Multi-Pseudo Regularized Label for Generated Data in Person Re-Identification

Yan Huang⬤, Jingsong Xu, Qiang Wu, *Member, IEEE*, Zhedong Zheng⬤,
Zhaoxiang Zhang⬤, *Senior Member, IEEE*, and Jian Zhang⬤, *Senior Member, IEEE*

Abstract—Sufficient training data normally is required to train deeply learned models. However, due to the expensive manual process for a labeling large number of images (*i.e.*, annotation), the amount of available training data (*i.e.*, real data) is always limited. To produce more data for training a deep network, generative adversarial network can be used to generate artificial sample data (*i.e.*, generated data). However, the generated data usually does not have annotation labels. To solve this problem, in this paper, we propose a virtual label called Multi-pseudo Regularized Label (MpRL) and assign it to the generated data. With MpRL, the generated data will be used as the supplementary of real training data to train a deep neural network in a semi-supervised learning fashion. To build the corresponding relationship between the real data and generated data, MpRL assigns each generated data a proper virtual label which reflects the likelihood of the affiliation of the generated data to pre-defined training classes in the real data domain. Unlike the traditional label which usually is a single integral number, the virtual label proposed in this paper is a set of weight-based values each individual of which is a number in (0,1] called multi-pseudo label and reflects the degree of relation between each generated data to every pre-defined class of real data. A comprehensive evaluation is carried out by adopting two state-of-the-art convolutional neural networks (CNNs) in our experiments to verify the effectiveness of MpRL. Experiments demonstrate that by assigning MpRL to generated data, we can further improve the person re-ID performance on five re-ID datasets, *i.e.*, Market-1501, DukeMTMC-reID, CUHK03, VIPeR, and CUHK01. The proposed method obtains +6.29%, +6.30%, +5.58%, +5.84%, and +3.48% improvements in rank-1 accuracy over a strong CNN baseline on the five datasets, respectively, and outperforms state-of-the-art methods.

Index Terms—Person re-identification, generated data, virtual label, semi-supervised learning.

Y. Huang, J. Xu, Q. Wu, and J. Zhang are with the Global Big Data Technologies Centre, School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: yan.huang-3@student.uts.edu.au; jingsong.xu@uts.edu.au; qiang.wu@uts.edu.au; jian.zhang@uts.edu.au).

Z. Zheng is with the Centre for Artificial Intelligence, School of Software, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: zhedong.zheng@student.uts.edu.au).

Z. Zhang is with the Research Center for Brain-Inspired Intelligence, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhaoxiang.zhang@ia.ac.cn).

Digital Object Identifier 10.1109/TIP.2018.2874715

I. INTRODUCTION

IN 2014S, Generative Adversarial Network (GAN) was proposed to generate data (images) with perceptual quality [1]. Since then, several improved approaches [2]–[4] were presented to further improve the quality of generated data. However, how to use the data is still an open question. Meanwhile, person re-identification (re-ID) is a challenging task of recognizing a person amongst different camera views. It is a typical computer vision problem that requires sufficient training data to learn a discriminative model. In the past few years, deep learning has demonstrated its performance in person re-ID by producing several state-of-the-art methods [5]–[10]. To this end, sufficient labeled training data is essential to train deeply learned models in a supervised learning fashion. Although some large datasets, *e.g.*, Market-1501 [11], DukeMTMC-reID [12], and CUHK03 [13] have been proposed. However, due to the expensive cost of data acquisition that needs to manually find corresponding labels of pedestrians who appear under different camera views, the number of images per ID in these datasets is still limited.

Using generated data to solve the problem of limited training data is a promising solution. Therefore, we attempt to use unlabeled data generated by GANs to improve the person re-ID performance further. In all existing methods by using GAN, there are two main challenging points in order to assure the better performance: 1) high quality data generated by GAN [2]–[4], 2) a better strategy to use the generated data into the training model [12]. Many works focus on the first point. This paper particularly focuses on the second point. We follow the same pipeline in [12] that incorporates generated data with real data to train deep models in a semi-supervised learning fashion. Compared with previous attempts [2], [14] that perform semi-supervised learning in the discriminator of GANs, sufficient unlabeled generated data will directly participate in training as the supplementary of limited labeled real data in our work.

In 2017s, a related work was first proposed in [12] that introduced a method called Label Smooth Regularization for Outliers (LSRO). This method assigns virtual labels to generated data with a uniform label distribution over all the pre-defined training classes. The uniform distribution considers weights of all the pre-defined training classes equally. More specifically, if the number of pre-defined training class is K, the weight of each class is equally divided into $1/K$.

(a)   (b)   (c) Fitting generated images (a) and (b) into pre-defined training classes using markers 'o' and '+', respectively.
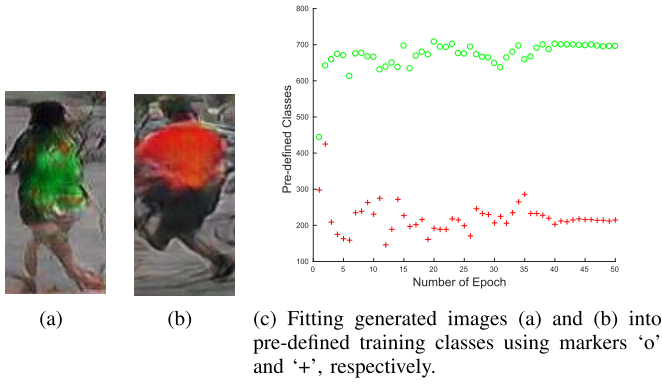
Fig. 1. Label distribution of pre-defined training classes (c) for generated images (a) and (b). Only the maximum predicted probability of pre-defined training classes is activated along with the training process (see (c)). Distinguishable label distributions can be observed between (a) and (b).

By doing so, LSRO shows two undesirable characteristics: 1) On the real data domain, the weights over all pre-defined training classes are identical. 2) On the generated data domain, all data share the same virtual label.

For the first fact, since every individual pre-defined training class of real data has the same weight, the data generated by GAN should be able to embed equal properties of all pre-defined training classes. However, during the actual GAN training process, only a random mini-batch of real data samples are used in each iteration. That is, only certain real data from some classes (not all pre-defined training classes) are used in GAN training in each iteration to generate artificial data following a continuous noise distribution [1], [2]. Consequently, the data distribution between the generated and real data is biased by equally utilizing the weights from all pre-defined training classes in the real data domain. **We need to assign certain type of label to generated data, which can reflect the proper weights of pre-defined training classes in GAN training on the different contributions to new data generated.** For the second fact, it may not be correct to assign the same label to certain different generated data if the generate data has the distinct visual differences. In that case, ambiguous predictions may happen in training. Figure 1(a) and 1(b) show two generated images with red and green clothes respectively. If we fit these two images into pre-defined training classes (only using the maximum predicted probability) through 50 training epochs, distinguishable label distribution can be observed in Figure 1(c). Therefore, using the same virtual label over all the generated data is improper. **We need to dynamically assign different virtual labels to each generated data.**

Although LSRO has demonstrated its effectiveness in [12], the above problems still limit its effectiveness. To solve this problem, a Multi-pseudo Regularized Label (MpRL) is proposed as a virtual label assigned to generated data. Unlike LSRO, main contributions of the proposed MpRL can be summarized in three-fold:

- Compared with LSRO using uniform label distribution, the proposed MpRL assigns each generated data a corresponding label which shows the likelihood of the affiliation of the generated data to all pre-defined

training classes. Thus, the relationship between the generated data and pre-defined training classes can be substantially built, which makes generated data more informative when they incorporate with the real data in training.

- By differentiating the different generated data, MpRL can inherently mitigate of ambiguous prediction in training. Intuitively, different generated data present distinct visual differences and should have different impacts to the training. The proposed method is to embed such characteristics into the training model.

- Qualitative analyses are given to the proposed MpRL. Also, comprehensive quantitative evaluations are carried out to verify the performance of the proposed MpRL not only on large but also on small-scale person re-ID datasets by adapting different CNN models. In addition, we also use two groups of generated data by different GAN models to evaluate the proposed method. Such comprehensive work was not presented in [12].

This paper is organized as follows. We first review some related works in Section II. In Section III, we begin to revisit the state-of-the-art virtual label used on generated data. Then the implementation details of the proposed MpRL are provided. A brief analysis is discussed to demonstrate why MpRL works better in Section IV. The experiments are shown in Section V. The conclusion is in Section VI.

