IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2019, accepted December 3, 2019, date of publication December 6, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958126

Semantic Constraint GAN for Person Re-Identification in Camera Sensor Networks

SHUANG LIU, (Senior Member, IEEE), TONGZHEN SI, XIAOLONG HAO, AND ZHONG ZHANG, (Senior Member, IEEE)

Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China

Corresponding author: Zhong Zhang (zhong.zhang8848@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501327 and Grant 61711530240, in part by the Natural Science Foundation of Tianjin under Grant 19JCZDJC31500 and Grant 17JCZDJC30600, in part by the Fund of Tianjin Normal University under Grant 135202RC1703, in part by the Open Projects Program of National Laboratory of Pattern Recognition under Grant 201800002, and in part by the Tianjin Higher Education Creative Team Funds Program.

ABSTRACT In this paper, we propose a novel data augmentation method named Semantic Constraint Generative Adversarial Network (SCGAN) for person re-identification (Re-ID) in camera sensor networks. The proposed SCGAN can generate multiple style pedestrian images with high-level semantic information. To this end, we design two types of semantic constraints, i.e., attention constraint and identity constraint. The attention constraint aims to restrict the significant areas in the attention map to be consistent before and after image transformation. The identity constraint focuses on keeping the identity of the generated pedestrian image to be the same as that of the real one. After generating pedestrian images using SCGAN, we combine them with the real pedestrian images to train the person Re-ID model. Since the proposed SCGAN increases the diversity of training samples, the generalization of Re-ID model is enhanced. We evaluate the proposed SCGAN on three large-scale person Re-ID databases, i.e., Market1501, CUHK03 and DukeMTMC-reID, and experimental results reveal that the proposed SCGAN yields consistent improvements over other methods.

INDEX TERMS Person re-identification, data augmentation, camera sensor networks.

I. INTRODUCTION

The sensor networks [1], [2] consist of various sensors, such as video cameras, microphones and so on, and they are widely used in video surveillance. Person re-identification (Re-ID) in camera sensor networks is a fundamental task in video surveillance, and it aims at matching an interested pedestrian from a gallery set collected under different cameras [3], [4]. In recent years, person Re-ID has attracted a lot of consideration in academia and industry because of its wide applications in multi-camera tracking, crowd counting, etc [5]–[7]. Although person Re-ID in camera sensor networks has been studied for many years, it still confronts many difficulties, such as complex lighting, viewpoint changes, various body poses and so on.

Recently, with the renaissance of deep learning [8], [9], a great deal of methods for person Re-ID have been debated to extract discriminative features [10], [11], construct robust models [12], [13] or both of them [14]. These methods are beneficial to the performance of person Re-ID in a certain range, but they remain several open issues. For example, there is a large gap between the training and test sets caused by large variations in body poses, illuminations, backgrounds and so on, which leads to the performance degeneration on the test set.

To address this limitation, the intuitive approach is to augment the number of training samples in order to promote the generalization ability of deep network model. However, it is prohibitively expensive to collect and annotate large-scale databases for person Re-ID. Alternatively, data augmentation is convenient to extend the training set without extra cost, and it is widely used in deep learning [15], [16]. The straightforward approaches of data augmentation include random flipping and random cropping, and they have been validated to be effective for person Re-ID [17], [18]. More recently, Generative Adversarial Network (GAN) and its variants have

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

been applied in many research fields [19], [20], and they are also utilized to augment the training samples for person Re-ID [21]–[23]. Then, the real training samples and the generated samples are both employed to train the deep network model. However, most existing GAN-based methods only focus on increasing the visual appeal for human, but ignore the constraint of semantic information between pedestrian images and generated ones, which degrades the performance of person Re-ID.

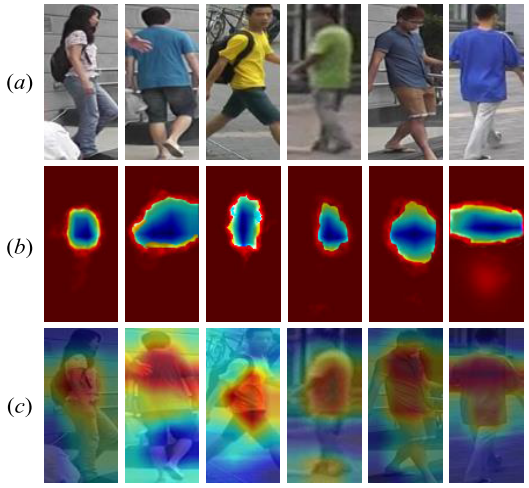

FIGURE 1. (a) Real pedestrian images, (b) the corresponding heatmaps where each pixel denotes the predicted true-class probability from CNN model when using a mask to occlude the pedestrian image and (c) attention maps of these pedestrian images.

