Fuzzy Multilayer Clustering and Fuzzy Label Regularization for Unsupervised Person Re-identification

Zhong Zhang, *Member, IEEE,* Meiyan Huang, Shuang Liu, *Member, IEEE,* Baihua Xiao and Tariq S. Durrani, *Fellow, IEEE*

***Abstract*—Unsupervised person re-identification has received more attention due to its wide real-world applications. In this paper, we propose a novel method named fuzzy multilayer clustering (FMC) for unsupervised person re-identification. The proposed FMC learns a new feature space using a multilayer perceptron for clustering in order to overcome the influence of complex pedestrian images. Meanwhile, the proposed FMC generates fuzzy labels for unlabelled pedestrian images, which simultaneously considers the membership degree and the similarity between the sample and each cluster. We further propose the fuzzy label regularization (FLR) to train the convolutional neural network (CNN) using pedestrian images with fuzzy labels in a supervised manner. The proposed FLR could regularize the CNN training process and reduce the risk of over-fitting. The effectiveness of our method is validated on three large-scale person re-identification databases, i.e., Market-1501, DukeMTMC-reID and CUHK03.**

***Index Terms*—Fuzzy multilayer clustering, fuzzy label regularization, unsupervised person re-identification.**

I. Introduction

PERSON re-identification, as a special kind of image retrieval, has received much attention in recent years due to its wide applications such as human behaviour analysis, multi-pedestrian association and cross camera tracking [1]–[3]. The aim of person re-identification is to search pedestrians who have the same identity with the probe in a large gallery. The person re-identification is an extremely challenging task, because pedestrian images are easily affected by illumination changes, various body postures and different camera angles.

With the development of deep learning, the convolutional neural network (CNN) has displayed potential for improving person re-identification performance [4]–[6]. The CNN is employed to learn deep features in an end-to-end way with various loss functions. The CNN-based methods for person re-identification are mainly divided into two groups according to loss functions. As for the first group, the CNN takes image pairs [7], triplets [8] or quadruplets [9] as input and determines whether two pedestrian images belong to the same identity or not by commonly utilizing contrastive loss or triplet loss [10]. This series of methods realizes the binary classification task for person re-identification and only requires implicit identity labels. As for the second group, person re-identification is considered as the multi-class classification task where the cross-entropy loss is often deployed to predict the identity probability [11], [12]. This group needs explicit identity labels for learning deep features. The similarity is finally calculated to retrieve pedestrian images from a large gallery for the pedestrian of interest. The two groups of methods have been proved to achieve promising performance on large-scale databases [13]–[15]. Hence, we adopt the CNN model for learning deep features.

Despite the prominent performance of CNN, it requires significant amounts of labelled samples in the training stage. However, the annotation process is labour intensive and tedious. Hence, some previous works [16]–[20] adopt unsupervised or semi-supervised strategies to alleviate the usage of labelled samples, but only effective on relatively small databases [21], [22]. At present, some CNN-based methods are developed for unsupervised person re-identification on large-scale databases. These methods are roughly divided into two categories. On one hand, some methods [23], [24] treat person re-identification as unsupervised domain adaptation which transfers the information from labelled source domain to

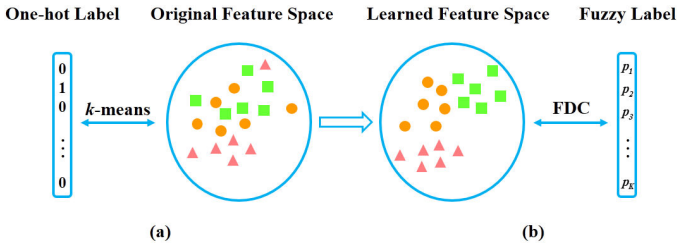

Fig. 1. (a) k-means clustering assigns each sample to the one-hot label in the original feature space. (b) The proposed FMC produces the fuzzy label for each sample in the learned feature space.

Z. Zhang, M. Huang and S. Liu are with Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China (e-mail: zhong.zhang8848@gmail.com; meiyanhuang7295@gmail.com; shuangliu.tjnu@gmail.com).

B. Xiao is with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: baihua.xiao@ia.ac.cn).

T. S. Durrani is with Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow Scotland, UK. (e-mail: t.durrani@strath.ac.uk).

This work was supported by National Natural Science Foundation of China under Grant No. 61501327 and No. 61711530240, Natural Science Foundation of Tianjin under Grant No. 17JCZDJC30600, the Fund of Tianjin Normal University under Grant No.135202RC1703, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 201700001 and No. 201800002, the China Scholarship Council No. 201708120039 and No. 201708120040, the Tianjin Higher Education Creative Team Funds Program and the NSFC-Royal Society grant.

unlabelled target domain. On the other hand, some researchers propose to assign pseudo labels to unlabelled samples for training CNN [25]–[27]. For example, Fan et al. [25] perform k-means clustering [28] on CNN-based features in an iterative way, and then assign labels according to clustering centers. However, k-means clustering is directly utilized in the original feature space for label prediction which is suboptimal for clustering as shown in Fig. 1 (a). In addition, k-means clustering only assigns each pedestrian image to one cluster (one-hot label), which easily results in over-fitting for unsupervised person re-identification.

In this paper, we propose a novel method named fuzzy multilayer clustering (FMC) for unsupervised person re-identification, which learns a suitable feature space for clustering using a multilayer perceptron and meanwhile produces fuzzy labels for unlabelled pedestrian images to reduce the risk of over-fitting as shown in Fig. 1 (b). Specifically, we first utilize irrelevant labelled pedestrian images to initialize the CNN model and then extract features from the CNN for unlabelled training pedestrian images. These pedestrian images are extremely complex due to changes in pedestrian appearances, illumination, captured views and so on. This results in a highly non-linear and inseparability feature space, and therefore it is hard to cluster them in the original feature space. To overcome this limitation, the proposed FMC first maps the features into a new feature space using a multilayer perceptron. As a result, the learned feature space is beneficial for clustering and label assignment. Afterwards, the proposed FMC assigns the fuzzy label to each pedestrian image in the learned feature space. The fuzzy label simultaneously considers the membership degree and the similarity between the sample and each cluster. Furthermore, we propose the fuzzy label regularization (FLR) to train the CNN using pedestrian images with fuzzy labels in a supervised manner, which could regularize the learning process of CNN and reduce the risk of over-fitting. Our method is an iterative process between clustering with FMC and the CNN training with fuzzy labels.

The contributions of this paper are summarized as follows:

1) The proposed FMC learns a new feature space for complex pedestrian images, and meanwhile assigns fuzzy labels to unlabelled samples.

2) The proposed FLR utilizes pedestrian images with fuzzy labels to train the CNN in a supervised manner, which could regularize the CNN learning process, while reducing the risk of over-fitting.

3) Experiments on three large-scale databases show that our method outperforms the state-of-the-art methods for unsupervised person re-identification.

II. Related Work

A. Fuzzy Clustering and Deep Clustering

Clustering collects samples with homogeneous features into the same cluster and distinguishes different clusters with heterogeneous features. k-means clustering [28] is one of the commonest clustering algorithms for unsupervised learning, and it assigns the one-hot label for each sample. Unlike k-means clustering, fuzzy c-means clustering [29], as one kind of fuzzy clustering, assigns soft labels for samples. Each element in the soft label represents the membership degree belonging to a certain cluster. There are many variants of fuzzy c-means clustering [30]–[34]. For example, Hathaway et al. [32] develop the generalized extensible fast fuzzy c-means algorithm (geFFCM), which focuses on the overall property of large database using statistics-based progressive sampling. In [33], the bit-reduced FCM (brFCM) adopts a binning strategy to accelerate fuzzy clustering. Havens et al. [34] propose the approximate kernel FCM (akFCM) which applies the sampled rows of the kernel matrix to constrain cluster centers to be linear combinations for fuzzy clustering.