II. RELATED WORK

A. Semi-Supervised Learning

Semi-supervised learning is halfway between supervised and unsupervised learning, which uses both labeled and unlabeled data to perform the learning task. It has been well investigated, and dozens of methods have been proposed in the literature. In image segmentation, a small number of strongly annotated images and a large number of weakly annotated images are incorporated to perform semi-supervised learning [15], [16]. For person identification in TV series, Bauml et al. [17] take labeled data and unlabeled data into account and constrain them in a joint formulation. To tackle multi-label image classification, Luo et al. [18] make use of unlabeled data in semi-supervised learning to boost the performance. In text classification, a region embedding is learned from unlabeled data to produce additional inputs to CNN [19].

Since obtaining training labels is expensive, previous semi-supervised works mainly focus on how to utilize sufficient unlabeled data with accessible labeled data to boost the performance. However, if the real data is scarce or hard to obtain, these methods may useless. Therefore, in this paper, we directly use existing data to generate unlabeled data by GAN. Further, we would like to show that these generated data can help improve discriminative model learning by assigning the proposed MpRL.

Also, several methods have proposed to assign virtual labels to unlabeled data in a semi-supervised learning fashion. In [20] and [14], a new class in the discriminator is taken as the virtual label (all-in-one) assigning to all the unlabeled data produced by the generator of GAN. The all-in-one method simply regards all generated data as an extra class. Let K represents the number of pre-defined training class

in the real data domain, then $K + 1$ is assigned to each generated data. Since these data are generated according to the distribution of real data, they tend to belong to the pre-defined training classes rather than a new one. To solve this problem, the one-hot pseudo label is proposed [21] that can assign a virtual label to generated data without using any extra class. The one-hot pseudo label utilizes the maximum predicted probability of the pre-defined training classes as the virtual label assigning to an unlabeled data. In training, the virtual label is dynamically assigned to the unlabeled data, so that the same data may receive a different label each time when it is fed into the network. Using the one-hot pseudo label, a generated image will be fitted into a specific pre-defined training class along with the training process, which may lead to over-fitting. To address this problem, Zheng *et al.* [12] introduce the LSRO that uses a uniform label distribution to regularize the network training for person re-ID. In this paper, the all-in-one [14], [20], one-hot pseudo [21], and LSRO [12] will be used as our comparison experiments. Amongst them, LSRO achieves the best performance in boosting the re-ID performance. Notably, we call the pseudo [21] as one-hot pseudo in this paper since only one pre-defined training class with the maximum predicted probability is activated in training.

B. Person Re-Identification

The person re-ID is selected to evaluate our MpRL based on two reasons. Firstly, in the past five years, there has been a tremendous increase in this research problem. It has drawn growing interest from academic researches to practical applications [22]. Secondly, compared with other computer vision tasks, acquiring labeled data is expensive for person re-ID. This inspires us to leverage generated data by GAN to solve the limited training data problem. In the past few years, two branches, including traditional and deep learning methods have demonstrated their performance for person re-ID.

In traditional methods, the task of person re-ID can be divided into two modules: feature extraction and metric learning. In feature extraction, Liao *et al.* [23] propose the local maximal occurrence feature to against viewpoint changes and handle illumination variations. Chen *et al.* [24] introduce a mirror representation to alleviate the view-specific feature distortion problem. Zheng *et al.* [11] present a bag-of-words descriptor that describes each person by a visual word histogram. In metric learning, Zheng *et al.* [25] use a relative distance comparison method to minimize the probability of a negative person image pair that has a larger distance than a positive pair. Liao and Li [26] propose logistic metric learning via an asymmetric sample weighting strategy. Li *et al.* [27] employ a locally-adaptive decision function that integrates traditional metric learning with a local decision rule. Yu *et al.* [28] learn an asymmetric metric that projects each view in an unsupervised learning fashion. Zhang *et al.* [29] utilize true motion cues for long-term person re-ID.

Unlike the above traditional methods that are manually designed to handle the person re-ID task. Deep learning discovers more implicit information in matching persons and achieves many state-of-the-art results.

In deep learning methods, to distinguish person appearance at the right spatial locations and scales, Qian *et al.* [6] propose a multi-scale deep learning model to learn discriminative features. Lin *et al.* [7] introduce a consistent-aware deep learning approach which seeks the globally optimal matching. Also, deep features over the full body and body parts are captured from local context knowledge by stacking multi-scale convolutions in [30]. Two-stream network [10], [31], triplet loss network [32], [33] and quadruplet network [34] have been designed for person re-ID.

Zheng *et al.* [8], [9] propose an identification (Identif) CNN. This network takes person re-ID as a multi-classification task, and a CNN embedding is learned to discriminate different identities in training. Beyond that, Zheng *et al.* [10] propose a Two-stream deep neural network. A verification function that separates two input images belonging to the same or different identities is considered to improve the performance of the Identif network further. In testing, the above two networks extract CNN embeddings in the last convolutional layer to compare the similarity between two inputs using squared Euclidean distance. Both of the two CNN networks have been utilized in [12] to investigate the improvement by adding generated data with LSRO virtual labels in training.

In this work, we adopt the Identif network [8], [9] and the Two-stream network [10] to verify the effectiveness of the proposed MpRL. Compared with the previous related work, our MpRL achieves better performance.

Boosting: Some methods have been proposed as a procedure to boost person re-ID performance further. Huang *et al.* [35] formulate person re-ID as a tree matching problem, and a complete bipartite graph matching is presented to refine the final matching result at the top layer of the tree. To study person re-ID with the manifold-based affinity learning, Bai *et al.* [36] introduce a manifold-preserving algorithm plunging into existing re-ID algorithms to enhance the performance. Re-ranking which exploits the relationships amongst initial ranking list in person re-ID has been studied to improve the performance [37]–[39]. Finally, human feedback in-the-loop is required that provides an instant improvement to re-ID ranking on-the-fly [40]–[42].

Unlike the above attempts, in this work, we attempt to use generated data to boost person re-ID performance on off-the-shelf CNNs by incorporating with the proposed MpRL. Although our main contribution is not to produce state-of-the-art person re-ID results. We also try to boost the performance of the Two-stream network [10] to outperform the results of several state-of-the-art methods by using our MpRL.

III. THE PROPOSED MULTI-PSEUDO REGULARIZED LABEL

In this section, we first revisit the state-of-the-art virtual label LSRO [12] for person re-ID. Then MpRL is introduced. Finally, three training strategies are given to the proposed MpRL.

A. LSRO for Person Re-ID Revisit

LSRO assumes that the generated data does not belong to any pre-defined training class and uses the uniform label

distribution on each of them to address over-fitting [12]. LSRO is inspired by label smoothing regularization (LSR) [43] which assigns less confidence on the ground-truth label and assigns small weights to other classes. Formally, giving a generated image g, its label distribution $q_{LSRO}^g(k)$ is defined as follows:

$$q_{LSRO}^g(k) = \frac{1}{K}, \tag{1}$$

where K is the number of pre-defined training classes in the real data domain, $k \in [1, ..., K]$ represents the k-th pre-defined training class. In training, the loss of LSRO to a generated image is defined as follows:

$$l_{LSRO} = -\frac{1}{K} \sum_{k=1}^{K} log(p(X_k)), \tag{2}$$

where X_k represents the output of k-th pre-defined training class, $p(X_k) \in (0, 1)$ is the softmax predicted probability of X_k belonging to the pre-defined training class k, defined as follows:

$$p(X_k) = \frac{e^{X_k}}{\sum_{j=1}^{K} e^{X_j}}. \tag{3}$$

- In Eq.2, the forward loss is as follows:

$$l_{LSRO} = -\frac{1}{K} \sum_{k=1}^{K} log(\frac{e^{X_k}}{\sum_{j=1}^{K} e^{X_j}})$$
$$= -\frac{1}{K} \sum_{k=1}^{K} (X_k) + log(\sum_{j=1}^{K} e^{X_j}). \tag{4}$$

- While, the backward gradient is as follows:

$$\frac{\partial l_{LSRO}}{\partial X_k} = -\frac{1}{K} + \frac{e^{X_k}}{\sum_{j=1}^{K} e^{X_j}}. \tag{5}$$