In this paper, we propose a novel data augmentation method named Semantic Constraint Generative Adversarial Network (SCGAN) to explicitly consider the semantic information in the process of image generation for person Re-ID in camera sensor networks. We first utilize SCGAN to generate pedestrian images with different camera styles, and combine them with real pedestrian images to train the person Re-ID model. As for SCGAN, it consists of two types of semantic constraints, i.e., attention constraint and identity constraint. As shown in Fig. 1 (b), only a part of pedestrian is crucial for the Re-ID performance, and therefore these regions are significant areas. Hence, we expect to maintain the significant areas in the process of image generation. Meanwhile, from Fig. 1 (c) we can see that the attention maps reflect where the deep network focuses on, and they could be utilized to locate the significant areas. Based on the above observations, we propose the attention constraint which restricts the significant areas in the attention map of real pedestrian image and generated pedestrian image to be consistent. As a result, the discriminative ability of generated pedestrian images is improved, which is beneficial to the person Re-ID performance.

The identity label contains high-level semantic information of pedestrians, and it is an important factor in the process of image generation. From Fig. 2, we can see that the generated pedestrian images usually achieve lower predicted probability

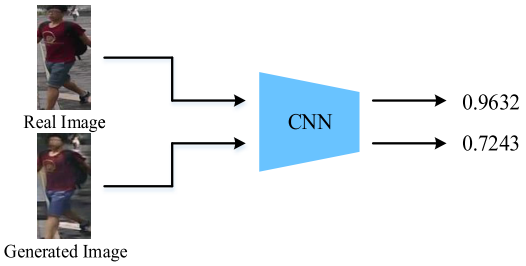

FIGURE 2. The predicted probabilities of real and generated pedestrian images.

than the real pedestrian images, which reveals the generated pedestrian images lose some semantic information. Thus, we propose the identity constraint to explicitly enforce the identity labels before and after generation. After adding the above two constraints, the generated pedestrian images not only possess different camera styles from real pedestrian images, but also enhance the discriminative areas and ensure the stability of person identity. With these generated pedestrian images, the diversity of training samples is enhanced, and therefore the generalization ability of deep model could be improved.

In a word, the contributions of the work are as follows:

1) The proposed SCGAN explicitly considers the semantic information using two constraints in the image generation process, where the generated pedestrian images could promote the generalization ability of deep model.

2) The proposed SCGAN does not require additional operations during the image generation process, which is free from the errors caused by extra algorithms.

3) The proposed SCGAN outperforms the state-of-the-art methods on three large-scale person Re-ID databases, i.e., Market1501 [24], CUHK03 [26] and DukeMTMC-reID [27].

II. RELATED WORK

A. PERSON RE-IDENTIFICATION

The sensor networks drive the applications in many research fields, such as human action recognition [28], soil moisture retrieval [25], [29]. Person Re-ID is a vital application in camera sensor networks, and it has shown tremendous progress when using Convolutional Neural Networks (CNNs) [30]–[33]. The CNN-based methods mainly focus on either feature extraction [34]–[39] or similarity measurement [12], [14], [40], and optimize them in an end-to-end method. As for feature extraction, Wu *et al.* [36] utilized smaller convolutional filters to learn features from holistic pedestrian images. Qian *et al.* [37] proposed a multi-scale deep learning model which mines holistic features from multiple scales. In addition, some methods focus on learning latent part features of pedestrian images to discover local cues. Yi *et al.* [38] partitioned pedestrian samples into three overlapped subregions and fed them into a CNN model to capture local cues of pedestrian images. Zhao *et al.* [34] exploited an off-the-shelf

pose estimation model to predict body joints, and then split each pedestrian image into several local regions to learn discriminative representations. Sun *et al.* [35] presented the Visibility-aware Part Model (VPM) to learn region-level features, and compared the pedestrian image pair using their shared regions for partial person Re-ID.

Many researchers study the similarity measurement to obtain robust person Re-ID models [12], [14], [40], [41]. Hermans *et al.* [40] employed the hardest positive and negative samples to calculate the loss value so as to generate a proper distance among pedestrian images. Chen *et al.* [12] introduced the quadruplet loss that treats four pedestrian images as a unit in order to raise inter-class variations and meanwhile reduce intra-class variations. Si *et al.* [41] introduced the Compact Triplet Loss (CTL) for person Re-ID, which makes the training samples be close to the corresponding feature centers and the different feature centers be away from each other.

B. GENERATIVE ADVERSARIAL NETWORKS

GANs [19] have obtained remarkable success in multiple tasks for instance image generation, image translation and style transfer [20], [42], [43]. A typical GAN model contains a generator and a discriminator. The generator learns to produce the generated images that are indistinguishable from the real images, while the discriminator wants to classify the real images and the generated ones. Radford *et al.* [44] utilized the convolutional layer to construct Deep Convolutional GANs (DCGANs) which could produce fake images to increase training samples. In [20], CycleGAN was introduced to learn a translation from the source domain to the target domain without paired images. Choi *et al.* [45] proposed the StarGAN to implement multi-domain image translations using a single generator and a single discriminator.

Due to the desirable characteristics of GANs, they have been utilized for person Re-ID so as to augment training samples. Zheng *et al.* [27] utilized DCGAN to generate pedestrian images and employed these generated images to fine-tune the person Re-ID model. They first demonstrated that generated pedestrian images could remarkably promote the performance of Re-ID model. Wei *et al.* [23] designed a new model named PTGAN to reduce the gap between the training and test sets. The PTGAN employed an extra algorithm, i.e., PSNet [46] to extract the foregrounds of pedestrian images in order to generate high-quality pedestrian images. Zhong *et al.* [22] utilized CycleGAN to produce the style-transferred images, and then combined real training samples to train the person Re-ID model.