Recently, deep clustering methods emerge for unsupervised learning. Tian et al. [35] utilize the stacked autoencoder to learn non-linear features for original graphs, and then perform k-means algorithm on them. Furthermore, Peng et al. [36] employ the autoencoder with sparsity prior to transform data into a non-linear latent space to adapt the local and global subspace structure for clustering. Chen et al. [37] develop a deep belief network (DBN) with nonparametric maximum margin clustering (NMMC) to obtain a discriminative clustering for deep features. In [38], Yang et al. learn latent representations from three deep networks and adopt k-means clustering on latent representations. Xie et al. [39] propose deep embedded clustering (DEC) to transform unlabelled data into a feature space with a deep network. Based on DEC, Guo et al. [40] take data structure into consideration by combining the clustering loss and deep network loss. Bhatia et al. [41] present a fuzzy graph clustering model by borrowing the idea from deep learning pipelines. However, there is no clustering algorithm to develop for unsupervised person re-identification which is expected to generate appropriate labels for unlabelled pedestrian images.

B. Unsupervised Person Re-Identification

Recently, unsupervised learning for person re-identification attracts more and more attention as it is closer to the real scenario. Hand-craft features can be directly deployed for unsupervised person re-identification because they do not need learning processing. Typical hand-crafted features including LOMO [42], SDALF [43] and ELF [44] are targeted at designing robust features for person re-identification. These features ignore the sample distribution of database. Some other methods for unsupervised person re-identification focus on salience learning. Zhao et al. [18] propose an unsupervised framework where salience detection in patch level is used to extract distinctive features without identity labels. Wang et al. [45] employ the generative probabilistic topic model to simultaneously discover salient image patches of person foreground appearance and remove background clutters surrounding the pedestrian. However, these methods are less effective for large-scale databases.

Inspired by the notable success of deep learning, some current works utilize deep features to represent unlabelled pedestrian images. They are mainly divided into two groups. The first group learns deep features from source and target databases using unsupervised domain adaptation technique,

and the second group predicts pseudo labels for unlabelled samples. For the first group, Wang et al. [46] propose a transferable model to learn the attribute-semantic and identity-discriminative feature spaces for the target domain. In [23], the similarity preserving generative adversarial network (SPGAN) is proposed to translate persons from the labelled source domain to the unlabelled target domain. Zhong et al. [24] simultaneously take camera invariance and domain connectedness into consideration for learning more generalized deep features on the target domain.

For the second group, Liu et al. [27] utilize the reciprocal nearest neighbor to learn cross-camera tracklet association and deep features for unsupervised video person re-identification. Wu et al. [26] present a dynamic sampling strategy to gradually estimate labels for relabelling samples and meanwhile update the deep model for one-shot person re-identification. Fan et al. [25] alternatively optimize the k-means clustering for label assignment and the CNN model on the unseen domain. Considering that k-means clustering in the original feature space can not perform well for unsupervised person re-identification, we aim to transfer the original feature space to a new one and generate fuzzy labels for unlabelled pedestrian images to reduce the risk of over-fitting.

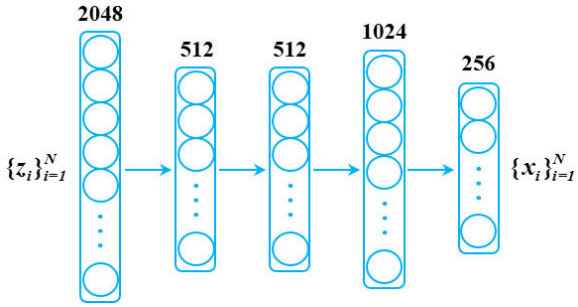

Fig. 2. The structure of the multilayer perceptron in FMC.

III. Fuzzy Multilayer Clustering

Many existing methods for person re-identification [24]–[26] perform clustering on the original feature space. However, features in the original feature space is of highly non-linear and inseparability due to complex pedestrian images. It often results in suboptimal clustering and inaccurate label assignment. To overcome the drawback, the proposed FMC utilizes a multilayer perceptron to learn a new feature space where the features are easily separated and are assigned appropriate labels. Afterwards, the proposed FMC assigns fuzzy labels for unlabelled samples to consider the membership degree and the similarity between sample and each cluster simultaneously. The proposed FMC iteratively optimizes the feature space and the fuzzy assignment using the following three steps. They are (1) to learn a mapping that transforms features into a new feature space; (2) to obtain the fuzzy assignment using learned features and cluster centers; (3) to compute the clustering loss for optimizing the feature mapping and cluster centers simultaneously.

A. Learning a New Feature Space

The proposed FMC first learns a new feature space using a multilayer perceptron. The structure of the multilayer perceptron in FMC is shown in Fig. 2, and it contains five fully-connected layers with 2048, 512, 512, 1024, and 256 neurons respectively. Let denote $\{z_i\}_{i=1}^N$ as the features of unlabelled training pedestrian images in the original feature space, where N is the number of unlabelled training pedestrian images. The dimension of z_i is 2,048. Note that we obtain z_i from a CNN model and we will introduce it in Section IV-A. We learn the multilayer perceptron in order to map z_i into a new feature space:

$$x_i = \phi_\theta(z_i) \tag{1}$$

where x_i is the feature of the i-th pedestrian image in the new feature space, and θ is the parameters of the multilayer perceptron. The learned feature x_i is the output of the multilayer perceptron, and therefore the dimension is 256.

B. Fuzzy Assignment

We employ the fuzzy c-means clustering [29] to cluster the learned features $\{x_i\}_{i=1}^N$. The fuzzy c-means clustering assigns x_i to each cluster with a certain membership degree by minimizing the following objective function:

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{j=1}^{K} a_{ij}^m ||x_i - c_j||^2 \\ &s.t. \sum_{j=1}^{K} a_{ij} = 1 \end{aligned} \tag{2}$$

where K is the pre-defined number of clusters, $m > 0$ is the fuzzy coefficient, c_j is the feature vector of the j-th cluster center, and a_{ij} represents the membership degree of the i-th pedestrian image belonging to the j-th cluster. The constraint indicates the sum of membership degree for each sample to all clusters is equal to 1. Here, we empirically set m to 2 in all experiments. a_{ij} and c_j are iteratively updated to optimize the objective function:

$$a_{ij} = \left[\sum_{k=1}^{K}\left(\frac{||x_i - c_j||}{||x_i - c_k||}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{3}$$

$$c_j = \frac{\sum_{i=1}^{N} a_{ij}^m x_i}{\sum_{i=1}^{N} a_{ij}^m} \tag{4}$$

In order to consider the membership degree and the similarity between pedestrian images and clusters, we propose the fuzzy assignment by modifying the t-distribution [59]. The fuzzy assignment of x_i to c_j is defined as:

$$f_{ij} = \frac{a_{ij}(1 + ||x_i - c_j||^2/\beta)^{-\frac{\beta+1}{2}}}{\sum_{k=1}^{K} a_{ik}(1 + ||x_i - c_k||^2/\beta)^{-\frac{\beta+1}{2}}} \tag{5}$$

where β is the degree of freedom. Here, we empirically set $\beta = 1$ in all experiments. The fuzzy assignment considers both the membership degree and the similarity of the pedestrian image to each cluster.

From Eq. (5), we derive that $f_{ij} \in [0,1]$ and $\sum_{j=1}^{K} f_{ij} = 1$. Hence, we regard the fuzzy assignment as the label distribution for unlabelled pedestrian images and define f_{ij} as the fuzzy label of the i-th pedestrian image belonging to the j-th identity. We utilize the fuzzy label instead of the one-hot label to generate pseudo labels for the subsequent supervised learning in Section IV-A.

C. Clustering Loss

We utilize a target distribution [39] to iteratively refine cluster centers according to high confidence assignments. The target distribution is defined as:

$$t_{ij} = \frac{f_{ij}^2/u_j}{\sum_{k=1}^{K} f_{ik}^2/u_k} \tag{6}$$

where $u_j = \sum_i f_{ij}$ is the j-th cluster frequency. In order to jointly learn the multilayer perceptron and cluster centers, we match the fuzzy assignment f_{ij} to the target distribution t_{ij}. We utilize the KL divergence as the loss function:

$$L = KL(T||F) = \sum_{i=1}^{N}\sum_{j=1}^{K} t_{ij} log \frac{t_{ij}}{f_{ij}} \tag{7}$$

where T and F are discrete probability distributions of t_{ij} and f_{ij}, respectively.

The proposed FMC learns the feature mapping and the cluster centers in a unified framework, which improves the accuracy of label assignment in the learned feature space. FMC jointly optimizes the membership degree, cluster centers and parameters of the multilayer perceptron as follows.