B. Multi-Pseudo Regularized Label

Like LSRO, we use the proposed MpRL to assign virtual labels to generated data when they are fed into the network. However, unlike LSRO, we do not set the virtual label as a uniform distribution over all pre-defined training classes (*i.e.*, $1/K$). The weights over all pre-defined training classes are different in the proposed MpRL. In this way, a dictionary α is built to record the weights. Compared with the LSRO (see Eq.1), for a generated image g, its label distribution is defined as follows:

$$q_{MpRL}^g(k) = \frac{\alpha_k}{K}, \tag{6}$$

where α_k represents the weight of k-th pre-defined training class in the dictionary α. The reason why different weights are considered in the proposed MpRL will be discussed in Section IV-C. Our MpRL does not belong to a specifi- cally pre-defined training class but is constituted by different weights from each of them. To obtain α_k, we first formulate the set of predicted probabilities $p(X)$ of a generated image over K pre-defined training classes as:

$$p(X) = \{p(X_k)|k \in [1, ..., K]\}. \tag{7}$$

Then, all elements in $p(X)$ are sorted from the minimum to maximum and saved to $p_s(X)$:

$$p_s(X) = \{p_s(X_n)|n \in [1, ..., K]\}, \tag{8}$$

where $p_s(X_1) == min(p(X))$ and $p_s(X_K) == max(p(X))$. α_k is obtained by taking the corresponding index of $p(X_k)$ in the set of $p_s(X)$:

$$\alpha_k = \phi_{p_s(X)}(p(X_k)), \tag{9}$$

where $\phi_{p_s(X)}(\cdot)$ returns the index of $p(X_k)$ in $p_s(X)$. By doing so, the corresponding relationship between real data and a generated image is built by utilizing different weights obtained through the predicted probabilities over all pre-defined training classes. Combining Eq.6 with Eq.9, the proposed MpRL can assign a multiple distributed virtual label to a generated image g when it is fed into the network in training:

$$q_{MpRL}^g = \left\{\frac{\alpha_k}{K}|k \in [1, ..., K]\right\}, \tag{10}$$

We call our method 'multi-pseudo' label because compared with the one-hot pseudo label that only the maximum predicted probability is activated, all the predicted probabilities are used in MpRL. To address over-fitting (*e.g.*, after several training iterations some weights from pre-defined training classes will become larger, while others may decrease to a pretty small value), Eq.10 regularizes the gap between two contiguous weights to $1/K$. In this way, the proposed MpRL retains the weights from all pre-defined training classes, even though some of them may not or just producing a tiny contribution to the generated data.

Combining the generated data with real data in training, we define the cross-entropy loss of the proposed MpRL as follows:

$$l_{MpRL} = -(1-y)log(p(X_c))$$
$$-y \cdot \lambda \cdot \sigma \sum_{k=1}^{K}(\frac{\alpha_k}{K} \cdot log(p(X_k))), \tag{11}$$

where c represents the ground-truth label of a real image, $\frac{\alpha_k}{K}$ is defined in Eq.6. λ is the parameter for the trade-off between losses of generated and real data. If not specified, we set λ to be 1. σ is a normalization factor. In Eq.11, if we sum up weights over K per-defined training classes ($\sum_{k=1}^{K} \frac{\alpha_k}{K}$), the total weight equals to $\frac{(1+K) \cdot K}{2}$. Therefore, to normalize weights over K pre-defined training classes, σ is set to $\frac{2}{1+K}$.

For a real image $y = 0$, Eq.11 is equivalent to softmax loss. For a generated image $y = 1$, only the MpRL is used. Overall, the network has two types of losses: one for real data and the other for generated data.

- In Eq.11, the forward loss is as follows:
  For a real image, $y = 0$:

$$l_{MpRL} = -log(\frac{e^{X_c}}{\sum_{j=1}^{K} e^{X_j}})$$
$$= -X_c + log(\sum_{j=1}^{K} e^{X_j}). \tag{12}$$

For a generated image, $y = 1$:

$$l_{MpRL} = -\lambda \cdot \sigma \sum_{k=1}^{K} (\frac{\alpha_k}{K} \cdot log(\frac{e^{X_k}}{\sum_{j=1}^{K} e^{X_j}}))$$

$$= -\lambda \cdot \sigma \sum_{k=1}^{K} (\frac{\alpha_k}{K} X_k - \frac{\alpha_k}{K} log(\sum_{j=1}^{K} (e^{X_j}))). \quad (13)$$

- While, the backward gradient is as follows:

For a real image, $y = 0$:

$$\frac{\partial l_{MpRL}}{\partial X_c} = -1 + \frac{e^{X_c}}{\sum_{j=1}^{K} e^{X_j}}. \quad (14)$$

For a generated image, $y = 1$:

$$\frac{\partial l_{MpRL}}{\partial X_k} = -\lambda \cdot \sigma \cdot \frac{\alpha_k}{K} (1 - \frac{e^{X_k}}{\sum_{j=1}^{K} (e^{X_j})}). \quad (15)$$

C. Training Strategy

To further investigate the effectiveness of the proposed MpRL, three different training strategies, including one static (constant virtual labels) and two dynamic (iteratively updated) approaches are introduced. Descriptions are as follows:

- **Static MpRL (sMpRL).** The sMpRL is assigned to each generated data before training the network. We use a pre-trained Identif network (see Section V-B2) to assign sMpRL. Specifically, 1) the Identif network is pre-trained on a target re-ID dataset; 2) Eq.3 is utilized to calculate the predicted probability over K pre-defined training classes for each generated data; 3) Eq.10 is used to assign each generated data with a sMpRL, and it remains unchanged during the whole training process. This implementation is similar to the LSRO except that we consider different weights over all pre-defined training classes instead of regarding them equally.

- **Dynamic MpRL-I (dMpRL-I): Dynamically Update MpRL from scratch.** During training, dMpRL-Is are dynamically assigned to each generated data using Eq.10, and they will be updated iteratively to change the likelihood of the affiliation of the generated data to all pre-defined training classes. Therefore, the same generated data may receive a different dMpRL-I each time when it is fed into the network. This dynamic progress starting from the first mini-batch fed into the network until the training is completed. Notably, generated data will assign random dMpRL-Is if they are involved in the first training iteration.

- **Dynamic MpRL-II (dMpRL-II): Dynamically Update MpRL from the intermediate point.** We try to assign dMpRL-IIs to generated data after 20 epochs when the CNN model becomes relatively stable, and also they will be updated iteratively. That is, in Eq.11 $y = 0$, and until after 20 epochs, it is set to 1. Also, the loss is set to 0.1 and 1 for the generated and real data respectively. Therefore, under this training strategy, λ is set to 0.1 in Eq.11. The detailed training strategy is shown in Algorithm 1.

Algorithm 1 The Training Strategy of the dMpRL-II: Dynamically Update MpRL From the Intermediate Point to Change the Likelihood of the Affiliation of the Generated Data to All Pre-Defined Training Classes Iteratively

---

Input: Real data set: R;
    Generated data set: G;
    Merged data set: $D = R \cup G$;
    Loss for the real data set: l_1;
    Loss for the generated data set: l_2.

1 **for** *number of training epochs* **do**
2   Shuffle D ;
3   **for** *number of training iterations in each epoch* **do**
4   Set $l_1 = 0$, $l_2 = 0$;
5   Sample minibatch from $D \rightarrow D'$;
6   Select real data R' from D';
7   Set $y = 0$ in Eq.11;
8   Calculate loss l_1 for R';
9   **if** *number of epochs* $\geq$ *20* **then**
10   Select generated data G' from D';
11   Assign MpRL to G' using Eq.10;
12   Set $y = 1$ in Eq.11;
13   Calculate loss l_2 for G';
14   Calculate the final loss $= l_1 + l_2 \times 0.1$;
15   Backward propagation;
16   Update parameters;
17 **final**;

IV. WHY MULTI-PSEUDO REGULARIZED LABEL WORKS BETTER?

We use the all-in-one [14], [20], one-hot pseudo [21], and LSRO [12] as our comparison experiments. Figure 2(b), (c) and (d) respectively illustrate the label distributions. Given a generated image, a new label that does not belong to any pre-defined training class is assigned to it by using the all-in-one (see Figure 2(b)). Using the one-hot pseudo, only the maximum predicted probability of pre-defined training classes is used as a virtual label (see Figure 2(c)). A uniform label distribution $1/K$ is utilized by the LSRO (see Figure 2(d)). The label distribution of MpRL is illustrated in Figure 2(e). The $\alpha = \{\alpha_k | k \in [1, ..., K]\}$ (defined by Eq.6 to Eq.9) is used to record the different weights over all the pre-defined training classes. In this section, the differences between MpRL and the other three virtual labels will be discussed in three aspects: 1) one-hot vs. multiple label distribution, 2) the same vs. different virtual labels, and 3) the same vs. different weights from pre-defined training classes. Three qualitative discussions are given to support the MpRL, while corresponding numerical evidence will be provided in experiments (see Section V).