The above-mentioned methods generate pedestrian images to enhance the number of training images, but they ignore the semantic constraint before and after generation, which leads to large semantic difference between generated pedestrian images and real ones. Unlike previous methods, our work aims to learn the semantic information in the translation process so as to maintain the discriminative areas and stabilize the pedestrian identities between generated pedestrian images and real ones.

C. ATTENTION MECHANISM

The attention model is able to automatically locate key object regions and represent them according to their appearances [47]. It extracts visual features from salient regions, and is widely used in various tasks, especially in image classification, multi-object tracking and visual question answering [48]–[50]. Currently, the attention mechanism is also applied in person Re-ID field to mine discriminative features. Zhou *et al.* [51] introduced a temporal attention model that can automatically pick out the most discriminative frame to improve feature performance for video-based person Re-ID. Yang *et al.* [52] proposed an attention-driven multi-branch network to learn intra-attention and inter-attention, which could guide the model to mine discriminative pedestrian representations. In this paper, we utilize the attention mechanism to select discriminative areas of pedestrian so as to constrain the semantic information in the generation process.

III. APPROACH

In this section, we first briefly review StarGAN [45] for person Re-ID in camera sensor networks. Then, we introduce the proposed semantic constraints, i.e., attention constraint and identity constraint, for SCGAN in detail. Finally, we explain how to apply SCGAN to train the deep person Re-ID model.

A. StarGAN FOR PERSON RE-ID

The pedestrian images are usually captured by multiple cameras and possess different camera styles as shown in Fig. 3 (a). Recent works [20], [23] have demonstrated camera styles are crucial for person Re-ID performance. They consider that pedestrian images with different camera styles distribute in different domains and employ generative adversarial networks, such as CycleGAN and PTGAN, to transfer camera styles among pedestrian images in order to enrich the training samples. However, these methods transform one style pair each time, and therefore they have limited scalability when meeting multiple camera styles in person Re-ID task.

Alternatively, StarGAN [45] could produce pedestrian images with different styles only training once, and therefore it is more suitable for generating pedestrian images for person Re-ID. StarGAN aims to learn one single generator to transfer all camera styles among pedestrian images. To this end, it is designed as a generator G and a discriminator D. The generator is formulated as $G(x, c_t) \rightarrow y$, where x and y are the real and generated pedestrian images respectively, and c_t is the target camera style. The discriminator strives to distinguish the pedestrian images belonging to the real or generated ones. To control the transformation of multiple camera styles, an auxiliary classifier is added into D so that the discriminator could also recognize the camera labels of pedestrian images. Specifically, StarGAN consists of three kinds of losses, i.e., adversarial loss, camera classification loss and reconstruction loss.

As for the adversarial loss, it makes the generated pedestrian images confuse the discriminator so that the generated pedestrian images are similar to the real pedestrian images:

$$L_{adv} = \mathbb{E}_x[logD_{src}(x)] + \mathbb{E}_{x,c_t}[log(1 - D_{src}(G(x, c_t)))], \quad (1)$$

where $D_{src}(x)$ represents the predicted probability that x belongs to the real pedestrian image.

The camera classification loss aims to distinguish the camera labels of pedestrian images, and it is defined as:

$$L_{cla} = r\mathbb{E}_{x,c_s}[-logD_{cls}(c_s|x)] + (1 - r)\mathbb{E}_{x,c_t}[-logD_{cls}(c_t|G(x, c_t))]. \quad (2)$$

where c_s is the camera label of the real pedestrian image, and $D_{cls}(c_s|x)$ expresses the predicted probability that x belongs to the camera label c_s. When the input image of D_{cls} is the real pedestrian image, $r = 1$, otherwise $r = 0$.

In addition, the reconstruction loss is applied to preserve the content of the real pedestrian image in the generation process:

$$L_{rec} = \mathbb{E}_{x,c_s,c_t}[||x - G(G(x, c_t), c_s)||_1], \quad (3)$$

where $||\cdot||_1$ indicates the $L1$ norm.

B. SEMANTIC CONSTRAINT GAN

Various GANs have been applied to image generation for person Re-ID. However, most of them only focus on increasing the visual effect of generated images, but ignore the semantic information, which leads to the large semantic difference between generated images and real ones. In order to guarantee the semantic consistency in the generation process, we propose two novel semantic constraints, i.e., attention constraint and identity constraint, based on StarGAN.

The significant areas are key to the discrimination of pedestrian representations, and therefore we expect the significant areas to keep consistency in the generation process. Since attention maps could reflect the areas that the deep network concerns, we utilize attention maps to calculate the significant areas. Specifically, we first construct a multi-classification deep network model F to perform a classification task. Then, the attention map of pedestrian image is extracted from F and it is formulated as:

$$R(x) = \sum_l |T_l(x)|^2, \quad (4)$$

where $T_l(x)$ illustrates the l-th feature map of x. It is worth mentioning that all operations are between elements. Afterwards, the significant areas could be formulated as:

$$A(x) = K_s(R(x)), \quad (5)$$

where $K_s()$ is the operation that sets the first $s\%$ large values to 1 and others to 0.