Step 1: We initialize the multilayer perceptron in FMC with the stacked autoencoder (SAE) and utilize the initialized multilayer perceptron to extract features $\{x_i\}_{i=1}^{N}$ in the new feature space. The fuzzy c-means clustering is utilized to cluster $\{x_i\}_{i=1}^{N}$ in order to initialize the membership degree a_{ij} and cluster centers $\{c_j\}_{j=1}^{K}$. Afterwards, we compute f_{ij} in Eq. (5) using initialized a_{ij} and $\{c_j\}_{j=1}^{K}$.

Step 2: After obtaining f_{ij}, we compute the loss in Eq. (7), and utilize the mini-batch stochastic gradient descent (SGD) to update parameters of the multilayer perceptron in FMC. With the updated multilayer perceptron, we traverse all the unlabeled pedestrian images and obtain the features $\{x_i\}_{i=1}^{N}$.

Step 3: We employ the fuzzy c-means clustering to cluster the features $\{x_i\}_{i=1}^{N}$ extracted from the updated multilayer perceptron to update the membership degree a_{ij} and cluster centers $\{c_j\}_{j=1}^{K}$ using Eq. (3) and Eq. (4). When $||A_M - A_{M-1}|| < \eta$, the iteration stops, where $A = [a_{ij}] \in \mathbb{R}^{N\times K}$, η is a small positive number, and M is the number of iterations. The updated a_{ij} and $\{c_j\}_{j=1}^{K}$ are used to compute f_{ij} using Eq. (5).

Step 4: We repeat Step 2 and Step 3 to optimize the membership degree, cluster centers and parameters of the multilayer perceptron in a unified framework. The iterations stop until the index change of all samples is less than a threshold between two consecutive updates, i.e., $\sum_i (l(x_i)_{old} \neq l(x_i))/N < \sigma$, which can ensure the algorithm convergence. Here, $l(x_i) = \arg\max_j f_{ij}$ indicates the maximum index of the fuzzy assignment for the i-th sample, N is the number of samples and σ is the threshold.

In practice, the termination condition of FMC can ensure the algorithm convergence. When satisfying the termination condition, the fuzzy assignment and parameters of the multilayer perceptron almost remain unchanged and the FMC converges.

TABLE I
THE STRUCTURE OF RESNET-50.

Name	Filters	Padding	Output Size
Conv_1	$\begin{bmatrix} 7\times 7, & 64 \end{bmatrix} \times 1$, stride 2	(3,3)	$128 \times 64 \times 64$
Max Pooling	3×3, stride 2	(1,1)	$64 \times 32 \times 64$
Conv_2	$\begin{bmatrix} 1\times 1, & 64 \\ 3\times 3, & 64 \\ 1\times 1, & 256 \end{bmatrix} \times 3$, stride 1	$\begin{bmatrix} (0,0) \\ (1,1) \\ (0,0) \end{bmatrix}$	$64 \times 32 \times 256$
Conv_3	$\begin{bmatrix} 1\times 1, & 128 \\ 3\times 3, & 128 \\ 1\times 1, & 512 \end{bmatrix} \times 4$, stride 2	$\begin{bmatrix} (0,0) \\ (1,1) \\ (0,0) \end{bmatrix}$	$32 \times 16 \times 512$
Conv_4	$\begin{bmatrix} 1\times 1, & 256 \\ 3\times 3, & 256 \\ 1\times 1, & 1024 \end{bmatrix} \times 6$, stride 2	$\begin{bmatrix} (0,0) \\ (1,1) \\ (0,0) \end{bmatrix}$	$16 \times 8 \times 1024$
Conv_5	$\begin{bmatrix} 1\times 1, & 512 \\ 3\times 3, & 512 \\ 1\times 1, & 2048 \end{bmatrix} \times 3$, stride 1	$\begin{bmatrix} (0,0) \\ (1,1) \\ (0,0) \end{bmatrix}$	$16 \times 8 \times 2048$
Average Pooling FC, $2,048$-dim Softmax			

IV. UNSUPERVISED PERSON RE-IDENTIFICATION USING FMC

A. Fuzzy Label Regularization

The flowchart for our method is shown in Fig. 3. We firstly initialize the CNN model using irrelevant labelled pedestrian images, where we treat the widely-used ResNet-50 [47] as CNN. The structure of ResNet-50 is illustrated in Table I. Conv_1 is a convolutional layer with the filter size of 7×7 followed by a max pooling with the size of 3×3. Conv_2-Conv_5 represent filter banks consisting of convolutional layers, and the number of filter banks are 3, 4, 6, 3, respectively. Suppose there are N unlabelled pedestrian images in the training set, and they are fed into ResNet-50. We extract features $\{z_i\}_{i=1}^{N}$ from the last fully-connected layer of ResNet-50 where the dimension of z_i is 2,048. Then, we utilize the proposed FMC to cluster $\{z_i\}_{i=1}^{N}$ and generate fuzzy labels for unlabelled pedestrian images. Afterwards, we select reliable samples with high confidence as the training samples. Finally, we propose FLR to train ResNet-50 using reliable samples with fuzzy labels in a supervised manner, which could regularize the learning process of ResNet-50 and reduce the risk of overfitting.

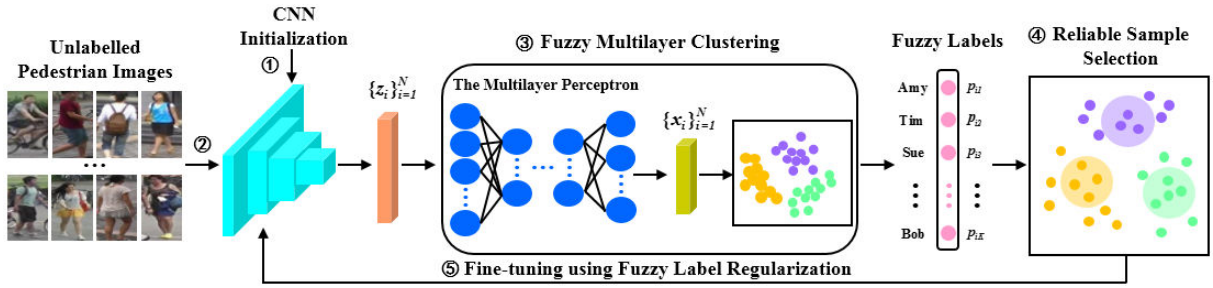

Fig. 3. The flowchart of our method for unsupervised person re-identification.

We employ the cross entropy as the loss function of ResNet-50:

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} p_{ij} log\ q_{ij} \tag{8}$$

where $q_{ij} \in [0,1]$ indicates the predicted probability of the i-th pedestrian image belonging to the j-th identity, and p_{ij} indicates the ground-truth identity of the i-th pedestrian image. Note that since the number of identities is unknown in unsupervised person re-identification, the proposed FMC assigns unlabelled pedestrian images into K clusters using fuzzy labels, where each cluster indicates one identity. Hence, K also represents the number of identities. q_{ij} is deduced from the Softmax function that is utilized to normalize the output of the fully-connected layer, and it can be directly outputted by ResNet-50. When we confirm the ground-truth identity of the i-th pedestrian image, p_{ij} can be written as:

$$p_{ij} = \begin{cases} 0, & j \neq s \\ 1, & j = s \end{cases} \tag{9}$$

where s is the ground-truth identity. From Eq. (9), we can see that the value of the cross-entropy loss only depends on the predicted probability of the ground-truth identity. However, the ground-truth label distribution is unknown for unsupervised person re-identification. Hence, we propose FLR to utilize fuzzy labels generated by FMC to optimize ResNet-50. Based on Eq. (5) and Eq. (8), the loss function in our method evolves to:

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} f_{ij} log\ q_{ij} \tag{10}$$

The proposed FLR pays attention to all clusters/identities by using fuzzy labels in cross-entropy loss, which could regularize the ResNet-50 training process and reduce the risk of over-fitting.