A. One-Hot vs. Multiple Label Distribution

The all-in-one and one-hot pseudo are two standard one-hot labels that assign a virtual label to each generated data outside (using a new class) and inside pre-defined training classes, respectively. Compared with the multiple label distribution that retains information from all pre-defined training classes,
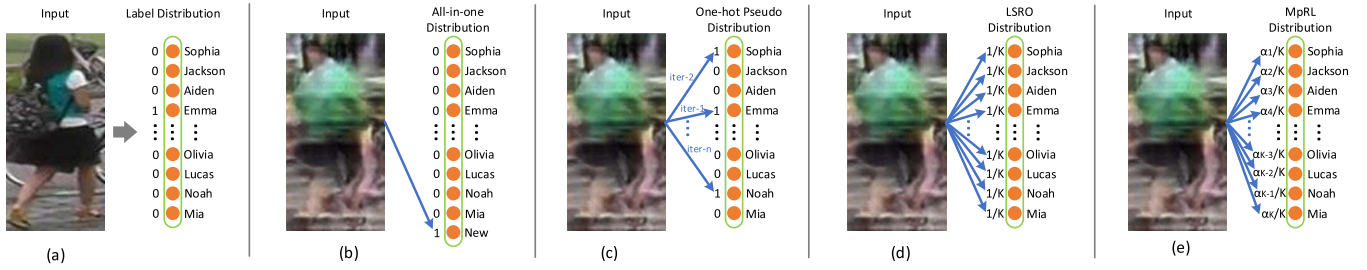
Fig. 2.   The label distributions of real and generated data. The ground-truth label is assigned to the real data (a). For a generated image, all-in-one (b) assigns a new label to it. One-hot pseudo (c) uses only one pre-defined training class with maximum predicted probability. LSRO (d) uses a uniform label distribution, while the proposed MpRL (e) considers different weights over all pre-defined training classes.

Fig. 3.   Examples of generated data and their corresponding representations in the real data domain. The left side shows four generated data with distinct visual differences (in red, yellow, white and green clothes). For each generated data, the right side gives ten nearest representations which represent each pre-defined training class in the real data domain.

the one-hot distribution may produce inadequate regularization power in training which is critical to prevent the network from over-fitting. In the one-hot distribution, the network may mislead to learn a discriminative feature on an infrequent data sample or class. While using multiple distributed label, the network will discourage to be tuned towards one particular class and thus reduces the chance of over-fitting [12], [43]. We device MpRL following the multiple label distribution. In Section V-F, corresponding experiments demonstrate the superiority by using the multiple label distribution.

B. The Same vs. Different Virtual Labels

Two strategies can be used to assign virtual labels to generated data: 1) using the same virtual label over all the generated data, 2) assigning different virtual labels to different generated data. Both all-in-one and LSRO follow the first strategy, while one-hot pseudo and MpRL go with the second one. Compared with the second strategy, assigning each generated data with the same label potentially leads to ambiguous predictions in training. In Figure 3, four different generated images with distinct visual differences (in red, yellow, white and green clothes) are given to find their top ten nearest representations which represent different pre-defined training classes in the real data domain. The four groups visually show clear differences. If we still assign the four generated images with the same virtual label in training, consequently, the network

TABLE I
COMPARISON BETWEEN VIRTUAL LABELS

Method	Label Distribution	Label Assigning	Weights on Pre-defined Classes
All-in-one [20], [14]	One-hot	Same	–
Pseudo [21]	One-hot	Different	–
LSRO [12]	Multiple	Same	Same
MpRL (ours)	Multiple	Different	Different

will mislead in identifying them. The proposed MpRL follows the second strategy that assigns each generated data with a weight-based virtual label according to different predicted probabilities. Corresponding experiments can be found in Section V-F to show that by assigning different virtual labels to generated data, the proposed MpRL can achieve better performance.

C. The Same vs. Different Weights From Pre-Defined Training Classes

LSRO assumes that the weight from each pre-defined training class should be identical. Thus a generated image is assumed to have the capability to simulate the distribution of all the pre-defined training classes equally. This is impractical when considering the actual GAN training process, for two reasons (details can be found in [1], [2]). First, in each training iteration, a mini-batch of random noise is fed into a generator to simulate another mini-batch of real data. This indicates that the generation capability of the inputs is limited in a small scope, specifically, within a mini-batch of real data. Secondly, normally the input random noise obeys a continuous distribution, *e.g.*, Gaussian distribution, while the distribution of real data is discrete. Consequently, complete mapping does not exist between inputs of the generator and the real data domain. Due to the above two reasons, bias exists between distributions of the output of the generator (generated data) and real data. Therefore, a generated image does not have the capability to embed equal properties of the distributions of all pre-defined training classes in the real data domain. To address the problem of LSRO, the proposed MpRL uses different weights from pre-defined training classes (see Section III-B). In experiments, we observe that the proposed MpRL can outperform the state-of-the-art LSRO method on three large and two small-scale person re-ID datasets (see Section V-F).

Through the above discussion, Table I summaries the properties between the proposed MpRL and other labels. Our MpRL takes the advantages of all the properties and achieves

Fig. 4.   Examples of generated (by DCGAN [2]) and real person images. (a)-(d) show the generated person images (first two rows) and real person images (the third row) on Market-1501, DukeMTMC-reID, CUHK03, VIPeR, and CUHK01, respectively.

better performance than others. The numerical evidence which shows the superiority of MpRL will be presented in Section V.

V. EXPERIMENTS

In this section five person re-ID datasets are used to verify the effectiveness of the proposed MpRL, including three large-scale datasets (Market-1501 [11], DukeMTMC-reID [12], and CUHK03 [13]) and two small-scale datasets (VIPeR [44] and CUHK01 [45]). We mainly evaluate the proposed MpRL using Market-1501 and VIPeR since they belong to different scales.

A. Person Re-ID Datasets

Market-1501 is collected from six cameras in Tsinghua University. It contains 12,936 training images and 19,732 testing images. The number of identities is 751 and 750 in the training and testing sets respectively. There is an average of 17.2 images per training identity. All the pedestrians are detected by the deformable part model (DPM) [46]. Both single and multiple query settings are used.

DukeMTMC-reID is collected from eight cameras. The original dataset is used for cross-camera multi-target pedestrian tracking [47]. We use the re-ID version benchmark [12] to evaluate our method. It contains 1,404 identities in which 702 identities for training and the remaining 702 identities for testing. The total training images are 16,522. In the testing set, one query image for each identity is picked up in each camera and put the remaining images in the gallery. There are 2,228 query images and 17,661 gallery images for the 702 testing identities.

CUHK03 is captured by six cameras on the CUHK campus. It contains 14,097 images of 1,467 identities, and each identity is observed by two disjoint camera views. There is an average of 9.6 training identity images in this set. CUHK03 contains two image settings: one is annotated by hand-drawn bounding boxes, and the other is produced by the DPM [46]. We use the detected bounding boxes and the single query setting.

VIPeR is a small-scale dataset that only contains 632 identities. Each identity has two images which are observed by two different camera views. There are 1,264 images in which half identities are for training and the remaining is for testing.

CUHK01 has 971 identities, each with four images captured from two disjoint camera views. There are totally 3884 images.

Two different settings can be found on this dataset: 1) 871 identities for training, and 2) 485 identities for training. We choose the latter one to verify the effectiveness of our approach since the scale of training data is much more limited than the former one. We use the multiple query setting in testing.

B. Experimental Setup

1) GAN Models for Generating Data: GAN simultaneously trains two models: a generator that simulates the distribution of real data, and a discriminator that estimates the probability that a image comes from the real data set rather than the generator [1]. We mainly use the DCGAN model [2] and follow the same settings in [12] for fair experimental comparisons. For the generator, 100-dim random noise is fed into a linear function to produce a tensor with size of $4 \times 4 \times 16$. Then, five deconvolutional functions with a kernel size of 5×5 and a stride of 2 are used to enlarge the tensor. A rectified linear unit and batch normalization are used after each deconvolution. Also, one deconvolutional layer with a kernel size of 5×5 and a stride of 1 are added to fine-tune the result followed by a tanh activation function. Finally, $128 \times 128 \times 3$ sized images can be generated. The input of the discriminator includes generated and real data. Five convolutional layers are used to classify whether the generated image is fake with a kernel size of 5×5 and a stride of 2. In the end, a fully-connected layer is added to perform a binary classification.

The Tensorflow [48] and DCGAN packages are used to train the GAN model. Only data from the training set are used. All the images are resized to 128×128 and randomly flipped before training. The adam stochastic optimization [49] is used with parameters $\beta 1 = 0.5, \beta 2 = 0.99$. The training stops after 30 and 60 epochs on large and small-scale re-ID datasets respectively. During testing, a 100-dim random vector ranged in $[-1, 1]$ with Gaussian distribution is fed into the GAN to generate a person image. Finally, all the generated data are resized to 256×256 and will be used to train CNN models with the proposed MpRL.