We aim to constrain the significant areas between the generated pedestrian images and the real ones. Hence, the attention constraint loss is formulated as:

$$L_{att} = \mathbb{E}_{x,c_t}[||(G(x, c_t) - x) \odot A(x)||_1], \quad (6)$$

where $\odot$ denotes the element-wise multiplication.

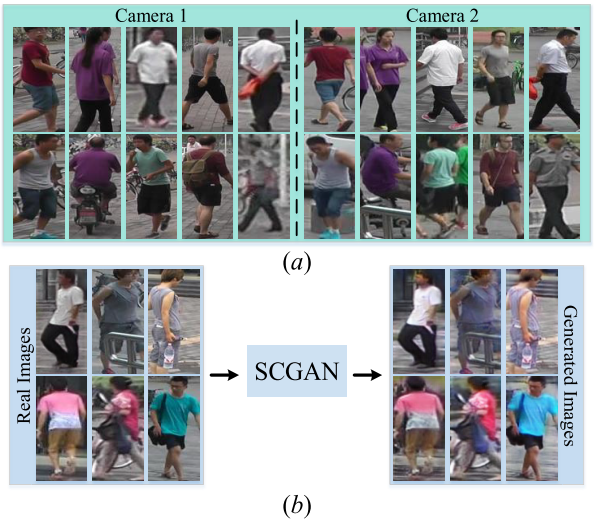

FIGURE 3. (a) Some pedestrian images from two different cameras on Market1501 and (b) some generated pedestrian images from a camera to the other camera by SCGAN.

The person Re-ID is usually treated as a multi-classification task, and therefore the identity labels are vital to the learning process. Additionally, they contain abundant semantic information. Hence, we propose the identity constraint to ensure the generated pedestrian image has the same identity with the real pedestrian image. The identity constraint forces G to generate different camera style pedestrian images with the same identity labels as the real ones, and it is defined as:

$$L_{ide} = r\mathbb{E}_{x,k}[-\sum_{k=1}^{K} q(k)\log(F_k(x))], + (1 - r)\mathbb{E}_{x,c_t,k}[-\sum_{k=1}^{K} p(k)\log(F_k(G(x, c_t)))], \quad (7)$$

where K is the number of identities, $F_k(x)$ expresses the predicted probability that x belongs to the k-th identity, and $q(k)$ is the true label distribution of the real pedestrian image:

$$q(k) = \begin{cases} 0 & k \neq g \\ 1 & k = g, \end{cases} \quad (8)$$

where g is the true identity. $p(k)$ indicates the predicted label distribution of the generated pedestrian image, and is written as:

$$p(k) = \begin{cases} 0 & k \neq \arg\max_b(F_b(x)) \\ 1 & k = \arg\max_b(F_b(x)), \end{cases} \quad (9)$$

In a word, the objective of SCGAN is formulated as:

$$\begin{aligned} L_D &= -L_{adv} + \lambda_{cla}L_{cla}|_{r=1} \\ L_G &= L_{adv} + \lambda_{rec}L_{rec} + \lambda_{att}L_{att} \\ &\quad + \lambda_{cla}L_{cla}|_{r=0} + \lambda_{ide}L_{ide}|_{r=0} \\ L_F &= \lambda_{ide}L_{ide}|_{r=1}. \end{aligned} \quad (10)$$

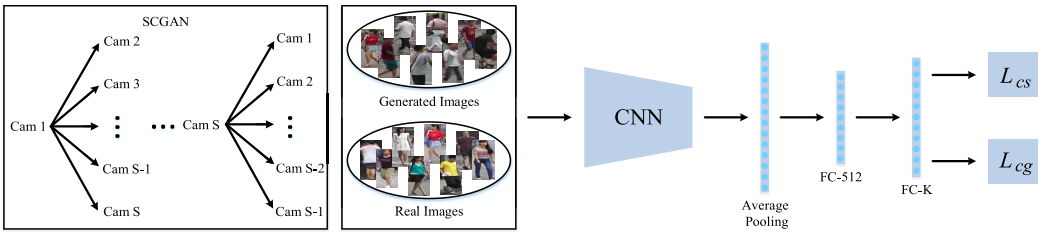

FIGURE 4. The flowchart of training the deep model using real and generated pedestrian images.

where λ_{cla}, λ_{rec}, λ_{att} and λ_{ide} are hyper-parameters that decide the relative importance of the corresponding loss, $|_{r=1}$ indicates that the loss function is calculated using real pedestrian images, and $|_{r=0}$ represents that the loss function is computed using generated pedestrian images. In all experiments, we set $\lambda_{cla} = 1$, $\lambda_{rec} = 10$, $\lambda_{att} = 1$ and $\lambda_{ide} = 1$. L_D, L_G and L_F express the objective functions of network D, G and F, respectively. We minimize L_D, L_G and L_F to optimize the three deep models, iteratively.