Since there are some noise or outliers in clustering results, ResNet-50 is likely to get trapped in a local optimization or oscillation. To alleviate this problem, we select reliable samples to train ResNet-50 rather than considering all training samples. Specifically, according to the fuzzy label, we first compute the index of the i-th pedestrian image:

$$l(x_i) = \arg\max_j f_{ij} \qquad j \in \{1, 2, \ldots, K\} \tag{11}$$

Then, we introduce a selection vector $r = [r_1, r_2, \ldots, r_N]$. If the i-th pedestrian image is chosen, r_i is equal to 1, otherwise, equal to 0. We formulate the selection criterion as:

$$\begin{aligned} &\min_r \sum_{j=1}^{K} \sum_{l(x_i)=j} r_i ||x_i - c_j|| - \xi ||r_i||_1 \\ &s.t.\ r_i \in \{0,1\}; \quad \sum_{l(x_i)=j} r_i \geq 1,\ \forall_j \end{aligned} \tag{12}$$

where ξ is a positive number. The constraint $\sum_{l(x_i)=j} r_i \geq 1$ is to guarantee that each cluster includes at least one reliable sample. From Eq. (12), the pedestrian images are selected as reliable samples when they are closer to the cluster centers. In other words, if the distance between the feature and its corresponding cluster center is less than ξ, this pedestrian image is selected as the reliable sample, otherwise discarded. With alternately updating between ResNet-50 and FMC, more reliable samples will be selected.

We can easily extend our method to semi-supervised learning for person re-identification. In order to integrate labelled pedestrian images into ResNet-50 training process, we directly add labelled pedestrian images into selected reliable samples in each iteration. The performance of semi-supervised learning for person re-identification will be analyzed in Section V-E.

B. Optimization Procedure

Our method is an iterative process between ResNet-50 training and FMC learning. The ResNet-50 is deployed to extract features $\{z_i\}_{i=1}^N$ for unlabelled pedestrian images. Then, the features $\{z_i\}_{i=1}^N$ are fed into the proposed FMC in order to generate fuzzy labels for the unlabelled pedestrian images. Note that we stop the FMC iteration when the label assignment change is less than a threshold σ between two consecutive updates. After selecting reliable samples, the proposed FLR utilizes them with fuzzy labels to train ResNet-50 in a supervised manner. The optimization procedure is presented in Alg. 1.

In practice, we first select the closest feature to the corresponding cluster center as the center feature, and then the cosine distance is employed to measure the similarity between features and center features in the new feature space. Hence, we are able to guarantee that each cluster contains at least one pedestrian image. The training process of ResNet-50 will converge until selected samples are saturated.

Algorithm 1: FMC for Unsupervised Person Re-identification

Input: Unlabelled pedestrian images;
Initialized ResNet-50;
Initialized multilayer perceptron with SAE;
Number of clusters: K;
Stopping threshold of FMC: σ;
Reliable sample selection threshold: ξ;
Initialized cluster centers $\{c_j\}_{j=1}^{K}$ and the membership degree a_{ij}.
Output: ResNet-50.
while *not convergence* **do**
  Extract features for unlabelled pedestrian images using ResNet-50: $\{z_i\}_{i=1}^{N}$;
  FMC
  while *not convergence* **do**
    (1) Compute the features $\{x_i\}_{i=1}^{N}$ using Eq. (1);
    (2) Update the multilayer perceptron in FMC using Eq. (7);
    (3) Cluster $\{x_i\}_{i=1}^{N}$ to update cluster centers $\{c_j\}_{j=1}^{K}$ and the membership degree a_{ij} using Eq. (3)-Eq. (4);
    (4) Obtain the index of the i-th pedestrian image $l(x_i)$ using Eq. (11)
    if $\sum_i (l(x_i)_{old} \neq l(x_i))/N < \sigma$ **then**
      stop
    end
  end
  Select reliable samples using Eq. (12);
  Train ResNet-50 using reliable samples with FLR.
end

TABLE II
THE STATISTICS FOR MARKET [48], DUKE [49] AND CUHK03 [4].

Databases	Cams	Training		Test			
		IDs	Images	Gallery		Query	
				IDs	Images	IDs	Images
Market	6	751	12,936	750	19,732	750	3,368
Duke	8	702	16,522	1,110	17,661	702	2,228
CUHK03	2	767	7,365	700	5,332	700	1,400

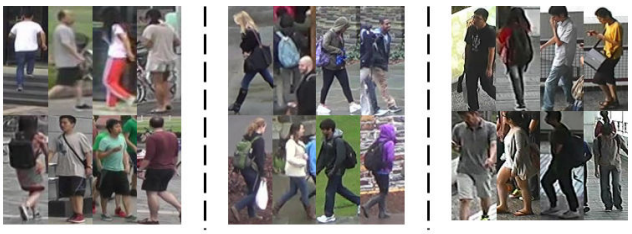

Fig. 4. Some pedestrian images from Market, Duke and CUHK03.

V. EXPERIMENTS

A. Databases and Settings

We evaluate our method on three large-scale person re-identification databases, i.e., Market-1501 [48], DukeMTMC-reID [49], and CUHK03 [4].

Market-1501 [48] includes 32,668 annotated pedestrian images of 1,501 identities captured from 6 different view points. All pedestrian images in Market-1501 are automatically cropped by the Deformable Part Model (DPM) [50], and are resized to 128×64. The database is partitioned into three parts: 751 identities with 12,936 pedestrian images in the training set, 750 identities with 19,732 pedestrian images in gallery, and 750 identities with another 3,368 pedestrian images in query. Note that the gallery and the query have the same identities but different images, and we retrieve the ground-truth images for each query from the gallery. For simplicity, we utilize "Market" to express the database.

DukeMTMC-reID [49] is a subset of DukeMTMC database [51], which contains 36,411 pedestrian images belonging to 1,812 identities which are collected by 8 disjoint cameras. This database is divided into three parts, i.e., training set, gallery set and query set. The training set includes 16,522 pedestrian images of 702 identities, and the gallery set contains 17,661 pedestrian images of the remaining 1,110 identities. The query set consists of 2,228 pedestrian images of the same 702 identities with the gallery. The pedestrian images in this database vary in size. For simplicity, we utilize "Duke" to express the database.

CUHK03 [4] contains 1,467 identities with 14,097 pedestrian images, and each identity is captured by two cameras. There are an average of 9.6 images for each identity. This database offers both manually labelled and DPM-detected bounding boxes. We evaluate on DPM-detected bounding boxes which are closer to the real scene. For fair comparison, we adopt the train/test protocol as [52]–[54]: 767 identities with 7,365 pedestrian images for training, 700 identities with 5,332 pedestrian images for gallery and the same 700 gallery identities with 1,400 pedestrian images for query. The pedestrian images in CUHK03 also vary in size.

The statistics and some pedestrian images are listed in Table II and Fig. 4. The query and gallery in Market and CUHK03 share the identities. In Duke, apart from the shared 702 identities, the gallery includes another 408 identities. We utilize the single query for all experiments and report rank-1, 5, 10, 20 accuracy and mean average precision (mAP) for three databases.

B. Implementation Details and Convergence analysis

According to the clustering loss, FMC is optimized by alternately updating the multilayer perceptron and cluster centers. We initialize the multilayer perceptron in FMC with the stacked autoencoder (SAE) which includes the greedy layer-wise pre-training and entire deep autoencoder training. Concretely, in the process of greedy layer-wise pre-training, the parameters are initialized by a zero-mean Gaussian distribution with a standard deviation of 0.01. Each fully-connected layer is pre-trained for 30,000 iterations with a dropout rate of 0.2. During the entire deep autoencoder training, we train the deep autoencoder for 60,000 iterations without dropout. For both layer-wise pre-training and entire deep autoencoder training, the batch size is set to 64, and the learning rate is

TABLE III
RANK-1, 5, 10, 20 ACCURACY (%) AND MAP (%) OF BASELINES. NOTE THAT (.) DENOTES THE IRRELEVANT DATABASE FOR RESNET-50 INITIALIZATION.

