Neural Computing and Applications
https://doi.org/10.1007/s00521-022-07071-1

ORIGINAL ARTICLE

Person Re-identification with pose variation aware data augmentation

Lei Zhang[1] · Na Jiang[2] · Qishuai Diao[1] · Zhong Zhou[1] · Wei Wu[1]

Received: 28 August 2021 / Accepted: 7 February 2022

Abstract
Person re-identification (Re-ID) aims to match a person of interest across multiple non-overlapping camera views. This is a challenging task, partly because a person captured in surveillance video often undergoes intense pose variations. Consequently, differences in their appearance are typically obvious. In this paper, we propose a pose variation aware data augmentation (PA^4) method, which is composed of a pose transfer generative adversarial network (PTGAN) and person re-identification with improved hard example mining (Pre-HEM). Specifically, PTGAN introduces a similarity measurement module to synthesize realistic person images that are conditional on the pose, and with the original images, form an augmented training dataset. Pre-HEM presents a novel method of using the pose-transferred images with the learned pose transfer model for person Re-ID. It replaces the invalid samples that are caused by pose variations and constrains the proportion of the pose-transferred samples in each mini-batch. We conduct extensive comparative evaluations to demonstrate the advantages and superiority of our proposed method over state-of-the-art approaches on Market-1501, DukeMTMC-reID, and CUHK03 dataset.

Keywords Person re-identification · Data augmentation · Generative adversarial network · Pose transfer · Hard example mining

1 Introduction

Person re-identification (Re-ID) aims to recognize and identify person images captured from non-overlapping camera views [1–6]. It is a fundamentally challenging task because a person's appearance undergoes intense variations in different poses. In recent years, this task has been widely studied for its widespread potential applications in video surveillance and virtual reality [7, 8].

Pose variations commonly exist in person Re-ID and produce severe misalignment between pedestrian images. Existing approaches address this task by learning discriminative descriptors for different poses. Traditional and deep learning approaches [9–13] suffer from over-fitting in close-sets (e.g., DukeMTMC-reID [14], Market-1501 [15], CUHK03 [16]). Consequently, owing to the development of generative adversarial networks (GANs) [17], several data augmentation approaches [18, 19] have been proposed to generate auxiliary samples in different poses to increase image diversity. However, GANs synthesize images that can be confusing because of noise and the appearance features of different persons, which adversely affects Re-ID training.

This study addresses person Re-ID problem from the perspective of pose variation data augmentation. As shown in Fig. 1, we are primarily inspired by the demand for a massive training data scale in person Re-ID. However, manual annotation is very time-consuming and prohibitively expensive. Therefore, if we can use a GAN to add more auxiliary images that are invariant under pose variations, we can (1) deal with the data insufficient issue in person Re-ID and (2) learn features which are invariant under pose variations.

In this paper, we propose a **P**ose v**A**riation **A**ware d**A**ta **A**ugmentation (PA^4) method to regularize model training. PA^4 is composed of a pose transfer generative adversarial

✉ Zhong Zhou
zz@buaa.edu.cn

[1] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, XueYuan Road No. 37, HaiDian District, Beijing 100191, China

[2] Information Engineering College, Capital Normal University, North Xisanhuan Road No. 105, Beijing 100089, China

Published online: 10 March 2022

Fig. 1 Motivation of our approach. Auxiliary training data with diverse pose information are synthesized using a generative adversarial network, which enhances the person Re-ID performance

network (PTGAN) and person re-identification with an improved hard example mining strategy (Pre-HEM). The works that are most relevant to ours [19, 20] propose to synthesize realistic person images using deep person image generation model. They design a GAN specifically to generate additional auxiliary data and introduce the LSR (Label Smooth Regularization) strategy for outliers. The generated images are then combined with real images to improve the performance of person Re-ID. Similarly, the proposed PTGAN and Pre-HEM are independent of each other. PTGAN aims to generate fake images, and Pre-HEM focuses on making full use of the images that are prepared in PTGAN stage. In contrast to references [19, 20], PTGAN considers the influence of pose variation on similarity measurement to generate high-quality auxiliary data in the generative adversarial stage. It pushes pedestrian images in significantly different poses further than those in similar poses to preserve better appearance features and sharper pose. During ReID training, we focus on the invalid samples that are a result of pose variations and propose an effective method of using the auxiliary data instead of obtaining an extended dataset directly.

Specifically, PTGAN introduces a similarity measurement module to generate realistic pedestrian samples that are conditional on poses. The similarity measurement module analyzes the similarity change in pose variations to better preserve appearance features, as shown in Fig. 2. In Fig. 2a, two pedestrian images with the same identity and similar poses appear similar, while the same pedestrian in significantly different poses appears to be quite different, as shown in Fig. 2b. Using the learned pose transfer generative adversarial model, our approach can transfer labeled training images to any given pose. During Re-ID training, we combine the real images and synthesized fake samples to obtain an extended dataset, which is beneficial in achieving a pose-invariant property and further reducing over-fitting. Pre-HEM aims to match the contributions of the real images and the synthesized fake images. It makes full use of the auxiliary data to enhance the discriminability of person Re-ID by (1) improving the manner in which the

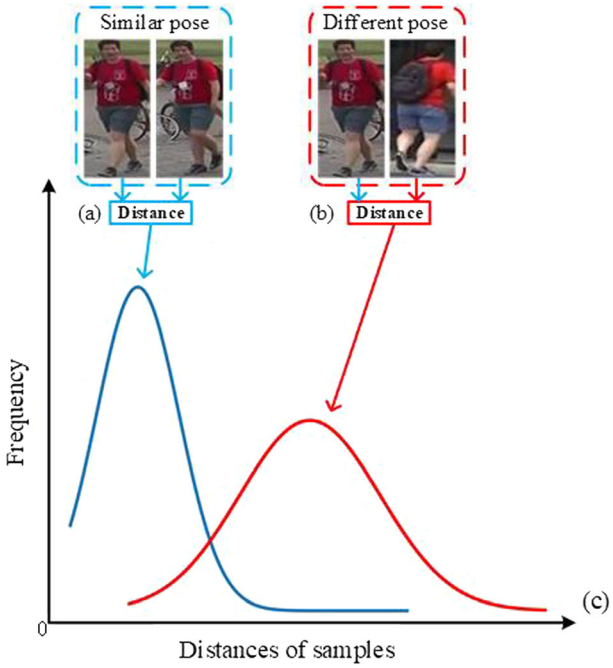


Fig. 2 Examples of distance changes in pose variations. **a**, **b** Indicate the same person in similar and significantly different poses, respectively. **c** Denotes the distance distribution from a pre-training approach, the similarity of the same person in similar poses is generally closer than in significantly different poses

synthesized fake images are used and (2) controlling the proportion of real and synthesized fake images.

Pre-HEM consists of hard example mining with replaceable samples (HEM-RS) and a novel adaptive margin triplet loss (AMTL). Training with triplet loss, we propose HEM-RS to optimize the manner in which samples are used. HEM-RS replaces the invalid examples via pose transfer to augment the number of effective triplets in each mini-batch. As an aid to HEM-RS, we introduce a valid AMTL to balance the contributions of the real and fake images. Specifically, by evaluating the impact of data proportions (ratio of the real images and the synthesized fake samples) on the Re-ID task, we note that the number of synthesized auxiliary images in a training mini-batch

influences performance improvement. In the HEM-RS stage, we replace the inferior examples using the pose transfer model in each mini-batch. As the model converges, the distributions of positive and negative pairs are pushed away, which increases the number of invalid samples. This leads to a steady increase in the proportion of synthesized fake samples and real images. Therefore, we introduce AMTL to constrain the proportion in a training mini-batch to enhance the performance of PA[4].