C. PERSON RE-ID WITH SCGAN

In order to increase the diversity of pedestrian images, we employ the proposed SCGAN to generate new training samples. Given a person Re-ID database captured by S cameras, we can generate S-1 kinds of camera styles for each pedestrian image using SCGAN. Hence, the training set is enlarged S times. Since each generated pedestrian image preserves the semantic information of the corresponding real one, the generated images are considered to have the same identity as the corresponding real ones. Some generated pedestrian images are shown in Fig. 3 (b) where we can see that the generated pedestrian images are different from the real ones. We combine these generated pedestrian images and real pedestrian images as the training samples to train a deep network.

After obtaining an expanded training set, we structure a person Re-ID model using modified ResNet-50 as shown in Table 1. We displace the last classification layer with a 512-dimensional fully connected layer (FC-512) and a K-dimensional fully connected layer (FC-K) where K is the total number of identities. It should be noticed that FC-512 is followed by Batch Normalization, LeakyReLU and Dropout. We treat person Re-ID as a classification task, and employ different losses for real and generated pedestrian images. The whole architecture is shown in Fig. 4.

As for real pedestrian images, we only consider the ground-truth label distribution and employ the cross-entropy loss:

$$L_{cs} = -\sum_{k=1}^{K} q(k)\log(l(k)) \tag{11}$$

where $l(k)$ denotes the predicted probability that the real pedestrian image belongs to the k-th identity, and $q(k)$ has the same meaning as Eq. 8.

TABLE 1. The structure of the modified ResNet-50. It is noted that the batch normalization is utilized after each convolutional layer.

Name	Filters	Stride
Conv1	$\left[7 \times 7, 64\right] \times 1$	(2, 2)
Max Pooling	3×3	(2, 2)
Conv2_X	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} (1,1) \\ (1,1) \\ (1,1) \end{bmatrix}$
Conv3_X	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} (1,1) \\ (2,2),(1,1) \\ (1,1) \end{bmatrix}$
Conv4_X	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} (1,1) \\ (2,2),(1,1) \\ (1,1) \end{bmatrix}$
Conv5_X	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} (1,1) \\ (1,1) \\ (1,1) \end{bmatrix}$
Average Pooling	24×8	(1, 1)
FC-512		
FC-K		

Although the generated pedestrian images contain semantic information of real pedestrian images, they are not perfect enough. To handle this situation, the label smoothing regularization (LSR) [53] was proposed to consider all distributions. Specifically, it assigns a large weight to the ground-truth identity, while a small weight to the other identities. This encourages the deep network model to have less trust in the ground-truth identity. Thus, we employ LSR to train the generated pedestrian images, and the label distribution of LSR is defined as:

$$LSR(k) = \begin{cases} \dfrac{\epsilon}{K} & k \neq g \\ 1 - \epsilon + \dfrac{\epsilon}{K} & k = g \end{cases} \tag{12}$$

where $\epsilon \in [0, 1]$ is the hyper-parameter to adjust the ratio of non-ground truth identity labels. Based on this label

TABLE 2. Comparison of the proposed SCGAN with other style-transferred methods.

Methods	Market1501 (%)		CUHK03 (labeled) (%)		CUHK03 (detected) (%)		DukeMTMC-reID (%)	
	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP
Baseline	88.7	72.9	53.5	49.8	52.2	48.9	77.8	61.5
Baseline+CycleGAN	91.1	74.7	54.6	50.5	53.4	50.2	79.5	63.6
Baseline+StarGAN	91.7	74.9	55.2	51.3	54.1	50.4	80.3	64.2
Baseline+SCGAN	**93.3**	**76.8**	**58.2**	**56.5**	**56.9**	**54.1**	**82.5**	**66.2**

distribution, the cross-entropy loss is reformulated as:

$$L_{cg} = -(1-\epsilon)log(l(g)) - \frac{\epsilon}{K}\sum_{k=1}^{K}\log(l(k)) \tag{13}$$

In the training process of deep model, we randomly select U real pedestrian images and V generated pedestrian images in one batch, and the total loss is written as:

$$L_T = \frac{1}{U}\sum_{i=1}^{U} L_{cs}^{i} + \frac{1}{V}\sum_{j=1}^{V} L_{cg}^{j} \tag{14}$$

where L_{cs}^{i} and L_{cg}^{j} denote the loss values of the i-th real pedestrian image and the j-th generated pedestrian image, respectively.

IV. EXPERIMENTS

In this section, we first describe three widely used person Re-ID databases including Market1501 [24], CUHK03 [26] and DukeMTMC-reID [27]. Then, we present the experimental details, and compare SCGAN with other style-transferred methods. Afterwards, we compare the proposed SCGAN with state-of-the-art methods. Finally, we evaluate important parameters.

A. DATABASES

Market1501 [24] possesses 32,668 pedestrian images with 1,501 identities and it is captured from six different cameras. These images are divided into the training samples and the test samples. The training samples contain 751 identities and the test samples include 750 identities. All pedestrian images are obtained by Deformable Parts Model (DPM) [55].