Figure 4 illustrates the generated and real data on the five different re-ID datasets. Although the generated data can be easily recognized as fake by human, they remain effective in improving the performance by adding the proposed MpRL as virtual labels in experiments.
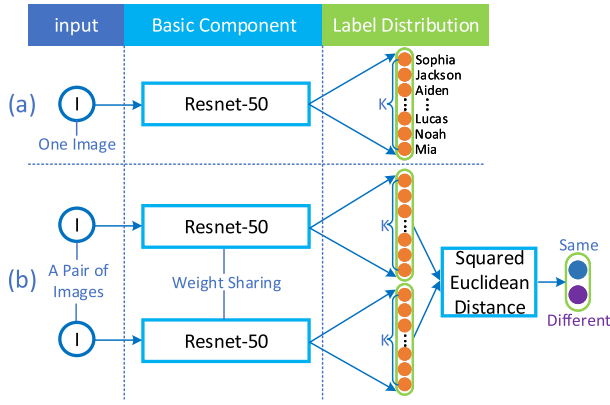
Fig. 5. (a) is the Identif network presented in [8] and [9], (b) is the Two-stream network introduced in [10]. Both networks use resnet-50 as a basic component of CNN.

2) CNNs for Evaluation:

We adopt two CNNs to evaluate the proposed MpRL. These two networks have been used to evaluate the performance of the all-in-one, one-hot pseudo, and LSRO labels in [12]. The first is an Identif network [8], [9] that takes person re-ID as a multi-classification task according to the number of pre-defined training classes in the real data domain. We use the Identif network as a baseline when only the real data is used. Furthermore, to compare the performance of different virtual labels, generated images are incorporated into real images as inputs. The second one is a Two-stream network [10] that combines the Identif network with a verification function to train the network. Given two input images, the verification function will classify them into two classes (belong to the same or different identities). We use this Two-stream network to achieve better results by adding generated data in training. In our experiment, both Identif and Two-stream networks use the pre-trained resnet-50 [50] as a basic component. We change the last fully-connected layer to have K neurons to predict K classes, where K is the number of pre-defined training classes. Since we do not need to add extra classes on generated data by using the proposed MpRL, the last fully-connected layer remains K neurons in training.

Figure 5(a) and Figure 5(b) respectively show the Identif and Two-stream networks. MpRLs are assigned to generated data when they are fed into the network. In the Two-stream network, squared Euclidean distance is used as a similarity measure between two outputs of the K neurons, and parameters are shared between the two resnet-50 components. Since generated images are unlabeled data that do not belong to any classes, only real images participate in the verification function.

The Matconvnet [51] package is used to implement the Identif network and the Two-stream network. All the images are resized to 256×256 before being randomly cropped into 224×224 with random horizontal flipping. A dropout layer is inserted before the final convolutional layer of the resnet-50. The dropout rate is set to 0.75 for Market-1501 and DukeMTMC-reID, and 0.5 for CUHK03, VIPeR, and CUHK01. We modify the fully-connected layer of resnet-50 to have 751, 702, 1,367, 316 and 485 neurons for Market-1501, DukeMTMC-reID, CUHK03, VIPeR, and CUHK01 respectively. For the verification function in the Two-stream network,

TABLE II
PERFORMANCE OF THE IDENTIF AND TWO-STREAM NETWORKS. ONLY THE REAL IMAGES ARE USED. RANK-1 ACCURACY AND mAP ARE LISTED

Dataset	CNN	mAP	rank-1
Market-1501	Identif [8], [9]	52.68%	74.08%
	Two-stream [10]	64.09%	81.83%
DukeMTMC-reID	Identif [8], [9]	42.20%	61.94%
	Two-stream [10]	51.04%	72.62%
CUHK03	Identif [8], [9]	68.36%	63.10%
	Two-stream [10]	85.20%	81.88%
VIPeR	Identif [8], [9]	46.38%	40.76%
	Two-stream [10]	59.38%	51.84%
CUHK01	Identif [8], [9]	63.60%	65.33%
	Two-stream [10]	76.38%	77.78%

a dropout layer with a rate of 0.9 is adopted after the similarity measure. Stochastic gradient descent is used on both networks with momentum 0.9. The learning rate is set to 0.1 and decay to 0.01 after 40 epochs, and we stop training after the 50-th and 60-th epochs on the Identif network and Two-stream network, respectively. For the Identif network, the batchsize is set to 64. For the Two-stream network, the batchsize is set to 32 and 48 on large and small-scale re-ID datasets respectively. During testing, for both networks, a 2,048-dim CNN embedding in the last convolutional layer of the resnet-50 is extracted. The similarity between two images is calculated by a squared Euclidean distance before ranking. Naturally, the small-scale dataset cannot train a network from the scratch. In order to build certain initial network parameters, we first use the three large scale re-ID datasets to pre-train two evaluation CNN models which we use in our experiments (*i.e.*, the Identif network and the Two-stream network). Then, small datasets VIPeR and CUHK01 along with the generated data (based on the proposed method in this paper) are to fine-tune the network.

C. The CNN Performance

Using the experimental setup in Section V-B, we train the Identif and Two-stream networks on Market-1501, DukeMTMC-reID, CUHK03, VIPeR and CUHK01, respectively. Table II shows the experimental results using the real data only. With the Identif (Two-stream) network, we obtain the rank-1 accuracy 74.08% (81.83%), 61.94% (72.62%), 63.10% (81.88%), 40.76% (51.84%), and 65.33% (77.78%) on Market-1501, DukeMTMC-reID, CUHK03, VIPeR, and CUHK01, respectively. The result shown in Table II is a baseline, and our goal is to improve the performance of the two networks by using the proposed MpRL with generated data in training.

D. Generated Data Improve the Performance of the Identif Network

We first give the result of the Identif network to evaluate our MpRL. Since the performance of the Two-stream network is higher, it will be used to compare with some state-of-the-art methods with the proposed MpRL in Section V-H. Table III shows that when we add 24,000 GAN generated images to train the Identif network on three large-scale datasets, our dMpRL-II significantly improves the re-ID performance

Dataset	Method	mAP	rank-1
Market-1501	baseline	52.68%	74.08%
	LSRO [12]	56.33%	78.21%
	Improvement	*+3.65%*	*+4.14%*
	dMpRL-II	58.59%	80.37%
	Improvement	**+5.91%**	**+6.29%**
DukeMTMC-reID	baseline	42.20%	61.94%
	LSRO [12]	46.66%	66.92%
	Improvement	*+4.46%*	*+4.98%*
	dMpRL-II	48.58%	68.24%
	Improvement	**+6.38%**	**+6.30%**
CUHK03	baseline	68.36%	63.10%
	LSRO [12]	71.60%	66.30%
	Improvement	*+3.24%*	*+3.20%*
	dMpRL-II	73.48%	68.68%
	Improvement	**+5.12%**	**+5.58%**
VIPeR	baseline	46.38%	40.76%
	LSRO [12]	49.94%	43.57%
	Improvement	*+3.56%*	*+2.81%*
	dMpRL-II	52.25%	46.60%
	Improvement	**+5.87%**	**+5.84%**
CUHK01	baseline	63.60%	65.33%
	LSRO [12]	64.47%	66.98%
	Improvement	*+0.87%*	*+1.65%*
	dMpRL-II	66.37%	68.81%
	Improvement	**+2.77%**	**+3.48%**

on the strong baseline of Market-1501. The improvements are +5.91% (from 52.68% to 58.59%) and +6.29% (from 74.08% to 80.37%) in mAP and rank-1 accuracy, respectively. For DukeMTMC-reID, +6.38% (from 42.20% to 48.58%) and +6.30% (from 61.94% to 68.24%) improvements are obtained in mAP and rank-1 accuracy, respectively. For CUHK03, the improvements are +5.12% (from 68.36% to 73.48%) and +5.58% (from 63.10% to 68.68%) in mAP and rank-1 accuracy, respectively. We also test the effectiveness of our proposed method on two small-scale datasets, including VIPeR and CUHK01. +5.87% (mAP) and +5.84% (rank-1) improvements can be observed on VIPeR by adding 1,200 generated images in training. Meanwhile, +2.77% (mAP) and +3.48% (rank-1) improvements can be observed on CUHK01 by adding 4,000 generated images in training. The above results indicate the proposed MpRL can effectively yield improvements over the baseline performance on both large and small-scale re-ID datasets.