There are four main contributions of this study:

(1) We introduce a novel PTGAN to produce realistic auxiliary samples for pose variation-aware data augmentation in the Re-ID task. The network learns to push images of pedestrians in significantly different poses further than those with similar poses to maintain better pose and appearance features;
(2) We propose the HEM-RS method to improve the manner in which the synthesized fake samples are used. This method aims to replace the invalid triplets that arise from pose variations via the learned PTGAN model to improve the learning of discriminative features;
(3) We propose an AMTL to restrict the proportion of synthesized fake images in a training mini-batch. It serves as an aid to HEM-RS to make full use of the synthesized fake images and enhance the person Re-ID performance;
(4) To evaluate our approach, we conduct extensive comparisons on Market-1501 [15], DukeMTMC-reID [14], and CUHK03 [16] to verify the performance of our proposed method.

2 Related work

In this section, we provide an overview of three topics: deep learning in person Re-ID, data augmentation-based person Re-ID and GANs.

2.1 Deep learning in Person Re-identification

Person Re-ID is a challenging task that has been extensively studied. Existing approaches are generally divided into two categories: (1) feature representation learning and (2) deep metric learning. We briefly review several of these approaches.

Traditional person Re-ID methods generally seek to construct color features and handcrafted features as representations [21, 22]. These types of features are simple and efficient; however, the level of discrimination is weak. Therefore, deep learning approach is widely used in Re-ID to extract discriminative feature representations. [23] first introduces deep learning to optimize the relationship between a pair of input pedestrian images. Since then, many deep learning approaches have been proposed, including Deepreid [16], IDE [1], and Gated Network [24]. In addition, Imani et al. [25] propose novel features for RGB-D person Re-ID. Pedestrians appear to be quite different because of pose variations that influence the accuracy of similarity metric. Person Re-ID introduces pose information to address this problem. Pose-based approaches usually take advantage of pose information to obtain position features of human body parts, which can align the body parts or extract the local feature. For example, [12, 26] first consider human body structure information in person Re-ID task. This information helps fuse the features that are extracted from the whole and part region pedestrian images. Sun et al. [27] present a Part-based Convolutional Baseline (PCB) that utilizes pose information and design a unified division on the convolutional layer to learn local features. In [28], VPM is introduced through self-supervision. Zhu et al. [29] propose an identity-guided human semantic parsing method using only person identity labels to locate both personal belongings and the pedestrian body parts at the pixel level.

Metric learning attract wide attention before the deep learning era. It aims to determine the similarity between two feature spaces and whether objects in those spaces are similar or dissimilar, such as DNS [11], XQDA [9], KISSME [10]. The approach of deep metric learning is to decrease the distance between similar feature spaces and augment the distance between dissimilar objects. These methods are effective for measuring image similarities in an adaptive manner. However, in large-scale galleries, they may have efficiency problems. With the success of deep learning in recent years, loss function design, which can guide feature representation learning, has replaced the role of metric learning. In existing approaches [14, 30–32], identity loss has been widely utilized. These works treat the person Re-ID training process as a classification task, that is, every identity is an independent class. Verification loss aims to optimize the pairwise relationship, using either binary verification loss or contrastive loss. Binary verification loss [16, 33] distinguishes the positive and negative of an input pedestrian image pair, and contrastive loss [34, 35] accelerates the relative pairwise distance comparison. Triplet loss [36] regards the training process of Re-ID as a retrieval ranking task. It aims to push the negative pairs away and pull the positive pairs closer by a predefined margin. Hu et al. [37] propose a triplet-batch-center loss (TBCL) to improve the generalization ability of loss functions. In [38], an online joint multi-metric adaptation approach is proposed to adopt the person Re-ID models obtained offline for online data.

The approaches mentioned above introduce multifarious schemes, the majority of which construct a complex framework and training objectives with different loss functions for person Re-ID. However, these methods do not consider the effect of pose variation on similarity measures. In addition, existing datasets are closed, and the number of training samples is limited. These issues may cause the learned model to over-fit easily and make the model difficult to converge.

2.2 Data augmentation in Person Re-identification

In contrast to the number of training samples, over-fitting might occur if a convolutional neural network (CNN) is excessively complex. Therefore, many regularization approaches have been introduced in deep learning, such as Batch Norm [39] and Dropout [40]. Batch Norm aims to reduce the internal covariate shift by normalizing layer inputs, and in doing so, accelerate the training of deep neural networks dramatically. Dropout is an effective method to improve CNN models by reducing over-fitting and is widely used in various recognition tasks. It serves as a means of efficiently combining exponentially different neural network architectures and randomly drops out units (hidden and visible) with a fixed probability in the training stage. In addition, the data augmentation method is an effective approach to alleviate the difficulty caused by conflicts between large variations of samples and the limited size of training datasets. McLaughlin et al. [41] propose the generation of various samples by utilizing background and linear transformations to enhance robustness. In [42], Random Erasing randomly chooses areas in the input images to augment occluded examples, which prevents the model from over-fitting. Similarly, Huang et al. [43] introduce adversarially occluded samples to augment the variation in training data.

2.3 Generative adversarial network

GANs [44] have achieved feasible result in synthesizing visually real images in recent years. A GAN is typically composed of a generator and a discriminator. The generator aims at synthesizing visually real images to deceive the discriminator, and the discriminator learns to identify whether a sample is real or fake. Subsequently, DCGAN [17] scales up a GAN with CNNs. In addition, in [45], a conditional generative adversarial network (cGAN) is proposed.

Since GANs have been proposed, many variants have been introduced to deal with diverse tasks, for example, style transfer, image-to-image translation, and pose-to-image generation. Image-to-image translation has attracted considerable attention in the field of computer vision. Isola et al. [46] learn a projection to accomplish style transfer using cGAN. The main drawback of this method is that it requires paired images. To alleviate this issue, CycleGAN is proposed for training with unpaired samples. The pose-to-image task aims to transfer an image into other poses, and style transfer attempts to transfer the style of an input image to another.

Furthermore, some approaches take advantage of the GAN to synthesize auxiliary samples for person Re-ID. In [35], a SPGAN is proposed to maintain self-similarity and domain-dissimilarity. It is composed of a CycleGAN and a Siamese network. In a cross-domain task, Wei et al. [47] introduce person transfer to bridge the domain gap. In addition, GANs have been used for data augmentation. Zheng et al. [48] propose a joint learning framework to couple image generation and person Re-ID training end-to-end.

3 Our proposed method

In this section, we describe our person Re-ID framework, that is, the pose variation aware data augmentation (PA^4) method. Specifically, PA^4 is composed of a PTGAN and Pre-HEM. PTGAN introduces a similarity measurement module to synthesize fake images for data augmentation (Sect. 3.1), Pre-HEM attempts to replace the inferior triplets (Sect. 3.2) and restrains the data proportions (ratio of real images and synthesized fake images) (Sect. 3.3) to enhance the Re-ID performance. The details of our proposed approach are as follows.

3.1 Pose transfer generative adversarial network

In this section, we present our PTGAN, designed from a data augmentation perspective. Given a Re-ID dataset, our method aims to learn pose-to-image translation models to generate pose-rich samples for training. In contrast to the previous GAN-based methods, to maintain appearance consistency in the pose transfer, we introduce a similarity measurement module that analyzes the variations in similarity when poses change. Figure 3 shows an overview of the PTGAN.