CUHK03 [26] includes 1,467 identities from 14,096 pedestrian images. This database is captured from two different cameras and has 9.6 pedestrian images for each identity. These pedestrian images contain two kinds of types, i.e., detected set extracted by DPM and labeled set extracted by hand-drawn. We employ the same training and test protocol as [56], [57]. The training set has 7,365 pedestrian images and the gallery set has 5,332 pedestrian images. The remaining pedestrian images are treated as the query set including 1,400 pedestrian images. We comprehensively verify the proposed SCGAN based on the two types of image settings.

DukeMTMC-reID [27] includes 1,812 identities from 36,411 pedestrian images and is captured by eight high-resolution cameras. All samples are divided into the training set, the test set and the query set. Specifically, the training set has 16,522 pedestrian images, the test set contains 17,661 pedestrian images, and the query set includes 2,228 pedestrian images.

B. IMPLEMENTATION DETAILS

SCGAN employs similar network architecture with StarGAN [45]. As for the generator network, it includes two convolutional layers, six residual blocks, and two deconvolution layers, and all convolutional layers are with a step size of two. As for the discriminator network, it contains two parts. One part utilizes PatchGAN [58] to classify whether a pedestrian image is real or generated, and the other part aims to recognize the camera label of a pedestrian image. In addition, the instance normalization [59] is only applied on the generator. In the training process of SCGAN, the pedestrian images are cropped to 128×64 and Adam solver [63] is adopted to optimize all models. The epoch number is set to 200.

After obtaining the generated pedestrian images, we combine them with the real pedestrian images to train the person Re-ID model. All pedestrian images are cropped to 384×128. We implement random cropping and random horizontal flipping on the real pedestrian images for data augmentation, but there is no data augmentation for generated pedestrian images. The SGD solver is employed to optimize the person Re-ID model and the number of epochs is set to 60. Since the generated image produced by the proposed SCGAN is several times of the real image, we utilize all real pedestrian images and randomly select some generated pedestrian images in one batch. We set the batch size to 56 where the number of real pedestrian images is 32 and the number of generated pedestrian images is 24. In addition, the dropout probability is set to 0.5, and the gradient of Leaky ReLU is 0.1. In the evaluation phase, we extract the output of the final pooling layer (512-dim) as the features of pedestrian images, and calculate the Euclidean distance as the similarity between pedestrian images.

C. COMPARISON WITH OTHER STYLE-TRANSFERRED METHODS

We comprehensively compare SCGAN with other style-transferred methods, and the results are summarized in Table 2. Baseline illustrates that we only utilize the real pedestrian images to train an IDE model [61]. For comparison, we employ the popular CycleGAN [20] and StarGAN [45] to obtain generated pedestrian images, and combine them with Baseline. From the table, our method

(Baseline+SCGAN) achieves the best performance, which verifies the effectiveness of SCGAN. It illustrates that generating pedestrian images with multiple styles is an effective data augmentation method.

Furthermore, SCGAN is superior to CycleGAN and StarGAN in promoting the performance of person Re-ID. It is because the proposed SCGAN can control the significant areas before and after image transformation by using the attention constraint. Meanwhile, the proposed SCGAN keep the identity information of pedestrian images in the generation process by adding the identity constraint. In a word, it can keep the high-level semantic information of the pedestrian image before and after transformation.

D. COMPARISON WITH THE STATE-OF-THE-ART METHODS

In this subsection, we compare the proposed SCGAN with the state-of-the-art methods on the three databases.

TABLE 3. Comparison of the proposed SCGAN with the state-of-the-art methods on the Market-1501 database.

Methods	rank-1 (%)	mAP (%)
BoW [24]	34.4	14.1
SSDAL [64]	39.4	19.6
LOMO+XQDA [65]	43.8	22.2
Gated [67]	65.9	39.6
IDE [61]	73.7	51.5
PIE [60]	79.3	56.0
PAR [11]	81.0	63.4
SVDNet [10]	82.3	62.1
LSRO [27]	84.0	66.1
TriNet [40]	84.9	69.1
FMN [68]	86.0	67.1
AOS [57]	86.5	70.4
APR [66]	87.0	66.8
REDA [69]	87.1	71.3
DPFL [62]	88.6	72.6
DML [54]	89.3	70.5
MLFN [30]	90.0	74.3
HA-CNN [31]	91.2	75.7
SCGAN	**93.3**	**76.8**

The comparison results on Market1501 are listed in Table 3. The proposed SCGAN attains the best results, i.e., 93.3% and 76.8% on rank-1 accuracy and mAP, respectively. This indicates that the proposed SCGAN as a data augmentation method could alleviate the over-fitting phenomenon produced by the lack of training samples to a certain extent. More importantly, the proposed SCGAN could generate multiple style pedestrian images with high-level semantic information.

We then verify the performance of the proposed SCGAN on CUHK03, and the results are illustrated in Table 4. The rank-1 accuracy and the mAP obtain 58.2% and 56.5% on the labeled set, respectively. In addition, the proposed SCGAN achieves 56.9% rank-1 accuracy, 54.1% mAP on the detected set. These results exceed all the state-of-the-art methods. This demonstrates that SCGAN could generate high quality pedestrian images to improve the performance of person Re-ID.

TABLE 4. Comparison of the proposed SCGAN with the state-of-the-art methods on the CUHK03 database.