E. Comparison With Different Implementations of MpRL

Three implementations are used in our experiments to demonstrate the effectiveness of the proposed MpRL (see Section III-B). We conduct this experiment using the Identif network. Table IV gives the comparisons on Market-1501. We observe that by dynamically updating the likelihood of the affiliation of the generated data to all pre-defined training classes in training, dMpRL-I (+4.74% and +4.87% improvements in mAP and rank-1 accuracy respectively) and dMpRL-II (+5.91% and +6.29% improvements in mAP and rank-1 accuracy respectively) achieve better improvements

compared with the sMpRL (+3.08% and +4.77% improvements in mAP and rank-1 accuracy respectively). This is because each generated data will receive a proper MpRL along with the discriminative power of the CNN getting better in training. Also, compared with dMpRL-I, dMpRL-II achieves the best improvement when the network becomes relatively stable after 20 training epochs.

F. Comparison With Existing Virtual Labels

To further evaluate the proposed MpRL, we compare it with other three competitive virtual labels: all-in-one, one-hot pseudo, and LSRO. Amongst them, LSRO [12] is the state-of-the-art method using generated data for person re-ID. Table IV provides the comparison results. We add a different number of generated data in training to show the improvement. By adding 30,000 and 18,000 generated images, the all-in-one achieves the best improvements in mAP (+3.51%) and rank-1 vaccuracy (+3.32%), respectively. The one-hot pseudo achieves +4.22% (mAP) and +3.87% (rank-1) improvements when 24,000 and 30,000 generated images are respectively added. Compared with them, LSRO obtains a better rank-1 accuracy improvement (+4.13%) when adding 24,000 generated images. However, the improvement of mAP (+3.65%) is slightly less than the one-hot pseudo. In this experiment, we use the same generated data over all the methods; the improvements are on par with that reported in [12]. Although the improvement of mAP (+3.08%) is less than other virtual labels by using sMpRL, we obtain better rank-1 accuracy improvements under all the implementations of the proposed MpRL (+4.77%, +4.87%, and +6.29%, respectively). dMpRL-I and dMpRL-II also outperform other methods in mAP by +4.74% and +5.91% respectively. By adding 24,000 generated images, dMpRL-II improves the mAP and rank-1 accuracy of the Identif network from 52.68% and 74.08% to 58.59% and 80.37%, respectively. Our method outperforms the previous state-of-the-art method LSRO to a certain degree (mAP: +3.65%→ +5.91%, rank-1 accuracy: +4.13%→ +6.29%). It can be observed that when 12,000 generated images are used, there is limited regularization capability to improve the re-ID performance over all the virtual labels. Meanwhile, if too many generated images are added in training, *e.g.*, 48,000, the performance is dropped since the network tends to converge towards the generated data instead of real data. To balance the number of generated data in training, we empirically set it to 24,000 over the three large-scale datasets we used.

In Table IV, it is clear to see that the multiple label distribution (LSRO and MpRL) can always outperform the one-hot label distribution (all-in-one and one-hot pseudo) in the rank-1 accuracy. The reason can be found in Section IV-A. Besides, we also find that compared with the way using the same label, assigning different labels to generated data can achieve better results in both multiple (MpRL vs. LSRO) and one-hot (one-hot pseudo vs. all-in-one) label distribution. The reason can be found in Section IV-B.

To further investigate the performance of the proposed MpRL, we also evaluate it on two small-scale re-ID datasets. Table V lists the result on VIPeR. Our dMpRL-II

TABLE IV

COMPARISON OF ALL-IN-ONE, ONE-HOT PSEUDO, LSRO, AND MPRLS UNDER DIFFERENT NUMBERS OF GENERATED DATA ON MARKET-1501 BY USING THE IDENTIF NETWORK. THE BEST IMPROVEMENT OF DIFFERENT METHODS IS HIGHLIGHTED IN **BOLD**. RANK-1 ACCURACY AND mAP ARE SHOWN

#GAN Img	All-in-one [20], [14]		One-hot Pseudo [21]		LSRO [12]		sMpRL		dMpRL-I		dMpRL-II	
	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1
0 (base)	52.68%	74.08%	52.68%	74.08%	52.68%	74.08%	52.68%	74.08%	52.68%	74.08%	52.68%	74.08%
12000	55.68%	76.96%	55.69%	76.52%	55.22%	77.17%	55.27%	77.73%	55.84%	77.88%	58.14%	79.22%
18000	55.59%	**77.40%**	56.04%	77.95%	55.28%	76.96%	55.05%	77.73%	56.21%	78.36%	58.31%	79.81%
24000	56.07%	77.21%	**56.90%**	77.62%	**56.33%**	**78.21%**	55.59%	**78.85%**	56.10%	77.79%	**58.59%**	**80.37%**
30000	**56.19%**	77.17%	56.54%	**77.95%**	55.40%	77.46%	**55.76%**	77.82%	57.15%	78.65%	57.69%	79.16%
36000	55.24%	75.92%	56.38%	77.42%	55.82%	77.91%	55.45%	78.32%	**57.42%**	**78.95%**	57.61%	79.90%
48000	53.98%	75.16%	55.86%	76.72%	54.87%	76.90%	55.02%	77.45%	56.01%	77.57%	57.03%	78.73%
improvement	+3.51%	+3.32%	+4.22%	+3.87%	+3.65%	+4.13%	+3.08%	+4.77%	+4.74%	+4.87%	**+5.91%**	**+6.29%**

TABLE V

COMPARISON OF LSRO AND THE PROPOSED dMpRL-II UNDER DIFFERENT NUMBERS OF GENERATED DATA ON VIPeR WITH THE IDENTIF NETWORK. THE BEST IMPROVEMENT OF DIFFERENT METHODS IS HIGHLIGHTED IN **BOLD**. RANK-1 ACCURACY AND mAP ARE LISTED

#GAN Img	LSRO [12]		dMpRL-II	
	mAP	rank-1	mAP	rank-1
0 (base)	46.38%	40.76%	46.38%	40.76%
600	48.98%	42.80%	48.59%	42.61%
1200	**49.94%**	**43.57%**	**52.25%**	**46.60%**
1800	49.41%	43.39%	50.51%	44.24%
2400	45.95%	40.65%	49.36%	43.77%
12000	43.34%	37.12%	44.25%	37.66%
improvement	+3.56%	+2.81%	**+5.87%**	**+5.84%**

TABLE VI

COMPARISON BETWEEN USING GENERATED DATA BY DCGAN AND WGAN-GP. TWO APPROACHES ARE USED, INCLUDING LSRO AND THE PROPOSED dMpRL-II. EXPERIMENTS CONDUCTED ON THREE DATASETS: MARKET-1501, DukeMTMC-reID, AND VIPeR. RANK-1 ACCURACY AND mAP ARE LISTED

Method	Market-1501			
	DCGAN [2]		WGAN-GP [4]	
	mAP	rank-1	mAP	rank-1
LSRO [12]	56.33%	78.21%	55.53%	78.32%
dMpRL-II	58.59%	80.37%	59.04%	79.75%
	DukeMTMC-reID			
LSRO [12]	46.66%	66.92%	46.79%	66.97%
dMpRL-II	48.58%	68.24%	49.30%	68.76%
	VIPeR			
LSRO [12]	49.41%	43.39%	48.47%	43.14%
dMpRL-II	52.25%	46.60%	52.16%	46.39%

improves the mAP and rank-1 accuracy on this dataset by +5.87% and +5.84% respectively when adding 1,200 generated images in training, and outperforms the LSRO method. Since VIPeR is a small dataset (only 632 images for training), adding too many generated images, *e.g.*, 12,000 leads to inferior results. Therefore, we set the number of generated data to approximate double that of the number of real data on small datasets. Specifically, we use 1,200 and 4,000 generated images for VIPeR and CUHK01 respectively. We mainly report the result on VIPeR by changing the number of generated data. The results of CUHK01 can be found in Table III and VII.

Using the Identif network, Table III shows comparison results between our dMpRL-II and LSRO on three large-scale datasets by adding 24,000 generated images. Also, two small-scale datasets are used to evaluate the proposed method by adding 1,200 and 4,000 images respectively. By using different weights from pre-defined training classes, dMpRL-II can always outperform previous state-of-the-art virtual label LSRO over the five datasets. The reason can be found in Section IV-C.