GANs focus on the synthesis of visually real images. These images can be easily confused as a result of noise and various appearance features in the fake images, which affects the results of data augmentation in person Re-ID. To address this problem, with the intention of preserving appearance consistency, we introduce a PTGAN, which is composed of three parts: a generator, a discriminator, and a similarity measurement module. The generator is an

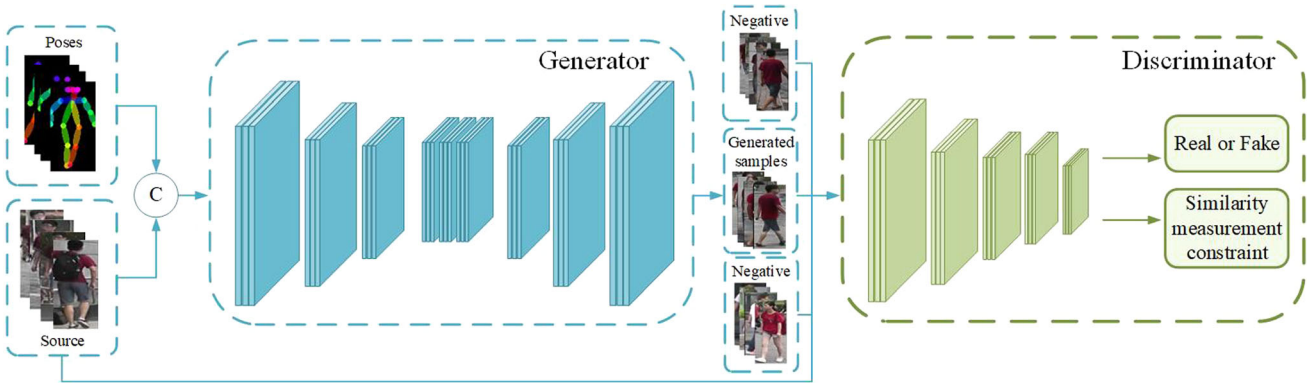


Fig. 3 Outline of our proposed PTGAN. Our method is composed of a generator G and a discriminator D. The generator attempts to transfer a pedestrian into diverse poses. The discriminator learns to identify whether a sample is real or fake

encoder-decoder architecture that aims to replace the pose of input images to synthesize virtually real images. The discriminator seeks to determine whether the input sample is real or fake. Herein, we argue that two images of the same pedestrian in similar poses appear similar, while the same pedestrian in significantly different poses appear quite different. Therefore, we propose a similarity measurement module that aims to improve the appearance features and pose of the synthesized images. PTGAN takes triplet images as inputs, including source, target and target pose images, which are expressed as x_s, x_t and p_t, respectively. Note that the pose images are extracted using a pre-trained OpenPose [49] model. During training, we first pass the triplet images through the generator to synthesize fake images with the mapping function $x_f = G(x_s, p_t)$, where x_f is the generated sample. Then we utilize x_f, x_s, and x_t, which are from the same person, and a random pedestrian image x_n from a different person to construct a quadruplet as the discriminator input. Therefore, in contrast to the general discriminator, our proposed approach also learns to preserve more of the appearance, which is important for person Re-ID. The objective loss of the proposed similarity measurement module is expressed as follows:

$$\begin{aligned}\mathcal{L}_s = \mathbb{E}\Big[\big\|f(x_f) - f(x_s)\big\|_2^2 - \big\|f(x_f) - f(x_n)\big\|_2^2 + \alpha_1\Big]_+ \\ + \mathbb{E}\Big[\big\|f(x_f) - f(x_t)\big\|_2^2 - \big\|f(x_f) - f(x_s)\big\|_2^2 + \alpha_2\Big]_+\end{aligned} \tag{1}$$

For clarity, we utilize $\mathbb{E}$ to denote $\mathbb{E}_{x,p\sim p_{data}}$, where $[\mathcal{L}]_+ = \max(\mathcal{L}, 0)$, and $f(x)$ is the feature extracted from the image x, α_1 and α_2 are margins.

In general, the complete objective loss of the proposed PTGAN is expressed as follows:

$$\mathcal{L}(G, D) = \mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_{L_1} + \lambda_2 \mathcal{L}_s \tag{2}$$

where $\mathcal{L}_{cGAN}(G, D)$ denotes the objective loss of *cGAN*, which can be formulated as:

$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) = \mathbb{E}[\log D(x_s, p_t)] \\ + \mathbb{E}[\log(1 - D(x_s, G(x_s, p_t)))]\end{aligned} \tag{3}$$

$\mathcal{L}_{L_1}$ is L1 loss which can be expressed as:

$$\mathcal{L}_{L_1} = \mathrm{E}\Big[\big\|x_f - G(x_s, p_t)\big\|_1\Big] \tag{4}$$

and λ_1, λ_2 are the weighting factors.

In the training stage, we sample as many source images, target images, and their poses as possible to construct triplets, which are passed through the generator. Then, together with the generator outputs, all images are taken as the discriminator inputs. During testing, we choose a pedestrian image from the person Re-ID dataset and arbitrary poses to synthesize fake training data. Herein, we refer to the synthesized person samples as pose transferred images or fake images. The experimental results verify that our approach can generate realistic and sharper pedestrian images. We combine the original images in the dataset with the synthesized fake images to augment the training set. Because every fake image introduces diverse pose variations, but preserves the content of the corresponding original image, it is recognized as the same identity as the original image. This allows our approach to use the fake images and their corresponding labels, combined with the original images, to train the person Re-ID model.

3.2 Hard example mining with replaceable sample

Our goal is to generate auxiliary samples for Re-ID training, as shown in Fig. 4. We analyze the previous data augmentation approaches that have achieved promising performance, and find that they directly combine the synthesized fake images with the real training data. In this study, we improve the manner in which the synthesized fake samples are used based on the hard example mining strategy. The triplet selection method directly affects the

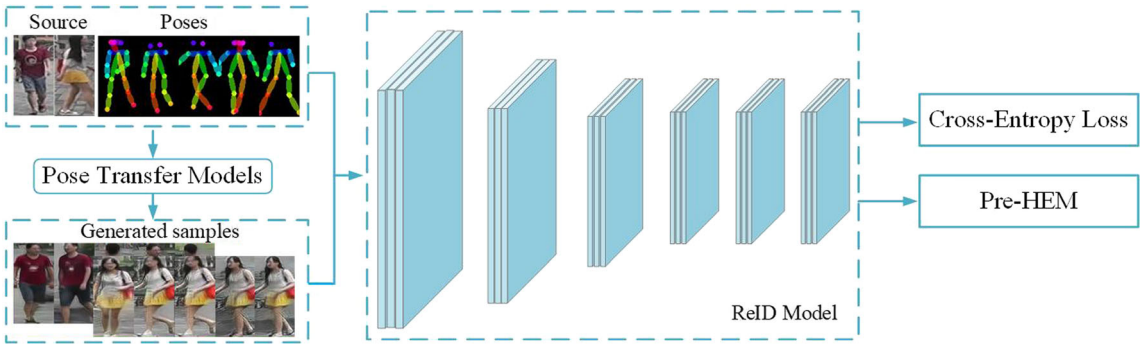


Fig. 4 Method of data augmentation in our proposed approach. The pose transfer models learn to generate realistic person images that are conditional on poses. Real images (top box) and fake images (bottom box) are combined for Re-ID training

performance of the Re-ID model. By analyzing the effect of pose changes on the similarity measurement, we optimize the hard example mining strategy with a replaceable sample (HEM-RS). In contrast to the previous hard example mining methods, HEM-RS attempts to replace the inferior triplets to optimize sample selection and increases the number of valid triplets. In this section, we first briefly present FaceNet [36] and then introduce our proposed approach.