Methods	labeled set (%)		detected set (%)	
	rank-1	mAP	rank-1	mAP
LOMO+XQDA [65]	14.8	13.6	12.8	11.5
IDE [61]	22.2	21.0	21.3	19.7
DPFL [62]	40.7	37.0	-	-
FMN [68]	42.6	39.2	40.7	38.1
SVDNet [10]	40.9	37.8	41.5	37.3
HA-CNN [31]	44.4	41.0	41.7	38.6
AOS [57]	-	-	47.1	43.3
TriNet [40]	49.9	46.7	50.5	46.5
MLFN [30]	54.7	49.2	52.8	47.8
REDA [69]	58.1	53.8	55.5	50.7
SCGAN	**58.2**	**56.5**	**56.9**	**54.1**

TABLE 5. Comparison of the proposed SCGAN with the state-of-the-art methods on the DukeMTMC-reID database.

Methods	rank-1 (%)	mAP (%)
LOMO+XQDA [65]	30.8	17.0
IDE [61]	65.2	45.0
LSRO [27]	67.7	47.1
TriNet [40]	72.4	53.5
APR [66]	73.9	55.6
FMN [68]	74.5	56.9
SVDNet [10]	76.7	56.8
AOS [57]	79.1	62.1
DPFL [62]	79.2	60.6
REDA [69]	79.3	62.4
Deep-Person [70]	80.9	64.8
SCGAN	**82.5**	**66.2**

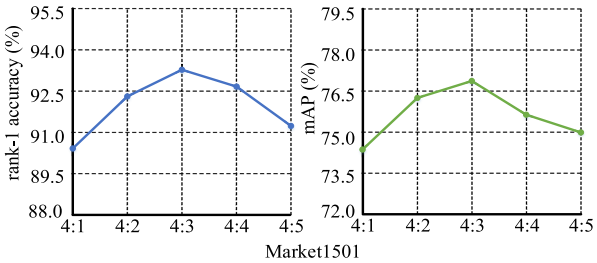

FIGURE 5. The influence of the number of the generated pedestrian images in one batch on Market1501.

As shown in Table 5, we also verify the proposed SCGAN on DukeMTMC-reID, and the results yield higher accuracies. Specifically, the proposed SCGAN achieves 82.5% rank-1 accuracy and 66.2% mAP. In addition, compared with the other methods, the proposed SCGAN yields the best results, which verifies the superiority of SCGAN once again.

E. PARAMETER ANALYSIS

In this work, we combine the generated pedestrian images and the real ones to jointly optimize the person Re-ID model. For example, since each identity on Market1501 is captured by six cameras, we generate other five kinds of camera styles for each pedestrian image. The number of the generated pedestrian images is very large, and we only utilize a part of them. The ratio of the real pedestrian images to the generated pedestrian images is important. We evaluate the ratio

on Market1501, i.e., $U : V$, and the effect are summarized in Fig. 5 where the proposed SCGAN obtains the best results when $U{:}V = 4 : 3$.

V. CONCLUSION

In this paper, we have proposed the data augmentation method named Semantic Constraint Generative Adversarial Network (SCGAN) for person Re-ID in camera sensor networks. The proposed SCGAN could maintain the semantic information to generate high quality pedestrian images. Specifically, we have proposed two kinds of semantic constraints named attention constraint and identity constraint. To take advantage of these images, we combine them with the real pedestrian images to train the deep model. Furthermore, we apply label smooth regularization (LSR) to softly label the generated pedestrian images. Experiments on the three databases show that the proposed SCGAN effectively increases the diversity of pedestrian images and meanwhile promotes the performance of person Re-ID.