Methods	Market					Duke					CUHK03				
	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP
ResNet (Market)	-	-	-	-	-	32.5	49.1	55.0	61.9	16.7	11.2	16.6	20.1	25.7	10.0
k-means+CEL (Market)	-	-	-	-	-	39.4	53.2	59.3	65.3	20.4	12.9	18.8	22.1	29.4	12.3
Fuzzy+FLR (Market)	-	-	-	-	-	41.9	56.3	62.3	68.4	23.0	14.7	21.0	25.4	32.2	14.1
MD (Market)	-	-	-	-	-	45.9	59.5	65.6	71.0	25.8	16.5	23.3	27.5	34.6	16.4
FMC+CEL (Market)	-	-	-	-	-	46.6	60.1	66.0	71.5	26.2	17.1	23.7	27.9	35.1	16.8
Ours (Market)	-	-	-	-	-	**48.0**	**62.3**	**68.1**	**73.3**	**27.8**	**19.1**	**25.7**	**30.1**	**37.4**	**18.6**
ResNet (Duke)	44.3	62.3	69.1	76.8	20.3	-	-	-	-	-	9.6	14.7	18.8	25.0	8.8
k-means+CEL (Duke)	52.4	69.0	75.1	80.2	24.3	-	-	-	-	-	10.4	15.7	20.0	26.1	9.6
Fuzzy+FLR (Duke)	56.1	72.3	78.3	83.4	26.5	-	-	-	-	-	11.7	16.9	21.7	28.7	11.0
MD (Duke)	60.1	76.4	81.9	86.4	29.8	-	-	-	-	-	15.0	20.3	25.1	32.6	14.1
FMC+CEL (Duke)	60.8	77.2	82.8	87.3	30.4	-	-	-	-	-	15.4	20.8	25.5	32.9	14.5
Ours (Duke)	**63.4**	**79.5**	**84.8**	**89.2**	**32.4**	-	-	-	-	-	**16.7**	**22.0**	**26.7**	**34.5**	**15.9**
ResNet (CUHK03)	41.8	58.9	66.3	74.0	19.9	24.2	37.8	45.1	51.7	12.4	-	-	-	-	-
k-means+CEL (CUHK03)	46.7	64.2	70.8	76.5	21.8	31.9	45.8	53.0	58.9	15.1	-	-	-	-	-
Fuzzy+FLR (CUHK03)	50.2	67.9	74.5	80.1	24.1	35.7	49.7	56.8	62.4	17.4	-	-	-	-	-
MD (CUHK03)	55.0	72.8	79.4	84.7	28.0	40.4	55.3	61.8	67.0	21.4	-	-	-	-	-
FMC+CEL (CUHK03)	55.4	73.3	79.8	85.2	28.4	41.0	55.8	62.4	67.5	21.9	-	-	-	-	-
Ours (CUHK03)	**57.9**	**75.9**	**82.3**	**87.5**	**30.4**	**43.6**	**58.7**	**65.2**	**69.9**	**24.0**	-	-	-	-	-

initialized as 0.1 and decays by a factor of 0.1 every 20,000 iterations. We utilize the trained encoder to initialize the multilayer perceptron in FMC. In the learning phase of FMC, the stochastic gradient descent (SGD) approach is employed to update parameters, and the learning rate is fixed to 0.01. The stopping threshold σ is set to 0.001. The features are normalized by l_2 norm before reliable sample selection. The threshold of reliable sample selection ξ is set to 0.85 for three databases. There are 751, 702, 767 training identities in Market, Duke and CUHK03 respectively, and therefore we set the number of clusters K to 750 for Market and CUHK03, and 700 for Duke.

ResNet-50 is trained for 30 epoches with the proposed FLR. We modify the fully-connected layer for adapting to different databases. All pedestrian images are resized to 256×128, and the batch size is set to 32. For data augmentation, we adopt random horizontal flipping and cropping. In the training process, SGD is exploited with a momentum of 0.9, and the learning rate is set to 0.001 without any decay. In the test stage, we utilize the last fully-connected layer of ResNet-50 as the feature representation for each pedestrian image.

We implement our algorithm based on the PyTorch deep learning platform. The main hardware environment is two NVIDIA TITAN XP GPUs with 12Gbytes memory and an 4-core Intel Xeon(R) CPU E5-1620 v4. In the training stage, our algorithm requires about 9 hours on Market. In the test stage, the proposed algorithm requires 3min21s and therefore the processing time for each pedestrian image is only about 60ms.

Rather than considering all samples simultaneously, our algorithm is implemented iteratively with the reliable samples in a meaningful order that facilitates learning. The meaningful order is determined by the confidence embed into learning. Therefore, the learning process of our method can be regarded as the self-paced learning (SPL). The SPL [56]–[58] could prevent latent variable models from getting stuck in a bad point and accelerate the speed of convergence. Our method selects reliable samples to fine tune ResNet-50, and more samples with high confidence are selected as the training samples. We consider the fuzzy labels of the unlabelled samples as the latent variables, and therefore our method belongs to latent variable model. Hence, our method could converge quickly. In practice, our method converges after 20 iterations.

C. Comparison with Baselines

Table III represents the results of baselines and our method. ResNet denotes that ResNet-50 is initialized on one of databases and directly tested on another database. k-means+CEL adopts the k-means clustering in the original feature space and assigns the one-hot label to the unlabelled pedestrian image. The samples with one-hot labels are then utilized to train ResNet-50 using the cross-entropy loss. In Fuzzy+FLR, we perform the fuzzy c-means clustering to cluster features in the original feature space and utilize the proposed FLR to train ResNet-50. MD utilizes the membership degree a_{ij} instead of f_{ij} to train ResNet-50. In FMC+CEL, we use the proposed FMC to learn cluster features and obtain one-hot labels according to Eq. (11). Then, we train ResNet-50 with the cross-entropy loss. From Table III, five conclusions can be made.

Firstly, our method is far superior to ResNet. For example, when ResNet-50 is initialized on Duke and tested on Market, our method yields $+19.1\%$ and $+12.1\%$ improvement in rank-1 accuracy and mAP. Due to the data distribution bias between different databases, ResNet-50 only initialized with the irrelevant database is not perform well on the other database.

Secondly, compared with k-means+CEL, our method notably improves the accuracy of unsupervised person re-identification. For example, when tested on Market and initialized on Duke and CUHK03, we observe the improvements of $+11.0\%$ and $+11.2\%$ in rank-1 accuracy, respectively. Hence, compared with the one-hot label, the fuzzy label could reduce the over-fitting.

Thirdly, it is obviously indicated that our method significantly improves Fuzzy+FLR on all three databases. For instance, our method improves the rank-1 accuracy from 50.2% to 57.9% on Market and 35.7% to 43.6% on Duke when initialized on CUHK03. This is because our method learns a new feature space using a multilayer perceptron for clustering, rather than clustering features in the original feature space. The results indicate that learning a new feature space is effective

TABLE IV
THE PERFORMANCE OF OUR METHOD COMPARED WITH THE STATE-OF-THE-ART METHODS FOR UNSUPERVISED PERSON RE-IDENTIFICATION WHEN INITIALIZED ON MARKET/DUKE AND TESTED ON MARKET/DUKE, RESPECTIVELY.

Methods	Market → Duke				Duke → Market			
	rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
LOMO [42]	12.3	21.3	26.6	4.8	27.2	41.6	49.1	8.0
BoW [48]	17.1	28.8	34.9	8.3	35.8	52.4	60.3	14.8
UMDL [17]	18.5	31.4	37.6	7.3	34.5	52.6	59.6	12.4
PTGAN [55]	27.4	-	50.7	-	38.6	-	66.1	-
PUL [25]	30.4	44.5	50.7	16.4	44.7	59.1	65.5	20.1
SPGAN [23]	41.1	56.6	63.0	22.3	51.5	70.1	76.8	22.8
TJ-AIDL [46]	44.3	59.6	65.0	23.0	58.2	74.8	81.1	26.5
HHL [24]	46.9	61.0	66.7	27.2	62.2	78.8	84.0	31.4
Ours	**48.0**	**62.3**	**68.1**	**27.8**	**63.4**	**79.5**	**84.8**	**32.4**

TABLE V
THE PERFORMANCE OF OUR METHOD COMPARED WITH THE STATE-OF-THE-ART METHODS FOR UNSUPERVISED PERSON RE-IDENTIFICATION WHEN INITIALIZED ON CUHK03 AND TESTED ON MARKET/DUKE, RESPECTIVELY.

Methods	CUHK03 → Market					CUHK03 → Duke				
	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP
PTGAN [55]	31.5	-	60.2	-	-	17.6	-	38.5	-	-
PUL [25]	41.9	57.3	64.3	70.5	18.0	23.0	34.0	39.5	44.2	12.0
HHL [24]	56.8	74.7	81.4	86.3	29.8	42.7	57.5	64.2	69.1	23.4
Ours	**57.9**	**75.9**	**82.3**	**87.5**	**30.4**	**43.6**	**58.7**	**65.2**	**69.9**	**24.0**

to deal with complex pedestrian features and therefore the learned features are suitable for clustering.