FaceNet first presents triplet loss. This ensures that a pedestrian image (anchor) of a particular identity is closer to all other images (positive) of the same person than to any pedestrian image (negative) of any other identity. Specifically, given an anchor image x_i^a, we choose the hardest positive x_i^p via argmax $\left\|f\left(x_i^a\right)-f\left(x_i^p\right)\right\|_2^2$ and select the hardest negative x_i^n via argmin $\left\|f\left(x_i^a\right)-f\left(x_i^n\right)\right\|_2^2$ similarly. To eliminate model collapse, FaceNet proposes semi-hard exemplars via $\left\|f\left(x_i^a\right)-f\left(x_i^p\right)\right\|_2^2<\left\|f\left(x_i^a\right)-f\left(x_i^n\right)\right\|_2^2$ to replace the hardest negatives . In this study, we focus on triplets that are invalid for person Re-ID training. In Fig. 5, the black dots denote the anchor images, the green and red dots are positive and negative images, respectively, and α is a margin. In Fig. 5a, the positives and negatives do not fall within the margin α. This type of triplet is called an ineffective examples, and is invalid for the training process. To address this issue, using the learned generator G above, we propose a hard example mining strategy with replaceable sample to augment effective triplets to improve the Re-ID performance.

We divide the images in the training dataset into three types: front, back, and side. Figure 5b shows the ineffective triplet $\left\{x_i^a, x_i^p, x_i^n\right\}$ and the corresponding poses $\left\{p^a, p^p, p^n\right\}$. We replace the pose of the negative images (N) with p^a directly with the learned generator and formulate it as:

$$x_i^N = G\left(x_i^n, p^a\right) \tag{5}$$

Thus, we can achieve $\left\|f\left(x_i^a\right)-f\left(x_i^N\right)\right\|_2^2<\left\|f\left(x_i^a\right)-f\left(x_i^n\right)\right\|_2^2$. Therefore, we acquire a harder negative example (image Nega). For the positive image (P), we first identify whether it and the anchor belong to the same subset. If they do, we do nothing. Otherwise, we transfer the positive sample to any one of the other two subsets expressed as x_i^P via the generator G. We can achieve $\left\|f\left(x_i^a\right)-f\left(x_i^P\right)\right\|_2^2>\left\|f\left(x_i^a\right)-f\left(x_i^p\right)\right\|_2^2$. Similarly, we acquire a harder positive sample (Posi). Then, we replace the original ineffective positive and negative with x_i^P and x_i^N to create a valid triplet $\left\{x_i^a, x_i^P, x_i^N\right\}$ for training.

3.3 Learning adaptive margin triplet loss

We evaluate the impact of the ratio of real images and synthesized fake images on Re-ID training, and the results are shown in Fig. 10. We note that the proportion of pose-transferred samples in a training mini-batch influences the performance gain. However, in the HEM-RS stage, we replace the invalid samples with the pose-transferred samples, which changes the data proportions in a mini-batch. Therefore, we propose an adaptive margin triplet loss (AMTL). In contrast to the previous triplet loss, AMTL seek to restrict the number of the pose-transferred samples in a training mini-batch to enhance the data augmentation performance in person Re-ID.

Triplet loss seeks to enforce a margin between the positives and negatives of each triplet. As the model converges, the distributions of positives and negatives are gradually pushed away, as shown in Fig. 6. This reduces the number of valid triplets, which increases the proportion of pose-transferred samples in a mini-batch. Intuitively, a larger margin can constrain the proportion and lead to better performance. However, [50] demonstrate that increasing the margin value of triplet loss does not work well. In contrast to [50], from the data augmentation perspective, we aim to consider the proportion of pose-

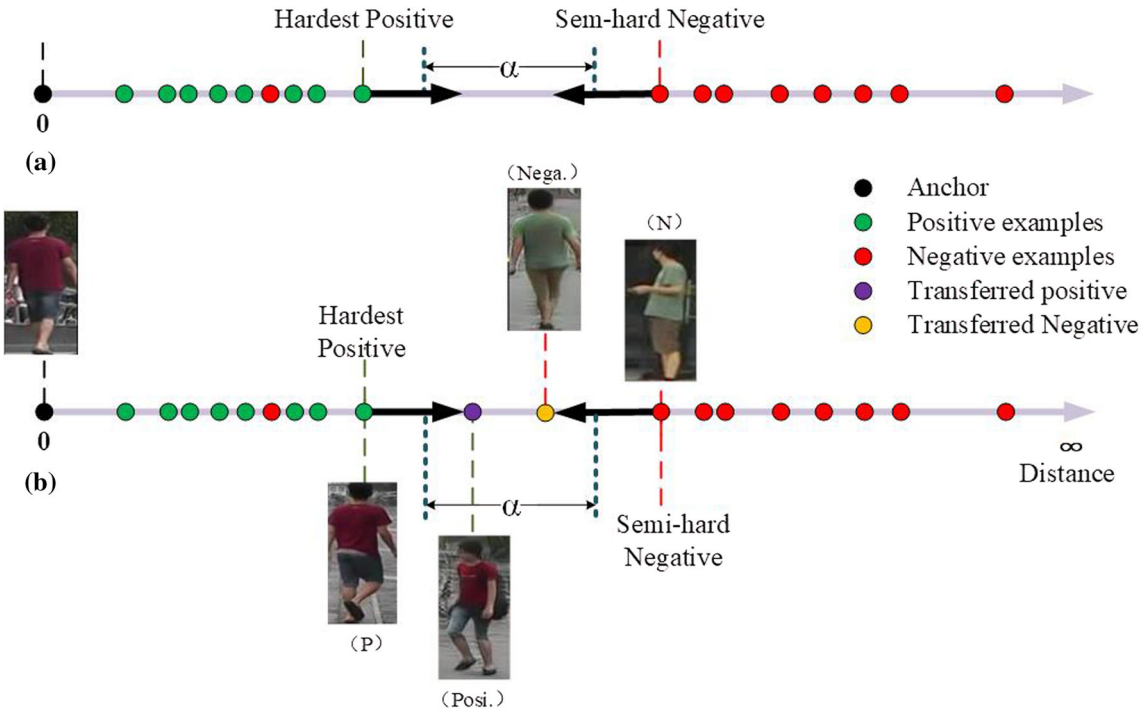


Fig. 5 **a** Ineffective triplets which do not lie within of the margin. This type of triplet is the focus of our approach. **b** Sample of introduced hard example mining with a replaceable sample

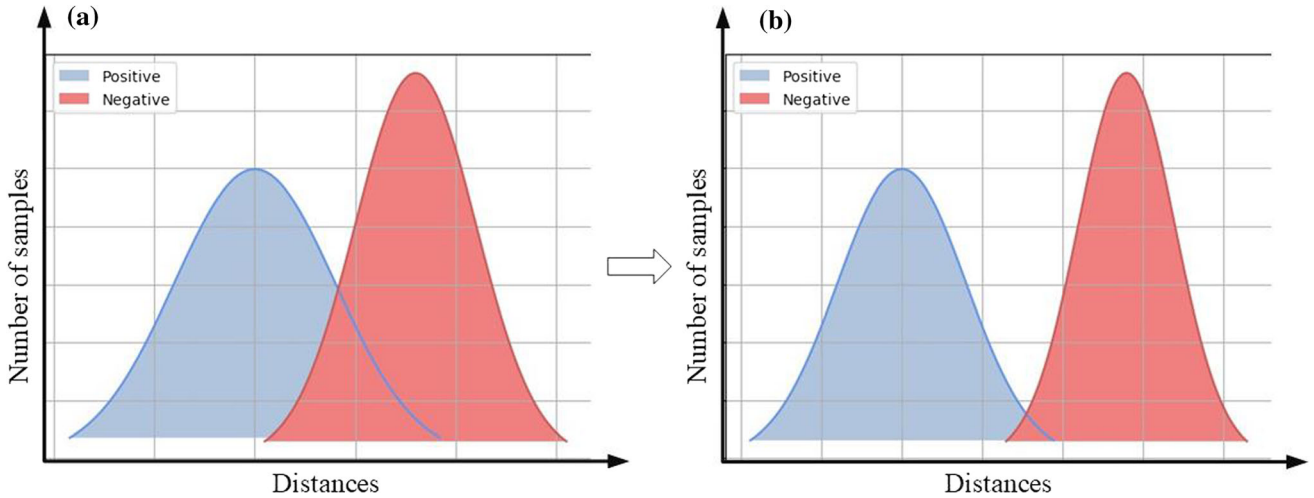


Fig. 6 Motivation of AMTL. The violet curves denote the distribution of the distances of positive sample pairs, while the red curves denote those of negative sample pairs. **a** Original distribution. **b** Distribution after training with the conventional triplet loss (colour figure online)