G. Comparison With Different GAN Models

In addition to the DCGAN, other GAN models such as WGAN-GP [4] has demonstrated its superior in generating high quality person images. We attempt to generate data using the WGAN-GP. Then, the relationship between the quality of generated images and our proposed MpRL can be testified by using different GAN models. In this experiment, two large and one small-scale datasets are used individually to generate images. Figure 6 shows the generated data by using different

GAN models. It can be observed that the WGAN-GP exhibits better capability of generating person images on these datasets. In order to verify the impacts of image quality created by different GAN approaches, we compare the performance of the proposed MpRL on the two different generated data sets. Table VI lists the comparison results. It is observed that by using generated data with different quality through different GAN approaches, the re-ID performance is not significantly affected. This is because these generated data are employed to improve the performance of CNN models by its regularization power instead of providing more actual subjects beyond the scope of the raw dataset in training. Therefore, better generated data can bring superior perceptual quality but cannot dramatically boost the effectiveness of regularizer although some marginal improvements can be observed.

H. Comparison With the State-of-the-Art Methods

Although the main contribution in this paper focuses on using the generated data to improve the performance of CNNs, but not on producing a state-of-the-art result, we still compare our result with several state-of-the-art methods. Table VII lists the comparison results. It is clear to see that although the performance of the original Two-stream network is competitive, it still be inferior to many methods such as Resnet+OIM [57], SSM [36], JLML [60], SVDNet [61], and PDC [62]. However, by incorporating with the proposed dMpRL-II, the Two-stream network achieves the state of the art compared with other methods on Market-1501,
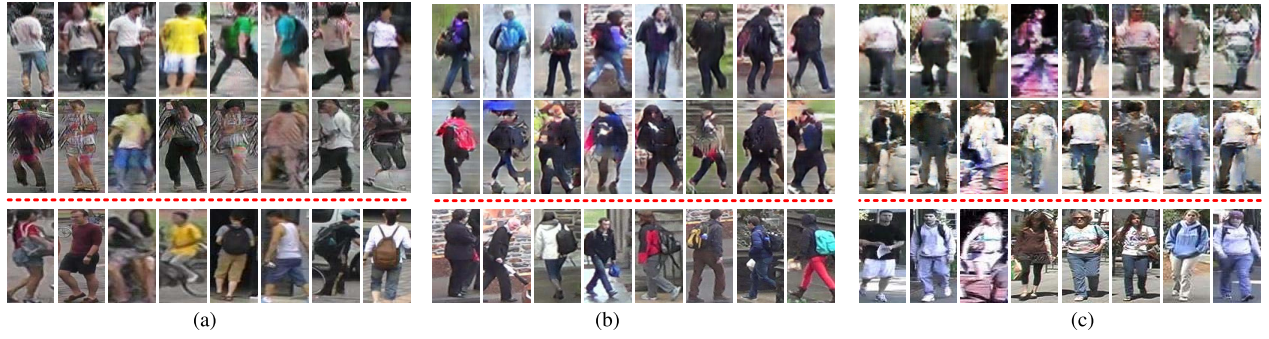
Fig. 6.   Examples of generated and real person images. (a)-(c) show the generated person images (first two rows) and real person images (the third row) on Market-1501, DukeMTMC-reID, and VIPeR respectively. Images in the first and second rows are respectively generated by the WGAN-GP [4] and the DCGAN [2].

TABLE VII

COMPARISON OF OUR RESULTS WITH THE PUBLISHED STATE-OF-THE-ART METHODS. THE BEST AND THE SECOND-BEST RESULTS ARE SHOWN IN **BOLD** AND UNDERLINE, RESPECTIVELY. RANK-1 ACCURACY AND MAP ARE LISTED. THE ReK MEANS RE-RANKING

Method		Large-Scale Datasets								Small-Scale Datasets	
		Market-1501				DukeMTMC-reID		CUHK03		VIPeR	CUHK01
		Single Query		Multiple Query		Single Query		Single Query (detected)		Single Query	Multiple Query
		mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	rank-1	rank-1
Gate-reID [52]	ECCV16	39.55%	65.88%	48.45%	76.04%	–	–	58.84%	68.10%	37.80%	–
SI-CI [53]	CVPR16	–	–	–	–	–	–	–	52.17%	35.76%	–
GOG+XQDA [54]	CVPR16	–	–	–	–	–	–	–	65.50%	49.70%	57.80%
SCSP [55]	CVPR16	26.35%	51.90%	–	–	–	–	–	–	53.54%	–
DNS [56]	CVPR16	35.68%	61.02%	46.03%	71.56%	–	–	–	54.70%	51.17%	69.09%
Resnet+OIM [57]	CVPR17	–	82.10%	–	–	–	68.10%	–	–	–	–
Latent Parts [30]	CVPR17	57.53%	80.31%	66.70%	86.79%	–	–	–	67.99%	38.08%	–
P2S [58]	CVPR17	44.27%	70.72%	55.73%	85.78%	–	–	–	–	–	–
ReRank [37]	CVPR17	63.63%	77.11%	–	–	–	–	–	–	–	–
CADL [7]	CVPR17	55.60%	80.90%	–	–	–	–	–	–	–	–
SpindleNet [59]	CVPR17	–	76.90%	–	–	–	–	–	–	53.80%	**79.90%**
SSM [36]	CVPR17	68.80%	82.21%	76.18%	88.18%	–	–	–	72.70%	53.73%	–
JLML [60]	IJCAI17	65.50%	85.10%	74.50%	89.70%	–	–	–	80.60%	50.20%	76.70%
SVDNet [61]	ICCV17	62.10%	82.30%	–	–	56.80%	76.70%	84.80%	81.80%	–	–
PDC [62]	ICCV17	63.41%	84.14%	–	–	–	–	–	78.29%	51.27%	–
Part Aligned [63]	ICCV17	63.40%	81.00%	–	–	–	–	–	81.60%	48.70%	75.00%
LSRO [12]	ICCV17	66.07%	83.97%	76.10%	88.42%	47.13%	67.68%	87.40%	84.60%	–	–
Identif [8], [9]		52.68%	74.08%	64.95%	82.06%	42.20%	61.94%	68.36%	63.10%	40.76%	65.33%
Identif+dMpRL-II		58.59%	80.37%	70.22%	86.47%	48.58%	68.24%	73.48%	68.68%	46.60%	68.81%
Two-stream [10]		64.09%	81.83%	73.65%	86.82%	51.40%	72.62%	85.20%	81.88%	51.84%	77.78%
Two-stream+dMpRL-II		67.53%	85.75%	77.85%	89.88%	58.56%	76.81%	87.53%	85.42%	**54.65%**	78.83%
Two-stream+dMpRL-II+ReK		**81.18%**	**87.96%**	**86.53%**	**90.97%**	**74.54%**	**81.28%**	**90.16%**	**88.00%**	53.22%	78.08%

DukeMTMC-reID, CUHK03 and VIPeR. To achieve better performance, after obtaining the rank list by sorting the similarity of gallery images to a query, a re-ranking method [37] is adopted to further boost our performance. The combination of the dMpRL-II and re-ranking on the Two-stream network achieves the best results on the three large-scale datasets. However, the re-ranking approach cannot further improve the performance of the two small-scale datasets with limited number of testing person identities. We find that the rank-1 accuracy of the DPFL method [64] proposed in the ICCV17 workshop is slightly higher than our result on Market-1501 (88.90% in single query and 92.30% in multiple query). However, DPFL uses an ensemble deep model with multiple granularity inputs for each image. Our Two-stream network just utilizes a single model and outperforms the DPFL on CUHK03 by a large margin in mAP even without re-ranking (mAP: 87.53% (our) vs. 78.10% (DPFL), rank-1: 85.42% (our) vs. 82.00% (DPFL)). Also, the performance of the Spindle [59] approach is slightly higher than ours on CUHK01 (79.90% *vs.* 78.83%). Since VIPeR and CUHK01 are two small-scale datasets, nine different person re-ID datasets are used to pre-train the SpindleNet model and then fine-tuning on the two small datasets respectively. We also use the fine-tuning strategy on these two datasets, but only three datasets are involved in the pre-training stage (see V-B2). Except for the CUHK01 dataset, our performance outperforms the SpindleNet on the other small-scale dataset VIPeR and the three large-scale re-ID datasets simultaneously.

VI. CONCLUSION

In this paper, we propose a new virtual label MpRL for the generated data by GAN. To train a CNN, MpRL is used as virtual label assigned to generated data. These data are used for semi-supervised learning. Two CNNs are adopted to show the effectiveness of the proposed MpRL. Experiments demonstrate that generated data can effectively improve the performance of the two CNNs trained with the proposed MpRL. Compared with the previous state-of-the-art method LSRO [12], MpRL can always achieve better improvements. In the future, considering the capability of GAN, we will continue to investigate virtual labels used on generated data for semi-supervised learning and apply the proposed method to other fields.

REFERENCES

[1] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.

[2] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–16.

[3] M. Arjovsky, S. Chintala, and L. Bottou. (2017). "Wasserstein gan." [Online]. Available: https://arxiv.org/abs/1701.07875

[4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 5767–5777.