Fourthly, MD drops from 63.4% to 60.1% in rank-1 accuracy and from 32.4% to 29.8% in mAP when initialized on Duke and tested on Market. When tested on Duke and CUHK03, MD is still inferior to our method. In a word, the fuzzy assignment f_{ij} is superior to the membership degree a_{ij} for unsupervised person re-identification.

Fifthly, compared with FMC+CEL, the performance of our method is consistently improved. When initialized on CUHK03, our method leads to an improvement of $+2.5\%$ in rank-1 accuracy and $+2.0\%$ in mAP on Market, and $+2.6\%$ in rank-1 accuracy and $+2.1\%$ in mAP on Duke. Similar conclusion is drawn when conducting tests on CUHK03. Different from FMC+CEL, our method proposes the FLR to train ResNet-50, which utilizes fuzzy labels to regularize the training process of ResNet-50 so as to reduce the risk of overfitting.

D. Comparison with Unsupervised Person Re-identification

We compare our method with the state-of-the-art methods for unsupervised person re-identification. The results of these methods tested on Market and Duke are detailed in Table IV. We first compare our method with two hand-crafted features, i.e., Local Maximal Occurrence (LOMO) [42] and Bag-of-Words (BoW) [48]. The two hand-crafted features are directly applied on the test set without any training process. From Table IV, we can see that they both achieve poor performance. We also compare our method with the existing unsupervised learning methods, including UMDL [17], PTGAN [55], SPGAN [23], TJ-AIDL [46] and HHL [24]. It is clear that our method outperforms the existing methods for unsupervised person re-identification, achieving a new state of the art. Specifically, when tested on Market, our method is higher than all competing methods, achieving 63.4% rank-1 accuracy and 32.4% mAP. When tested on Duke, our method yields the rank-1 accuracy of 48.0% and the mAP of 27.8%, surpassing other unsupervised learning methods as well. Table V presents the results when initialized on CUHK03, and our method achieves the best results once again.

In order to provide experimental explanation about different initializations, the experiments initialized on Duke and tested on Market are implemented 20 times. The mean and standard deviation of rank-1 accuracy are 0.6343 and 0.0028. The mean and standard deviation of mAP are 0.3239 and 0.0023. The rank-1 accuracy and mAP of the second best method are 0.622 and 0.314 as shown in Table IV. We employ the t-test to verify the statistical significance of our methods, and the results show that both rank-1 accuracy (p-value=4.75×10^{-14}) and mAP (p-value=9.98×10^{-14}) of our method are significantly better than other methods. Hence, our method can obtain very stable results and converge to reliable results when given various kinds of initializations.

E. Semi-supervised Person Re-identification

We briefly introduce the semi-supervised learning for person re-identification in Section IV-A. For the semi-supervised learning, the labelled pedestrian images of H identities are used for ResNet-50 training when selecting reliable samples. We set H to 25 and 50, respectively. The semi-supervised learning results are presented in Table VI, from which two major observations can be drawn.

Firstly, comparison to the unsupervised learning, the semi-supervised learning that adds 25 or 50 identities as supervised information for ResNet-50 training consistently improves the accuracy of person re-identification. When tested on Duke, the semi-supervised learning with 25 identities surpasses the unsupervised learning by $+2.8\%$ in rank-1 accuracy and $+2.4\%$ in mAP. Similarly, the semi-supervised learning obtains $+2.3\%$

TABLE VI
THE PERFORMANCE OF SEMI-SUPERVISED LEARNING FOR PERSON RE-IDENTIFICATION.

Semi-supervised	Market → Duke					Duke → Market				
	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP
Ours	48.0	62.3	68.1	73.3	27.8	63.4	79.5	84.8	89.2	32.4
Ours (25ID)	50.8	62.7	72.7	78.5	30.2	65.7	82.5	87.4	92.2	34.1
Ours (50ID)	54.6	72.6	77.5	83.9	32.8	68.9	85.4	90.7	94.8	36.6

and $+1.7\%$ improvement in rank-1 accuracy and mAP when tested on Market with 25 identities as supervised information.

Secondly, we can observe that utilizing more identities leads to better performance for semi-supervised learning. When tested on Duke, utilizing 50 identities is $+3.8\%$ and $+2.6\%$ higher than utilizing 25 identities in rank-1 accuracy and mAP. On the Market test set, the improvement brought by 50 identities is $+3.2\%$ in rank-1 accuracy and $+2.5\%$ in mAP higher than 25 identities. These results indicate that the performance of semi-supervised learning will increase with more labelled training pedestrian images but at the cost of expensive labelling process.

F. Important Parameters

There are three important parameters in our method, i.e., the feature dimension d of the output of the multilayer perceptron in FMC, the pre-defined number of clusters K and the reliable sample selection threshold ξ. We analyze results with 25 whole iterations.

Firstly, d is the feature dimension of the output of the multilayer perceptron in FMC. In order to choose the optimal d, we test on Market and initialize on Duke and CUHK03, respectively. The value of d is chosen from $\{64, 128, 256, 512, 1024\}$, and the results are shown in Fig. 5. From Fig. 5, $d = 256$ achieves the best results and the conclusions can be generalized to other databases.

Secondly, we evaluate the influence of K. Fig. 6 shows the rank-1 accuracy and mAP with different K when tested on one database and initialized on the other two databases, respectively. According to Fig. 6, the number of clusters K is determined by the one achieving the best performance and therefore we set the number of clusters K to 750 for Market and CUHK03, and 700 for Duke. Since we do not know the number of identities in practice, the number of clusters is always different from but approximates the number of identities. For instance, Market includes 751 training identities, as we expect, the rank-1 accuracy and mAP yield the best results when $K = 750$.

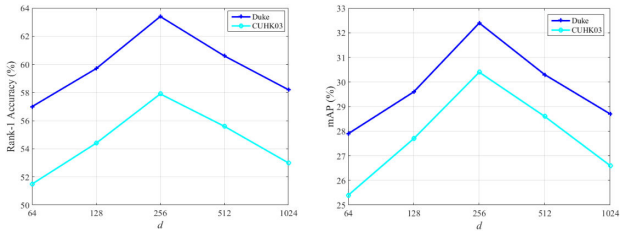

Fig. 5. Rank-1 accuracy and mAP vary with different d.

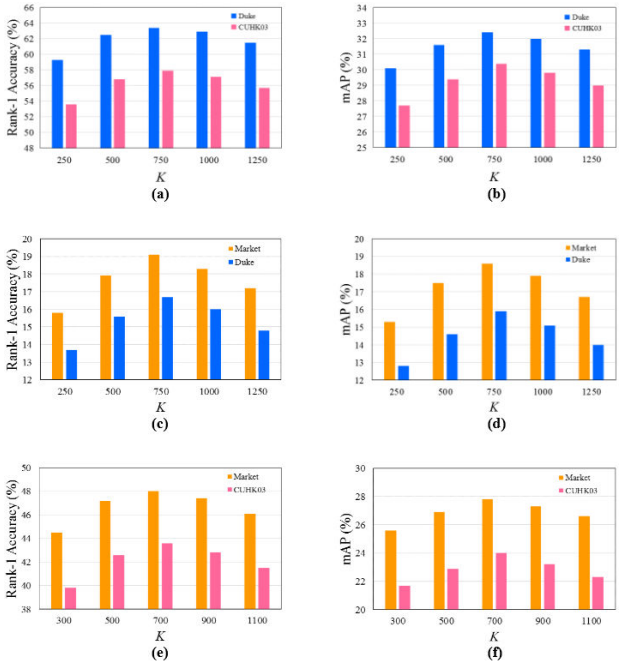

Fig. 6. Rank-1 accuracy and mAP with different K when tested on one database and initialized on the other two databases, respectively.