transferred samples. Therefore, as an aid to HEM-RS, we propose a novel strategy called AMTL. The training procedure is illustrated in Fig. 7. Unlike the original triplet loss, we learn the margin in an adaptive manner. By repeatedly restricting a stable proportion of pose-transferred samples in each mini-batch, the margin between the positives and negatives in the triplet is changed adaptively. Specifically, we consider an important parameter in Pre-HEM, that is, the ratio of $\frac{M}{N}$, where M and N denote the number of real and synthesized fake training images in a mini-batch, respectively. This parameter describes the fraction of the fake images used in the training stage. We seek to balance this ratio in the mini-batch to generate an adaptive margin. Therefore, we introduce a margin penalty, which is expressed as:

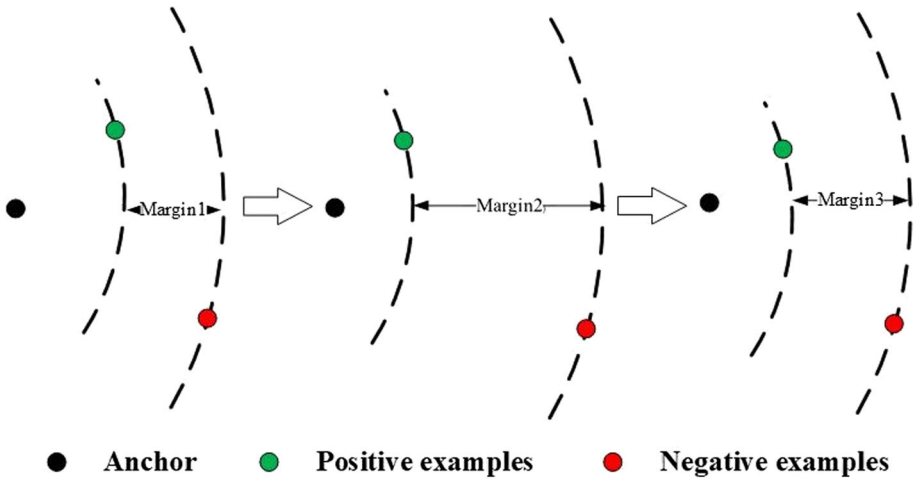


Fig. 7 AMTL training procedure

$$penalty = \begin{cases} \dfrac{M+T}{N-T} & \dfrac{M+T}{N-T} > \beta_2 \\ 0 & \beta_1 < \dfrac{M+T}{N-T} < \beta_2 \end{cases} \tag{6}$$

T is the number of instances that are replaced in HEM-RS, and β_1 and β_2 aim to constrain the proportion. In conclusion, the total adaptive margin can be expressed as:

$$margin = \alpha * (1 + penalty) \tag{7}$$

where α is the margin of the original triplet loss. During training, in the primary stage, there are few replaceable triplets. As the model converges, the distributions of positives and negatives are pushed away, which increases the number of inferior triplets. This leads to a change in the proportion of real images and pose-transferred samples. We calculate a parameter that serves as a penalty for the next mini-batch to constrain the ratio after replacing samples in each mini-batch. Thus, the margin is self-adaptive based on the penalty, which can dynamically constrain the proportion of real images and synthesized fake images.

In the experiments, the proposed approach significantly improves the generalization ability of the Re-ID model.

4 Experimental results

We evaluate the proposed framework, PA[4], on three public datasets. We describe comparisons with state-of-the-art approaches and provide in-depth studies. We perform an ablation experiment to validate each component of our approach. Extensive experiments reveal that our approach consistently generates realistic images and more importantly shows results comparable to overall benchmarks of the competing approaches.

4.1 Datasets and evaluation metrics

Market-1501 [15]. Market-1501 is an image-based person Re-ID dataset. It is composed of 32,668 pedestrian images from 1501 persons observed from 6 camera views. Pedestrians are detected using deformable part model (DPM). The dataset is composed of two parts: 12,936 images from 751 persons for training and 19,732 images from 750 persons for testing. During training, there are 17.2 images per pedestrian on average. During testing, 3368 images from 750 persons are utilized as queries to search the suitable pedestrians.

DukeMTMC-reID [14]. DukeMTMC-reID is composed of 36,411 images from 1404 persons observed from 8 camera views. During training, there are 16,522 images from 702 identities. During testing, there are 25.2 images per pedestrian on average.

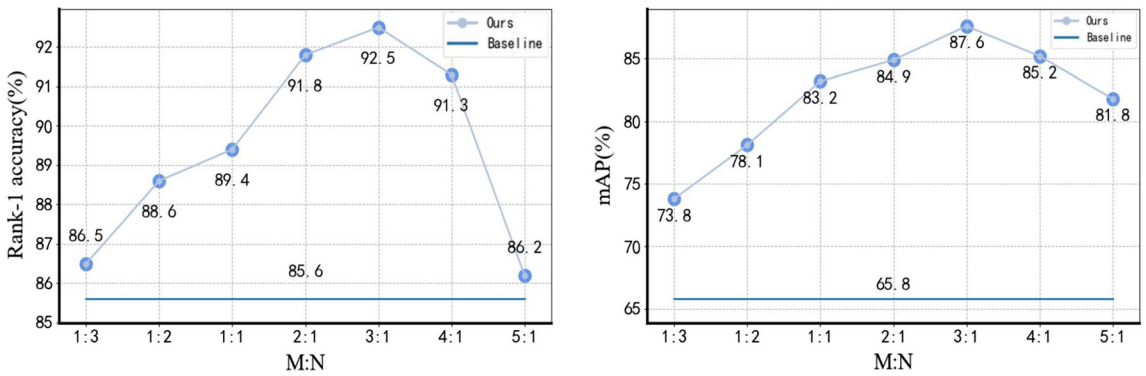
CUHK03 [16]. CUHK03 contains 1467 identities and 28,192 bounding boxes captured by two camera views. Two types of annotations are provided in this dataset: manually labeled pedestrian bounding boxes and DPM-detected bounding boxes. 767 identities are used for training and 700 identities are used for testing. The labeled dataset contains 7368 training images, 5328 gallery and 1400 query images for testing, while the detected dataset contains 7365 images for training, 5332 gallery and 1400 query images for testing.

Evaluation Metrics In our experiments, two evaluation metrics are utilized to appraise the re-ID performance quantitatively. One is Cumulative Matching Characteristic (CMC) [51] at rank-1. The other is mean average precision (mAP) [1]. The CMC scores reflect the retrieval precision, while mAP reflects recall.

Fig. 8 Comparison of the synthesized fake images and real images on Market-1501 across the different approaches. Images in each part are the real image, a random pose, Camstyle [52], Deformable [53], PN_GAN [20], ours, and Ground-Truth from left to right

(a) Market-1501. **(b)** DukeMTMC-reID. **(c)** CUHK03.