[5] Y. Huang, H. Sheng, Y. Zheng, and Z. Xiong, "DeepDiff: Learning deep difference features on human body parts for person re-identification," *Neurocomputing*, vol. 241, pp. 191–203, Jun. 2017.

[6] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Jul. 2017, pp. 5399–5408.

[7] J. Lin, L. Ren, J. Lu, J. Feng, and J. Zhou, "Consistent-aware deep learning for person re-identification in a camera network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5771–5780.

[8] L. Zheng, Y. Yang, and A. G. Hauptmann. (2016). "Person re-identification: Past, present and future." [Online]. Available: https://arxiv.org/abs/1610.02984

[9] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person re-identification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1–2.

[10] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person reidentification," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 1, p. 13, 2016.

[11] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.

[12] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3754–3762.

[13] 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. (CVPR)*, Jul. 2014, pp. 152–159.

[14] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2234–2242.

[15] S. Hong, H. Noh, and B. Han, "Decoupled deep neural network for semi-supervised semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 1495–1503.

[16] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1635–1643.

[17] M. Bauml, M. Tapaswi, and R. Stiefelhagen, "Semi-supervised learning with constraints for person identification in multimedia data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3602–3609.

[18] Y. Luo, D. Tao, B. Geng, C. Xu, and S. J. Maybank, "Manifold regularized multitask learning for semi-supervised multilabel image classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 523–536, Feb. 2013.

[19] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 919–927.

[20] A. Odena. (2016). "Semi-supervised learning with generative adversarial networks." [Online]. Available: https://arxiv.org/abs/1606.01583

[21] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn. (ICML)*, vol. 3, 2013, pp. 1–2.

[22] S. Gong and T. Xiang, "Person re-identification," in *Visual Analysis of Behaviour*. London, U.K.: Springer, 2011, pp. 301–313.

[23] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2015, pp. 2197–2206.

[24] Y. Chen, W. Zheng, and J. Lai, "Mirror representation for modeling view-specific transform in person re-identification," in *Proc. Int. Joint. Conf. Artif. Intell. (IJCAI)*, 2015, pp. 3402–3408.

[25] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 649–656.

[26] S. Liao and S. Z. Li, "Efficient PSD constrained asymmetric metric learning for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3685–3693.

[27] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2013, pp. 3610–3617.

[28] H.-X. Yu, A. Wu, and W.-S. Zheng, "Cross-view asymmetric metric learning for unsupervised person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 994–1002.

[29] P. Zhang, Q. Wu, J. Xu, and J. Zhang, "Long-term person re-identification using true motion from videos," in *Proc. IEEE Winter Conf. Appl. Comput. Vis (WACV)*, Mar. 2018, pp. 494–502.

[30] D. Li, X. Chen, Z. Zhang, and K. Huang, "Learning deep context-aware features over body and latent parts for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 384–393.

[31] M. Geng, Y. Wang, T. Xiang, and Y. Tian. (2016). "Deep transfer learning for person re-identification." [Online]. Available: https://arxiv.org/abs/1611.05244

[32] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.

[33] 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. (CVPR)*, Jun. 2016, pp. 1335–1344.

[34] 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. (CVPR)*, Jul. 2017, pp. 1–8.

[35] Y. Huang, H. Sheng, and Z. Xiong, "Person re-identification based on hierarchical bipartite graph matching," in *Proc. IEEE Int. Conf. Image. Process. (ICIP)*, Sep. 2016, pp. 4255–4259.

[36] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1–7.

[37] 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. (CVPR)*, Jul. 2017, pp. 3652–3661.

[38] J. Garcia, N. Martinel, C. Micheloni, and A. Gardel, "Person re-identification ranking optimisation by discriminant context information analysis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1305–1313.

[39] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni, "Discriminant context information analysis for post-ranking person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1650–1665, Apr. 2017.

[40] S. Ali, O. Javed, N. Haering, and T. Kanade, "Interactive retrieval of targets for wide area surveillance," in *Proc. ACM Int. Conf. Multimedia. (ACMMM)*, 2010, pp. 895–898.

[41] H. Wang, S. Gong, X. Zhu, and T. Xiang, "Human-in-the-loop person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 405–422.

[42] C. Liu, C. Change Loy, S. Gong, and G. Wang, "Pop: Person re-identification post-rank optimisation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 441–448.

[43] 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. (CVPR)*, Jun. 2016, pp. 2818–2826.

[44] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 262–275.

[45] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2012, pp. 31–44.

[46] 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.

[47] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 17–35.

[48] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.

[49] D. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 770–778.

[51] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia. (ACMMM)*, 2015, pp. 689–692.

[52] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 791–808.

[53] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1288–1296.

[54] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1363–1372.

[55] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1268–1277.

[56] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1239–1248.

[57] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3376–3385.

[58] S. Zhou, J. Wang, J. Wang, Y. Gong, and N. Zheng, "Point to set similarity based deep feature learning for person reidentification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 6, Jul. 2017, pp. 1–10.

[59] H. Zhao *et al.*, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1077–1085.

[60] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proc. Int. Joint. Conf. Artifi. Intelli. (IJCAI)*, 2017, pp. 2194–2200.

[61] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3800–3808.

[62] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Pose-driven deep convolutional model for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3980–3989.

[63] L. Zhao, X. Li, J. Wang, and Y. Zhuang, "Deeply-learned part-aligned representations for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3239–3248.

[64] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 2590–2600.
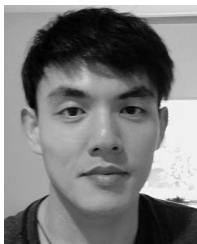
Yan Huang received the B.S. degree from the School of Computer Science and Engineering, Sichuan University, China, and the M.S. degree from Beihang University, China. He is currently pursuing the Ph.D. degree with the Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia. His research interests include deep learning and person re-identification.
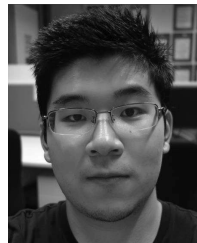

Jingsong Xu received the B.S. degree in computer science and the Ph.D. degree in pattern recognition from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2007 and 2014, respectively. He is currently a Research Fellow with the Global Big Data Technologies Center, University of Technology Sydney, Ultimo, NSW, Australia. His research interests include computer vision, pattern recognition, and machine learning.

Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2004. He is currently an Associate Professor and a Core Member of the Global Big Data Technologies Centre, University of Technology Sydney.

His research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His application fields where the research outcomes are applied span over video security surveillance, biometrics, video data analysis, and human–computer interaction. His research outcomes have been published in many premier international conferences, including ECCV, CVPR, ICIP, and ICPR, and the major international journals, such as IEEE TIP, IEEE TSMC-B, IEEE TCSVT, IEEE TIFS, PR, PRL, and *Signal Processing*.

Zhedong Zheng received the B.S. degree from Fudan University, China, in 2016. He is currently pursuing the Ph.D. degree with the University of Technology Sydney, Australia. His research interests include image retrieval and person re-identification.

Zhaoxiang Zhang received the B.S. degree in electronic science and technology from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2009.

In 2009, he joined the School of Computer Science and Engineering, Beihang University, Beijing, as an Assistant Professor, from 2009 to 2011, as an Associate Professor, from 2012 to 2015, and as the Vice Director of the Department of Computer Application Technology from 2014 to 2015. In 2015, he returned to the Institute of Automation, Chinese Academy of Sciences, as a Full Professor. His current research interests include computer vision, pattern recognition, machine learning, and brain-inspired neural network and brain-inspired learning.

Dr. Zhang is an Associate Editor or a Guest Editor of some internal journals, such as *Neurocomputing*, *Pattern Recognition Letters*, and the IEEE ACCESS.

Jian Zhang received the B.S. degree in electronics from East China Normal University, China, the M.S. degree in computer science from Flinders University, Australia, and the Ph.D. degree in electrical engineering from the University of New South Wales (UNSW), Australia. From 2004 to 2011, he was a Principal Researcher and a Project Leader with Data61, Australia, and a Conjoint Associate Professor with the School of Computer Science and Engineering, UNSW. He is currently an Associate Professor with the Global Big Data Technologies Centre, University of Technology Sydney, Australia. He has authored or co-authored over 140 paper publications, book chapters, and six issued U.S. and Chinese patents. His current interests include social multimedia signal processing, large-scale image and video content analytics, retrieval and mining, 3D-based computer vision, and intelligent video surveillance systems.

Dr. Zhang was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2006 to 2015. He has been an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and the *EURASIP Journal on Image and Video Processing* since 2016.