. Wang, "3-D PersonVLAD: Learning deep global representations for video-based person reidentification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3347–3359, Nov. 2019.
[25] L. Wu, Y. Wang, H. Yin, M. Wang, and L. Shao, "Few-shot deep adversarial learning for video-based person re-identification," *IEEE Trans. Image Process.*, vol. 29, pp. 1233–1245, Sep. 2019.
[26] X. Zhu, X.-Y. Jing, F. Ma, L. Cheng, and Y. Ren, "Simultaneous visual-appearance-level and spatial-temporal-level dictionary learning for video-based person re-identification," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 7303–7315, Nov. 2019.
[27] F. Ma, X.-Y. Jing, Y. Yao, X. Zhu, and Z. Peng, "High-resolution and low-resolution video person re-identification: A benchmark," *IEEE Access*, vol. 7, pp. 63426–63436, 2019.
[28] Z. Zhuang, H. Ai, L. Chen, and C. Shang, "Cross-resolution person re-identification with deep antithetical learning," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 233–248.
[29] Y.-J. Li, Y.-C. Chen, Y.-Y. Lin, X. Du, and Y.-C. F. Wang, "Recover and identify: A generative dual model for cross-resolution person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2019, pp. 8090–8099.
[30] X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, "Multi-scale learning for low-resolution person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3765–3773.
[31] S. Mao, S. Zhang, and M. Yang, "Resolution-invariant person re-identification," in *Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI)*, Macao, China, Aug. 2019, pp. 883–889, doi: 10.24963/ijcai.2019/124.
[32] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[33] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.
[34] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 63–79.
[35] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
[36] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4539–4547.
[37] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.
[38] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," 2018, *arXiv:1812.10477*. [Online]. Available: https://arxiv.org/abs/1812.10477
[39] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[40] K. Yu, C. Dong, L. Lin, and C. Change Loy, "Crafting a toolchain for image restoration by deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2018, pp. 2443–2452.

[41] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 701–710.
[42] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[43] Z. Wang, R. Hu, Y. Yu, J. Jiang, C. Liang, and J. Wang, "Scale-adaptive low-resolution person re-identification via learning a discriminating surface," in *Proc. IJCAI*, 2016, pp. 2669–2675.
[44] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 262–275.
[45] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. BMVC*, 2011, vol. 1, no. 2, p. 6.
[46] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.
[47] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline *in vitro*," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3754–3762.

YAN HA received the M.S. degree in computer science and technology from Hebei University, Baoding, China, in 2016, where she is currently pursuing the Ph.D. degree in management engineering. Her research interests include deep learning, trusted computing, and computer vision.

JUNFENG TIAN was born in 1965. He received the Ph.D. degree from the China University of Technology. He is currently a Professor with Hebei University. He is also a Professor and the Deputy Secretary General of the Open System Specialized Committee, China Computer Federation, a young and middle-aged Specialist with outstanding contribution, Hebei, and a young and middle-aged Teacher with outstanding contribution, Hebei. He is also with the Key Laboratory on High Trusted Information System in Hebei Province. He has been involved in teaching and research in the fields of distributed computing and network technology for many years. His research interests include information security and trusted computing.

QIAOWEI MIAO is currently pursuing the bachelor's degree in computer science and technology with Hebei University, Baoding, China. His research interests include image super resolution and pattern recognition.



QI YANG is currently pursuing the bachelor's degree in network engineering with Hebei University, Baoding, China. His research interests include deep learning and pattern recognition.

JIAAO GUO is currently pursuing the bachelor's degree in information security with Hebei University, Baoding, China. His research interests include machine learning and computer vision.

RUOHUI JIANG is currently pursuing the bachelor's degree in computer science and technology with Hebei University, Baoding, China. His research interest is on computer vision.