Fig. 9 Examples of our proposed PTGAN on three public datasets (Market-1501, DukeMTMC-reID, and CUHK03)



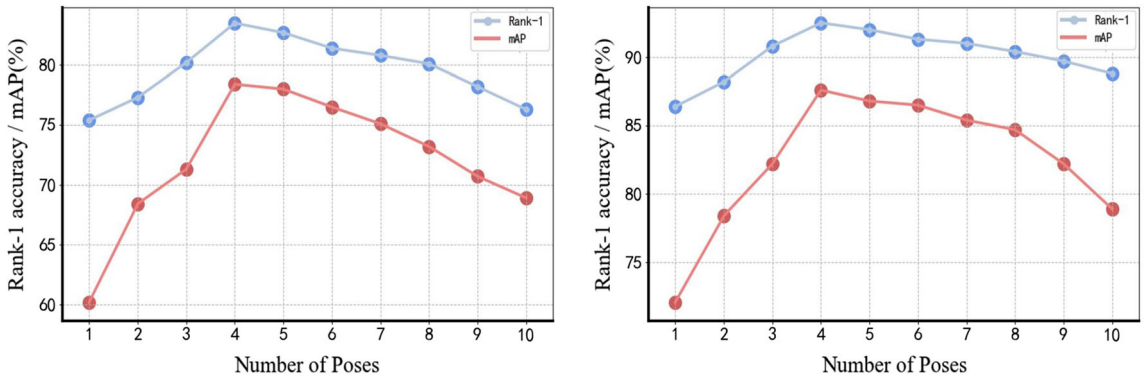
(a) Rank-1. **(b)** mAP.

Fig. 10 Evaluation with different ratios of real data and fake data (M : N) in a training batch on Market-1501

4.2 Implementation details

The proposed approach comprises two components. First, we train the PTGAN to synthesize sharper and more realistic images in diverse poses. Second, we introduce a novel method of synthesized fake images used to extract robust features in person Re-ID, which includes HEM-RS and an AMTL.

Training of PTGAN. Our proposed PTGAN is implemented using the PyTorch framework, which utilizes a network similar to that in PN-GAN [20]. We adopt the learned model in [49] to extract the pose images. The generator inputs are pairs of appearances and poses. We utilize the Adam optimizer to train the network, and the learning rate is 0.0002 for the generator and 0.0001 for the discriminator. The dropout ratio is set to 0.5. In the



(a) Market-1501. **(b)** DukeMTMC-reID.

Fig. 11 Impact of the number of poses on the ReID performance. **a** Market-1501. **b** DukeMTMC-reID

Table 1 Comparison of top matching ranks (%) and mAP (%) on Market-1501 and DukeMTMC-reID

Dataset	Market-1501		DukeMTMC-reID	
Metric	Rank-1	mAP	Rank-1	mAP
KISSME+BoW [15]	44.4	20.8	25.1	12.1
DNS [11]	55.4	29.9	–	–
Verif-Classif [33]	66.1	71.2	–	–
IDE [1]	72.5	46.0	65.2	45.0
OIM [12]	–	–	68.1	47.4
LSRO [14]	84.0	66.1	67.7	47.1
DJL [54]	85.1	65.5	–	–
PT [19]	87.7	68.9	78.5	56.9
PN-GAN [20]	89.4	72.6	73.6	53.2
Camstyle [18]	89.5	71.6	78.3	57.6
FD-GAN [55]	90.5	77.7	80.0	64.5
Part-aligned [56]	91.7	79.6	84.4	69.3
Mancs [57]	93.1	82.3	84.9	71.8
CtF [58]	93.7	84.9	87.6	74.8
FSAM [59]	94.6	85.6	86.4	75.7
DG-net [48]	94.8	86.0	86.6	74.8
GPS [60]	95.2	87.8	88.2	78.7
ABD-net [61]	95.6	88.3	89.0	78.6
SCSN [62]	95.7	88.5	90.1	79.0
Baseline [52]	94.1	85.7	86.2	75.9
PA[4]	**96.2**	**88.7**	**90.7**	**80.1**

Table 2 Comparison of top matching ranks (%) and mAP (%) on CUHK03 dataset

Dataset	Labeled		Detected	
Metric	Rank-1	mAP	Rank-1	mAP
PCB [27]	–	–	63.7	57.5
MGN [63]	68.0	67.4	68.0	66.0
Mancs [57]	69.0	63.9	65.5	60.5
CASN [32]	73.7	68.0	71.5	64.4
MHN [64]	77.2	72.4	71.7	65.4
Auto-ReID [65]	77.9	73.0	73.3	69.3
DenSem [66]	78.9	75.2	78.2	73.1
Pyramid [67]	78.9	76.9	78.9	74.8
Top-DB-Net [68]	79.4	75.4	77.3	73.2
SONA [68]	81.8	79.2	79.1	76.3
ArNet [69]	83.2	80.8	81.4	77.2
Baseline [52]	78.7	72.4	76.5	70.9
PA[4]	**84.3**	**81.3**	**82.4**	**78.1**

Table 3 Evaluation of the effect of our approaches

Dataset	Market-1501		DukeMTMC-reID	
Metric	Rank-1	mAP	Rank-1	mAP
Baseline [52]	94.1	85.7	86.2	75.9
+Fake images(w/o)	94.4	86.1	87.2	76.2
+Fake images(w)	95.1	86.3	88.3	76.7
HEM-RS	95.7	88.0	89.9	78.2
AMTL	96.2	88.7	90.7	80.1
Baseline [18]	85.6	65.8	72.3	51.8
+Fake images(w/o)	88.3	70.5	76.8	67.5
+Fake images(w)	89.5	73.5	78.3	70.5
HEM-RS	90.9	86.0	82.7	75.5
AMTL	92.5	87.6	83.5	78.4

experiments, λ_1 and λ_2 are 1.0 and 0.1, respectively. The margins α_1 and α_2 are set to 1.0 and 0.75, respectively. Our PTGAN models are converged in 20h on the Market-1501 dataset with one 1080Ti GPU.

Improving Re-ID performance with the proposed approach We utilize the generator to transfer each image in the dataset into four classic poses and combine them with the real data to acquire an augmented training set. We

utilize the strong baseline [52] without center loss and IDE* [18] as our baseline. During the training of PA^4, the same setting of the baseline model is used, and we arbitrarily choose M real images and N synthesized fake images in a mini-batch. If not specified, we use $M : N = 3 : 1$. In addition, we optimize the Re-ID model using triplet loss and label smoothing regularization strategy. The margin is set to 0.3 for the augmented fake images. The epsilons are set to 0.5 and 0.1 for real images and the synthesized fake data, respectively. In addition, β_1 and β_1 are set to $\frac{1}{4}$ and $\frac{1}{2}$, respectively. We resize each input image to 256×128 pixels and pad the resized images to 10 pixels with zero values, which are subsequently cropped into 256×128 pixels. In addition, we flip each image horizontally in the training set. The Adam optimizer are used to train the network. The original learning rate is set to 0.0005, and 10 epochs are utilized to increase it to 0.005 linearly. We train the model for 200 epochs.

4.3 Evaluation

Effectiveness of PTGAN In this section, we evaluate the effectiveness of the proposed PTGAN. In Fig. 8, we compare our approach with state-of-the-art approaches that are based on data augmentation, including Camstyle [52], Deformable [53] and PN_GAN [20]. The results verify that the synthesized fake images preserve the main characteristics of the pedestrian. In Fig. 9, we provide six visual comparisons on three public datasets, (a) Market-1501, (b) DukeMTMC-reID, and (c) CUHK03. In each dataset, we visualize two pose transfer examples. For each example, the first two images are the real image and pose. The next two samples are the target image and pose. The final sample is the image synthesized using the proposed approach. The images synthesized using our approach are sharper and more realistic, because we introduce the similarity measurement module, which enables the PTGAN to more effectively learn pose information and appearance features.