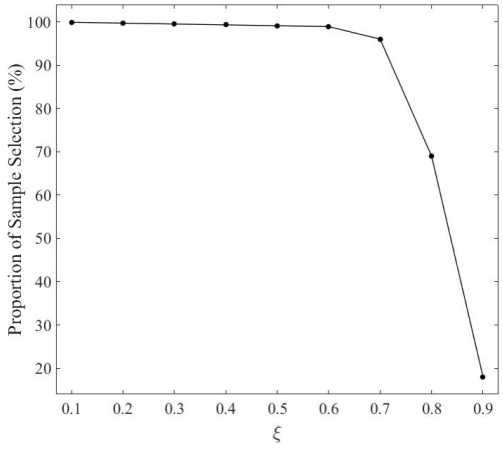

Fig. 7. The proportion of selected reliable samples in the first whole iteration with different ξ.

Thirdly, we evaluate the influence of ξ. ResNet-50 is initialized on Duke and tested on Market. The cosine distance is utilized to measure the similarity between two features, and therefore ξ belongs to $[-1, 1]$. Since ResNet-50 employs rectified linear unit (ReLU) with the value range from 0 to 1 as the activation function, the range of ξ is changed to $[0, 1]$. The smaller the value of ξ, the more reliable samples are selected. Fig. 7 shows the proportions of selected reliable samples with

TABLE VII
RANK-1 ACCURACY (%) AND MAP (%) ON THREE DATABASES WITHOUT RELIABLE SAMPLE SELECTION.

Accuracy	Market → Duke	Market → CUHK03	Duke → Market	Duke → CUHK03	CUHK03 → Market	CUHK03 → Duke
rank-1	42.5	15.2	57.2	12.7	51.3	37.5
mAP	23.6	14.5	27.1	12.2	24.9	18.9

TABLE VIII
RANK-1 ACCURACY (%) AND MAP (%) ON THREE DATABASES BY COMPUTING t_{ij} WITHOUT f_{ij} SQUARE.

Accuracy	Market → Duke	Market → CUHK03	Duke → Market	Duke → CUHK03	CUHK03 → Market	CUHK03 → Duke
rank-1	46.1	17.4	61.3	15.2	56.0	41.6
mAP	26.1	16.9	31.1	14.4	29.2	22.7

different ξ in the first whole iteration. We can observe that nearly all training pedestrian images are selected as reliable samples when $\xi < 0.7$. Therefore, we narrow the range of ξ and choose the value ξ from $\{0.70, 0.75, 0.80, 0.85, 0.90\}$. Fig. 8 and Fig. 9 represent the performance under different ξ. Generally, performance improves with increasing iterations, which verifies the effectiveness of our method for unsupervised person re-identification. Especially, $\xi = 0.85$ achieves the best performance.

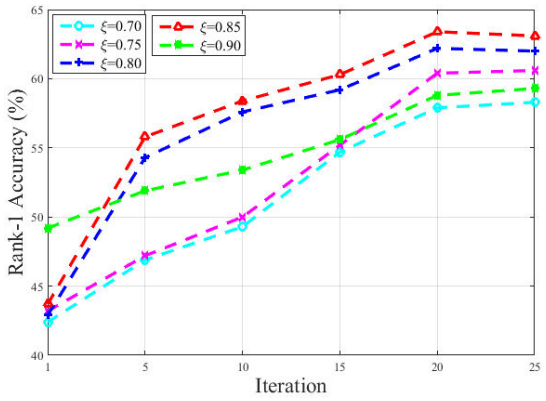

Fig. 8. Rank-1 accuracy with different ξ during 25 whole iterations (d=256, K=750).

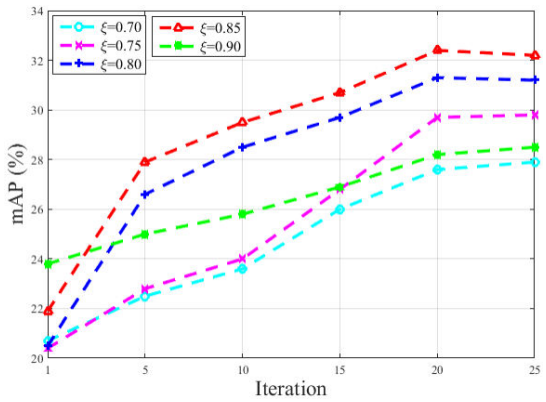

Fig. 9. mAP with different ξ during 25 whole iterations (d=256, K=750).

Furthermore, we discuss the situation when $\xi = 0$. That is, we train ResNet-50 without reliable sample selection. The results are presented in Table VII. Comparison with our method in Table III, the performance without reliable sample selection drops on three databases. For example, when initialized on Duke and tested on Market, our method yields the rank-1 accuracy of 63.4% and mAP of 32.4%, but our method drops to 57.2% in rank-1 accuracy and 27.1% in mAP without reliable sample selection. Hence, the reliable sample selection could exclude some noise in the clustering results.

In addition, we conduct experiments on three databases to validate the superiority of computing the target distribution t_{ij} in Eq. (6) with f_{ij} square to computing t_{ij} without f_{ij} square. The results of computing t_{ij} without f_{ij} are presented in Table VIII. Compared with our method in Table III, our method improves $1.5\% \sim 2.1\%$ in rank-1 accuracy and $1.2\% \sim 1.7\%$ in mAP on three databases than computing t_{ij} without f_{ij} square. This is because computing t_{ij} with f_{ij} square makes samples with higher confidence close to cluster centers, while makes samples with lower confidence far away from cluster centers. As a result, the target distribution pays more attention to samples with high confidence. Hence, t_{ij} in Eq. (6) needs f_{ij} square.

G. Discussion on the Structure of the Multilayer Perceptron

We conduct several experiments with different structures of multilayer perceptron to detail the determination of the number of fully-connected layers and the pattern of hidden layer sizes.

Firstly, we discuss the number of fully-connected layers. The number of neurons in the first and last layers remains 2048 and 256. In the experiments, the number of fully-connected layers is set to 3, 4, 5, 6, 7, respectively, and the corresponding neuron number of layers is set to {2048, 512, 256}, {2048, 512, 1024, 256}, {2048, 512, 512, 1024, 256}, {2048, 512, 512, 1024, 1024, 256} and {2048, 512, 512, 1024, 1024, 256, 256}. We present the experimental results tested on Market and initialized on Duke in Table IX. From Table IX, when the number of fully-connected layer is equal to 5, the results achieve best and the conclusions can be generalized to other databases. Note that in our experiments, we test many kinds of multilayer perceptron structures, and only report the results of several representative ones.

Secondly, we fix the number of fully-connected layers to 5 and analyze the pattern of hidden layer sizes except for the feature dimension of the final output. We list six kinds of multilayer perceptron structures which are {2048, 1024, 1024, 512, 256}, {2048, 1024, 512, 512, 256}, {2048, 1024, 512, 256, 256}, {2048, 512, 512, 1024, 256}, {2048, 512, 1024, 1024, 256} and {2048, 512, 1024, 256, 256}. Table X shows

TABLE IX
RANK-1 ACCURACY (%) AND MAP (%) TESTED ON MARKET AND INITIALIZED ON DUKE WITH DIFFERENT NUMBERS OF FULLY-CONNECTED LAYERS (FC).

FC	3	4	5	6	7
	{2048, 512, 256}	{2048, 512, 1024, 256}	{2048, 512, 512, 1024, 256}	{2048, 512, 512, 1024, 1024, 256}	{2048, 512, 512, 1024, 1024, 256, 256}
rank-1	54.3	57.6	**63.4**	60.2	58.9
mAP	27.6	29.3	**32.4**	30.7	29.8

TABLE X
RANK-1 ACCURACY (%) AND MAP (%) TESTED ON MARKET AND INITIALIZED ON DUKE WITH DIFFERENT PATTERNS OF HIDDEN LAYER SIZES.

Accuracy	{2048, 1024, 1024, 512, 256}	{2048, 1024, 512, 512, 256}	{2048, 1024, 512, 256, 256}	{2048, 512, 512, 1024, 256}	{2048, 512, 1024, 1024, 256}	{2048, 512, 1024, 256, 256}
rank-1	57.2	59.3	57.8	**63.4**	61.1	58.7
mAP	28.9	30.2	29.3	**32.4**	31.2	29.8

the rank-1 accuracy and mAP tested on Market and initialized on Duke with different patterns of hidden layer sizes. The structure of the multilayer perceptron with {2048, 512, 512, 1024, 256} yields the best results. Hence, we choose the multilayer perceptron containing five fully-connected layers with 2048, 512, 512, 1024 and 256 neurons respectively. Note that in our experiments, we test many kinds of multilayer perceptron structures with 5 layers, and only report the results of several representative ones.