REFERENCES

[1] J. Liang and C. Mao, "Distributed compressive sensing in heterogeneous sensor network," *Signal Process.*, vol. 126, pp. 96–102, Sep. 2016.
[2] X. Yu and J. Liang, "Genetic fuzzy tree based node moving strategy of target tracking in multimodal wireless sensor network," *IEEE Access*, vol. 6, pp. 25764–25772, 2018.
[3] S. Tan, F. Zheng, and L. Shao, "Dense invariant feature based support vector ranking for person re-identification," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2015, pp. 687–691.
[4] Z. Wang, R. Hu, C. Chen, Y. Yu, J. Jiang, C. Liang, and S. Satoh, "Person reidentification via discrepancy matrix and matrix metric," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 3006–3020, Oct. 2018.
[5] J. Berclaz, F. Fleuret, and P. Fua, "Multi-camera tracking and atypical motion detection with behavioral maps," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 112–125.
[6] W. Ge and R. T. Collins, "Marked point processes for crowd counting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2913–2920.
[7] M. Ye, J. Li, A. J. Ma, L. Zheng, and P. C. Yuen, "Dynamic graph co-matching for unsupervised video-based person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2976–2990, Jun. 2019.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[9] J. Liang, X. Yu, and H. Li, "Collaborative energy-efficient moving in Internet of Things: Genetic fuzzy tree versus neural networks," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6070–6078, Aug. 2019.
[10] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 3820–3828.
[11] L. Zhao, X. Li, J. Wang, and Y. Zhuang, "Deeply-learned part-aligned representations for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 3219–3228.
[12] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 403–412.
[13] D. Wu, S.-J. Zheng, W. Z. Bao, X.-P. Zhang, C.-A. Yuan, and D.-S. Huang, "A novel deep model with multi-loss and efficient training for person re-identification," *Neurocomputing*, vol. 324, pp. 69–75, Jan. 2019.
[14] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person re-identification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 1–20, 2017.
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 135–153.
[17] Z. Zhang, H. Zhang, and S. Liu, "Coarse-fine convolutional neural network for person re-identification in camera sensor networks," *IEEE Access*, vol. 7, pp. 65186–65194, 2019.
[18] F. Ma, X. Jing, X. Zhu, Z. Tang, and Z. Peng, "True-color and grayscale video person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 115–129, May 2019.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[20] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired imageto-image translation using cycle-consistent adversarial networks," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.
[21] S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 393–400, Aug. 2019.
[22] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camstyle: A novel data augmentation method for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1176–1190, Mar. 2019.
[23] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 79–88.
[24] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1116–1124.
[25] J. Liang and F. Zhu, "Soil moisture retrieval from UWB sensor data by leveraging fuzzy logic," *IEEE Access*, vol. 6, pp. 29846–29857, 2018.
[26] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.
[27] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 3754–3762.
[28] Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi, "Cross-view action recognition via a continuous virtual path," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2690–2697.
[29] J. Liang, X. Liu, and K. Liao, "Soil moisture retrieval using UWB echoes via fuzzy logic and machine learning," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3344–3352, Oct. 2018.
[30] X. Chang, T. M. Hospedales, and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2109–2118.
[31] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2285–2294.
[32] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, and Q. Tian, "Deep representation learning with part loss for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2860–2871, Jun. 2019.
[33] S. Liu, X. Hao, and Z. Zhang, "Pedestrian retrieval via part-based gradation regularization in sensor networks," *IEEE Access*, vol. 6, pp. 38171–38178, 2018.
[34] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, "Spindle Net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 907–915.
[35] Y. Sun, Q. Xu, Y. Li, C. Zhang, Y. Li, S. Wang, and J. Sun, "Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 393–402.
[36] L. Wu, C. Shen, and A. van den Hengel, "PersonNet: Person re-identification with deep convolutional neural networks," Jan. 2016, *arXiv:1601.07255*. [Online]. Available: https://arxiv.org/abs/1601.07255
[37] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5409–5418.
[38] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for practical person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2014, pp. 34–39.
[39] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1335–1344.
[40] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," Mar. 2017, *arXiv:1703.07737*. [Online]. Available: https://arxiv.org/abs/1703.07737

[41] T. Si, Z. Zhang, and S. Liu, "Compact triplet loss for person re-identification in camera sensor networks," *Ad Hoc Netw.*, vol. 95, Dec. 2019, Art. no. 101984.
[42] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 95–104.
[43] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2414–2423.
[44] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–16.
[45] Y. Choi, Y. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.
[46] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2881–2890.
[47] L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1791–1802, May 2019.
[48] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3156–3164.
[49] Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu, "Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4836–4845.
[50] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6077–6086.
[51] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan, "See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6776–6785.
[52] F. Yang, K. Yan, S. Lu, H. Jia, X. Xie, and W. Gao, "Attention driven person re-identification," *Pattern Recognit.*, vol. 86, pp. 143–155, Feb. 2019.
[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[54] A. Dunn and N. Robles, "Polynomial partition asymptotics," Apr. 2017, *arXiv:1705.00384*. [Online]. Available: https://arxiv.org/abs/1705.00384
[55] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[56] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3652–3661.
[57] H. Huang, D. Li, Z. Zhang, X. Chen, and K. Huang, "Adversarially occluded samples for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5098–5107.
[58] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
[59] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," Jul. 2016, *arXiv:1607.08022*. [Online]. Available: https://arxiv.org/abs/1607.08022
[60] L. Zheng, Y. Huang, H. Lu, and Y. Yang, "Pose-invariant embedding for deep person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4500–4509, Sep. 2019.
[61] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," Oct. 2016, *arXiv:1610.02984*. [Online]. Available: https://arxiv.org/abs/1610.02984
[62] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 2590–2600.
[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[64] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Deep attributes driven multi-camera person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 475–491.
[65] S. Liao, Y. Hu, X. Zhu, and S. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2197–2206.
[66] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, "Improving person re-identification by attribute and identity learning," *Pattern Recognit.*, vol. 95, pp. 151–161, Nov. 2019.
[67] R. R. Varior, M. Haloi, and G. Wang, "Gated Siamese convolutional neural network architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 791–808.
[68] G. Ding, S. Khan, Z. Tang, and F. Porikli, "Feature mask network for person re-identification," *Pattern Recognit. Lett.*, to be published, doi: 10.1016/j.patrec.2019.02.015.
[69] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," Aug. 2017, *arXiv:1708.04896*. [Online]. Available: https://arxiv.org/abs/1708.04896
[70] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, "Deep-person: Learning discriminative deep features for person re-identification," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107036.