Parameter analysis In this paper, we use the learned generator to transfer each image in the dataset into four classic poses. The generated fake images are then combined with real data to obtain an extended training set, in which the number of fake images is much larger than that of real images. Initially, we directly train the person Re-ID model on the extended training set. However, the performance does not achieve the expectation. Excessive noise might produce negative effect. Therefore, we involve a parameter to balance the contribution of effective information and noise in the synthesized fake images, that is, the proportion of $\frac{M}{N}$, where M and N indicate the number of real images and synthesized fake images in each mini-batch, respectively. This parameter indicates the proportion of fake samples used for training. By adopting IDE* [18] as the base network, we evaluate the impact of the ratio in PA^4. Figure 10 shows the experimental results obtained by varying this parameter. It is clear that PA^4 with different ratios $\frac{M}{N}$ has consistently better performance than the baseline model. When the number of synthesized samples is greater than that of real images ($\frac{M}{N} < 1$) in each mini-batch, our method achieves approximately 1% and 8% gains in rank-1 and mAP accuracy, respectively. In contrast, when $\frac{M}{N} > 1$, the proposed method produces more than 6% and 18% gains in rank-1 and mAP accuracy. When $\frac{M}{N} = 3$, the proposed approach achieves the best performance. To fully demonstrate the effectiveness of PA^4, in Fig. 11, we evaluate how the proportion of the synthesized images for each image in the training set influences the performance of the person Re-ID model. For every image, we measure the influence of 1 to 10 synthesized fake images on results. The experimental results for 4

(a) (b)

Fig. 12 Retrieve results of our approach. The yellow bounding boxes indicate the probe images, and others indicate the retrieve results. Among them, the green bounding boxes denote the results from the same persons as the probe images and the red images are from different persons (colour figure online)

samples of each image obtain the best results. As the number of generated fake images increases further, the results decrease slightly.

Performance comparison on public datasets In this section, we compare three types of person Re-ID approaches with our method. As shown in Table 1, the best results are in bold. And the works most similar to our proposed PTGAN are PN_GAN [20] and DG-net [48]. On Market-1501 and DukeMTMC-reID, compared with PN_GAN, our approach achieves gains of 6.8% and 17.1% for rank-1, respectively. Our method also performs better than DG-net by 1.4% and 4.1% for rank-1, respectively. Our approach outperforms other methods that do not utilize a data augmentation strategy. Specifically, rank-1 is improved from 94.1 to 96.2% for Market-1501 and from 86.2 to 90.7% for DukeMTMC-reID without the re-ranking scheme. In addition, we construct extensive experiments to further demonstrate the performance of our proposed method on CUHK03. Note that this dataset is more challenging than the previous ones because it contains fewer samples and has limited viewpoint variations. As shown in Table 2, the proposed PA^4 achieves the optimal rank-1 and mAP scores compared with state-of-the-art methods, which are in bold. Specifically, PA^4 exceeds ArNet by 1.1% and 0.5% in rank-1 and mAP on the labeled dataset, respectively. On the detected dataset, it surpasses ArNet by 1.0% and 0.9% in rank-1 and mAP, respectively.

Effectiveness of the proposed approaches In this section, we verify the effectiveness of the proposed approaches for person Re-ID training. We choose the ResNet50 baselines [52] and [18] for comparison with our approaches. In Table 3, the abbreviations w and w/o are used to denote "with" and "without" the proposed similarity measurement module, respectively. Among both the baselines, our approaches significantly improve the performance. Specifically, on Market-1501, the rank-1 of our method achieves improvements of 2.1% and 6.9%, which represent improvements from 94.1 to 96.2% and 85.6 to 92.5%, respectively. On DukeMTMC-reID dataset, the improvements are 4.5% and 11.2%, which are improved from 86.2 to 90.7% and 72.3 to 83.5%, respectively. This is because the introduced similarity measurement module can enable PTGAN to more effectively learn pose information and appearance features. Pre-HEM replaces the invalid examples and balances the contribution of real and fake images, which can make full use of the generated images.

The performance of our approaches can also be verify in the person Re-ID retrieval results. As shown in Fig. 12a, some pedestrians of different identities that are in similar poses cannot be retrieved correctly. Figure 12b reveals the adaptability of our approaches to pose variation, which further proves that our method can achieve a pose-invariant property.

5 Conclusion

In this study, we propose a pose variation aware data augmentation (PA^4) approach to regularize model training for person Re-ID. This approach is composed of a PTGAN and Pre-HEM. PTGAN introduces a similarity measurement module to synthesize virtually real pedestrian samples that are conditional on poses. Real images and synthesized fake images are combined to form a new extended dataset to enhance person Re-ID from data augmentation perspective. Pre-HEM attempts to improve the manner in which synthesized fake images are used. It focuses on replacing the invalid triplets caused by pose variations with PTGAN and constrains the proportion of real images and synthesized fake images, to make full use of the auxiliary images to enhance the person Re-ID performance. We conduct extensive experiments on three challenging contemporary person Re-ID datasets to validate the performance of our proposed approach. The results prove that our approach can achieve pose-invariant properties and reduce the over-fitting problem.

Acknowledgements This work was supported by the National Key R&D Program of China (Grant No. 2018YFB2100603) and the National Natural Science Foundation of China (Grant No. 61872024). The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Zheng L, Yang Y, Hauptmann AG (2016) Person re-identification: past, present and future. *arXiv preprint*arXiv:1610.02984
2. Wang Z, Ruimin H, Yi Y, Jiang J, Liang C, Wang J (2016) Scale-adaptive low-resolution person re-identification via learning a discriminating surface. In: IJCAI, vol 2, p 6
3. Bedagkar-Gala A, Shah SK (2014) A survey of approaches and trends in person re-identification. Image Vis Comput 32(4):270–286
4. Vezzani R, Baltieri D, Cucchiara R (2013) People reidentification in surveillance and forensics: a survey. ACM Comput Surv (CSUR) 46(2):1–37
5. Gao J, Qing L, Li L, Cheng Y, Peng Y (2021) Multi-scale features based interpersonal relation recognition using higher-order graph neural network. Neurocomputing 456:243–252
6. Hongyang G, Guangyuan F, Li J, Zhu J (2021) Auto-reid+: searching for a multi-branch convnet for person re-identification. Neurocomputing 435:53–66

7. Chen L, Yang H, Qiling X, Gao Z (2021) Harmonious attention network for person re-identification via complementarity between groups and individuals. Neurocomputing 453:766–776
8. Zhao Q (2011) 10 scientific problems in virtual reality. Commun ACM 54(2):116–118
9. Liao S, Yang H, Zhu X, Li Stan Z (2015) Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2197–2206
10. Yan Y, Ni B, Song Z, Ma C, Yan Y, Yang X (2016) Person re-identification via recurrent feature aggregation. In: European conference on computer vision, pp 701–716. Springer
11. Zhang L, Xiang T, Gong S (2016) Learning a discriminative null space for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1239–1248
12. Chi S, Li J, Zhang S, Xing J, Gao W, Tian Q (2017) Pose-driven deep convolutional model for person re-identification. In: Proceedings of the IEEE international conference on computer vision, pp 3960–3969
13. Jiang Na, Liu Junqi, Sun Chenxin, Wang Yuehua, Zhou Zhong, Wei Wu (2018) Orientation-guided similarity learning for person re-identification. In: 2018 24th International conference on pattern recognition (ICPR), pp 2056–2061. IEEE
14. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE international conference on computer vision, pp 3754–3762
15. Zheng L, Liyue SL, Tian SW, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: Proceedings of the IEEE international conference on computer vision, pp 1116–1124
16. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 152–159
17. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint*arXiv:1511.06434
18. Zhong Z, Zheng L, Zheng Z, Li S, Yang Y (2018) Camera style adaptation for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5157–5166
19. Liu J, Ni B, Yan Y, Zhou P, Cheng S, Hu J (2018) Pose transferrable person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4099–4108
20. Qian X, Fu Y, Xiang T, Wang W, Qiu J, Yang W, Jiang Y-G, Xue X (2018) Pose-normalized image generation for person re-identification. In: Proceedings of the European conference on computer vision (ECCV), pp 650–667
21. Mignon A, Pcca FJ A new approach for distance learning from sparse pairwise constraints. In: 2012 IEEE conference on computer vision and pattern recognition
22. Zhao R, Ouyang W, Wang X (2013) Unsupervised salience learning for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3586–3593
23. Yi D, Lei Z, Liao S, Li S Z (2014) Deep metric learning for person re-identification. In: 2014 22nd International conference on pattern recognition, pp 34–39. IEEE
24. Varior R R, Haloi M, Wang G (2016) Gated siamese convolutional neural network architecture for human re-identification. In: European conference on computer vision, pp 791–808. Springer
25. Imani Z, Soltanizadeh H (2018) Histogram of the node strength and histogram of the edge weight: two new features for rgb-d person re-identification. Sci China Inf Sci 61(9):1–14
26. Zhao H, Tian M, Sun S, Shao J, Yan J, Yi S, Wang X, Tang X (2017) Spindle net: person re-identification with human body region guided feature decomposition and fusion. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1077–1085
27. Sun Y, Zheng L, Yang Y, Tian Q, Wang S (2018) Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the European conference on computer vision (ECCV), pp 480–496
28. Sun Y, Qin X, Li Y, Zhang C, Li Y, Wang S, Sun J (2019) Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 393–402
29. Zhu K, Guo H, Liu Z, Tang M, Wang J (2020) Identity-guided human semantic parsing for person re-identification. arXiv preprintarXiv:2007.13467
30. Sun Y, Zheng L, Deng W, Wang S (2017) Svdnet for pedestrian retrieval. In: Proceedings of the IEEE international conference on computer vision, pp 3800–3808
31. Li DW, Huang KQ et al (2018) Adversarially occluded samples for person re-identification
32. Zheng M, Karanam S, Ziyan W, Radke RJ (2019) Re-identification with consistent attentive siamese networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5735–5744
33. Zheng Z, Zheng L, Yang Y (2017) A discriminatively learned cnn embedding for person reidentification. ACM Trans Multimedia Comput, Commun, Appl (TOMM) 14(1):1–20
34. Varior RR, Shuai B, Lu J, Xu D, Wang G (2016) A siamese long short-term memory architecture for human re-identification. In: European conference on computer vision, pps 135–153. Springer
35. Deng W, Zheng L, Ye Q, Kang G, Yang Y, Jiao J (2018) Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 994–1003
36. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 815–823
37. Bin H, Jiwei X, Wang X (2021) Learning generalizable deep feature using triplet-batch-center loss for person re-identification. Sci China Inf Sci 64(2):1–2
38. Zhou J, Bing S, Ying W (2020) Online joint multi-metric adaptation from frequent sharing-subset mining for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2909–2918
39. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprintarXiv:1502.03167
40. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
41. McLaughlin N, Del Rincon JM, Miller P (2015) Data-augmentation for reducing dataset bias in person re-identification. In: 2015 12th IEEE International conference on advanced video and signal based surveillance (AVSS), pp 1–6. IEEE
42. Zhong Z, Zheng L, Kang G, Li S, Yang Y (2020) Random erasing data augmentation. In: AAAI, pp 13001–13008
43. Huang Houjing, Li Dangwei, Zhang Zhang, Chen Xiaotang, Huang Kaiqi (2018) Adversarially occluded samples for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5098–5107
44. Goodfellow I, Pouget-Abadie J, Mirza M, Bing X, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial

nets. In: Advances in neural information processing systems, pp 2672–2680
45. Mirza M, Osindero S (2014) Conditional generative adversarial nets. *arXiv preprint*arXiv:1411.1784
46. Isola P, Zhu J-Y, Zhou T, Efros Alexei A (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
47. Wei L, Zhang S, Gao W, Tian Q (2018) Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 79–88
48. Zheng Z, Yang X, Zhiding Y, Zheng L, Yang Y, Kautz J (2019) Joint discriminative and generative learning for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2138–2147
49. Cao Z, Simon T, Wei S-E, Sheikh Y (2017) Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7291–7299
50. Zhang Y, Zhong Q, Ma L, Xie D, Shiliang P (2019) Learning incremental triplet margin for person re-identification. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 9243–9250
51. Wang X, Doretto G, Sebastian T, Rittscher J, Peter T (2007) Shape and appearance context modeling. In: 2007 ieee 11th international conference on computer vision, pp 1–8. Ieee
52. Luo H, Gu Y, Liao X, Lai S, Jiang W (2019) Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 0–0
53. Siarohin A, Sangineto E, Lathuilière S, Sebe N (2018) Deformable gans for pose-based human image generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3408–3416
54. Zhong Z, Zheng L, Cao D, Li S (2017) Re-ranking person re-identification with k-reciprocal encoding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1318–1327
55. Ge Y, Li Z, Zhao H, Yin G, Yi S, Wang X et al (2018) Fd-gan: pose-guided feature distilling gan for robust person re-identification. In: Advances in neural information processing systems, pp 1222–1233
56. Suh Y, Wang J, Tang S, Mei T, Kyoung ML (2018) Part-aligned bilinear representations for person re-identification. In: Proceedings of the European conference on computer vision (ECCV), pp 402–419
57. Wang C, Zhang Q, Huang C, Liu W, Wang X (2018) Mancs: a multi-task attentional network with curriculum sampling for person re-identification. In: Proceedings of the European conference on computer vision (ECCV), pp 365–381
58. Wang G, Gong S, Cheng J, Hou Z (2020) Faster person re-identification. In: European conference on computer vision, pp 275–292. Springer
59. Hong P, Tao W, Ancong W, Han X, Zheng W-S (2021) Fine-grained shape-appearance mutual learning for cloth-changing person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10513–10522
60. Nguyen Binh X, Nguyen Binh D, Do T, Tjiputra E, Tran Quang D, Nguyen A (2021) Graph-based person signature for person re-identifications. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3492–3501
61. Chen T, Ding S, Xie J, Yuan Y, Chen W, Yang Y, Ren Z, Wang Z (2019) Abd-net: attentive but diverse person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8351–8361
62. Chen X, Canmiao F, Zhao Y, Zheng F, Song J, Ji R, Yang Y (2020) Salience-guided cascaded suppression network for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3300–3310
63. Lin Y, Zheng L, Zhedong Zheng YW, Zhilan H, Yan C, Yang Y (2019) Improving person re-identification by attribute and identity learning. Pattern Recogn 95:151–161
64. Chen B, Deng W, Jiani H (2019) Mixed high-order attention network for person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 371–381
65. Quan R, Xuanyi Dong YW, Zhu L, Yang Y (2019) Auto-reid: searching for a part-aware convnet for person re-identification. In: Proceedings of the IEEE international conference on computer vision, pp 3750–3759
66. Zhang Z, Lan C, Zeng W, Chen Z (2019) Densely semantically aligned person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 667–676
67. Zheng F, Deng C, Sun X, Jiang X, Guo X, Zongqiao Y, Huang F, Ji R (2019) Pyramidal person re-identification via multi-loss dynamic training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8514–8522
68. Quispe R, Pedrini H (2021) Top-db-net: top dropblock for activation enhancement in person re-identification. In: 2020 25th International conference on pattern recognition (ICPR), pp 2980–2987. IEEE
69. Zhang S, Zhang L, Wang W, Xiaofu W (2020) Asnet: asymmetrical network for learning rich features in person re-identification. IEEE Signal Process Lett 27:850–854