VI. CONCLUSION

In this paper, we have proposed a novel method termed FMC for unsupervised person re-identification. The proposed FMC possesses two advantages. Firstly, to alleviate the influence of complex pedestrian images, FMC utilizes a multilayer perceptron to map features into a new feature space which is beneficial for clustering. Secondly, FMC assigns fuzzy labels for unlabelled pedestrian images instead of one-hot labels, which could simultaneously consider the membership degree and the similarity between the sample and each cluster. Afterwards, the proposed FLR utilizes pedestrian images with fuzzy labels to train ResNet-50, which regularizes ResNet-50 learning process and meanwhile reduces the risk of over-fitting. Experiments on three benchmark person re-identification databases demonstrate that our method achieves a new state of the art.

REFERENCES

[1] T. Banerjee, J.-M. Keller, M. Skubic, E. Stone, "Day or night activity recognition from video using fuzzy clustering techniques", *IEEE T. Fuzzy Syst.*, vol. 22, no. 3, pp. 483-493, Jun. 2014.

[2] Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, C. Shi, "Cross-view action recognition via a continuous virtual path", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2690-2697, 2013.

[3] F. Zheng, L. Shao, "Learning cross-view binary identities for fast person re-identification", *International Joint Conference on Artificial Intelligence*, pp. 2399-2406, 2016.

[4] W. Li, R. Zhao, T. Xiao, X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152-159, 2014.

[5] L. Wu, C. Shen, A. Hengel, "Personnet: Person re-identification with deep convolutional neural networks", *arXiv preprint arXiv:1601.07255*, 2016.

[6] T. Xiao, S. Li, B. Wang, L. Lin, X. Wang, "End-to-end deep learning for person search", *arXiv preprint arXiv:1604.01850*, 2016.

[7] R.-R. Varior, M. Haloi, G. Wang, "Gated siamese convolutional neural network architecture for human re-identification", *European Conference on Computer Vision*, pp. 791-808, 2016.

[8] H. Liu, J. Feng, M. Qi, J. Jiang, S. Yan, "End-to-end comparative attention networks for person re-identification", *IEEE T. Image Process.*, vol. 26, no. 7, pp. 3492-3506, 2017.

[9] W. Chen, X. Chen, J. Zhang, K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 403-412, 2017.

[10] A. Hermans, L. Beyer, B. Leibe, "In defense of the triplet loss for person re-identification", *arXiv preprint arXiv:1703.07737*, 2017.

[11] Z. Zheng, L. Zheng, Y. Yang, "A discriminatively learned CNN embedding for person re-identification", *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 1, pp. 13, Dec. 2017.

[12] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, "Beyond part models: Person retrieval with refined part pooling", *arXiv preprint arXiv:1711.09349*, 2017.

[13] Z. Zhang, M. Huang, "Learning local embedding deep features for person re-identification in camera networks", *Eurasip J. Wirel. Comm.*, vol. 2018, no. 1, pp. 85, 2018.

[14] T. Xiao, H. Li, W. Ouyang, X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1249-1258, 2016.

[15] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng,"Person re-identification by multi-channel parts-based CNN with improved triplet loss function", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1335-1344, 2016.

[16] Y. Yang, S. Li, L. Wen, S. Lyu, "Unsupervised learning of multi-level descriptors for person re-identification", *Association for the Advance of Artificial Intelligence*, pp. 4306-4312, 2017.

[17] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1306-1315, 2016.

[18] R. Zhao, W. Ouyang, X. Wang, "Unsupervised salience learning for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3586-3593, 2013.

[19] X. Ma, X. Zhu, S. Gong, X. Xie, J. Hu, K.-M. Lam, Y. Zhong, "Person re-identification by unsupervised video matching", *Pattern Recogn.*, vol. 65, pp. 197-210, May 2017.

[20] E. Kodirov, T. Xiang, Z. Fu, S. Gong, "Person re-identification by unsupervised l_1 graph learning", *European Conference on Computer Vision*, pp. 178-195, 2016.

[21] D. Gray, S. Brennan, H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking", *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, pp. 1-7, 2007.

[22] T. Wang, S. Gong, X. Zhu, S. Wang, "Person re-identification by video ranking", *European Conference on Computer Vision*, pp. 688-703, 2014.

[23] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 994-1003, 2018.

[24] Z. Zhong, L. Zheng, S. Li, Y. Yang, "Generalizing a person retrieval model hetero- and homogeneously", *European Conference on Computer Vision*, pp. 1-17, 2018.

[25] H. Fan, L. Zheng, Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning", *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018.

[26] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang, Y. Yang, "Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5177-5186, 2018.

[27] Z. Liu, D. Wang, H. Lu, "Stepwise metric promotion for unsupervised video person re-identification", *IEEE International Conference on Computer Vision*, pp. 2448-2457, 2017.
[28] J. MacQueen, "Some methods for classification and analysis of multivariate observations", *Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297, 1967.
[29] J.-C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters", *Journal of Cybernetics*, vol. 3, no. 3, pp. 32-57, Sep. 1973.
[30] R. Ding, X. Wang, K. Shang, B. Liu, F. Herrera, "Sparse representation-based intuitionistic fuzzy clustering approach to find the group intra-relations and group leaders for large-scale decision making.", *IEEE T. Fuzzy Syst.*, 2018.
[31] X. Bai, Y. Wang, H. Liu, S. Guo, "Symmetry information based fuzzy clustering for infrared pedestrian segmentation", *IEEE T. Fuzzy Syst.*, Vol. 26, no. 4, pp. 1946-1959, Sep. 2017.
[32] R. Hathaway, J. Bezdek, "Extending fuzzy and probabilistic clustering to very large data sets", *Comput. Statist. Data Anal.*, vol. 51, no. 1, pp. 215-234, Nov. 2006.
[33] S. Eschrich, J. Ke, L. Hall, D. Goldgof, "Fast accurate fuzzy clustering through data reduction", *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 262-269, Apr. 2003.
[34] T. Havens, R. Chitta, A. Jain, R. Jin, "Speedup of fuzzy and possibilistic kernel c-means for large-scale clustering", *IEEE International Conference on Fuzzy Systems*, pp. 463-470, 2011.
[35] F. Tian, B. Gao, Q. Cui, E. Chen, T. Liu, "Learning deep representations for graph clustering", *Association for the Advance of Artificial Intelligence*, pp. 1293-1299, 2014.
[36] X. Peng, S. Xiao, J. Feng, W. Yau, Z. Yi, "Deep subspace clustering with sparsity prior", *International Joint Conference on Artificial Intelligence*, pp. 1925-1931, 2016.
[37] G. Chen, "Deep learning with nonparametric clustering", *arXiv preprint arXiv:1501.03084*, 2015.
[38] L. Yang, X. Cao, D. He, C. Wang, X. Wang, W. Zhang, "Modularity based community detection with deep learning", *International Joint Conference on Artificial Intelligence*, pp. 2252-2258, 2016.
[39] J. Xie, R. Girshick, A. Farhadi, "Unsupervised deep embedding for clustering analysis", *International Conference on Machine Learning*, pp. 478-487, 2016.
[40] X. Guo, L. Gao, X. Liu, J. Yin, "Improved deep embedded clustering with local structure preservation", *International Joint Conference on Artificial Intelligence*, pp. 1753-1759, 2017.
[41] V. Bhatia, R. Rani, "DFuzzy: A deep learning-based fuzzy clustering model for large graphs", *Knowl. Inf. Syst.*, vol. 57, no. 1, pp. 159-181, Oct. 2018.
[42] S. Liao, Y. Hu, X. Zhu, S.-Z. Li, "Person re-identification by local maximal occurrence representation and metric learning", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2197-2206, 2015.
[43] L. Bazzani, M. Cristani, V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification", *Comput. Vis. Image Und.*, vol. 117, no. 2, pp. 130-144, Feb. 2013.
[44] D. Gray, H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features", *European Conference on Computer Vision*, pp. 262-275, 2008.
[45] H. Wang, S. Gong, T. Xiang, "Unsupervised learning of generative topic saliency for person re-identification", *British Machine Vision Association*, pp. 1-11, 2014.
[46] J. Wang, X. Zhu, S. Gong, W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification", *arXiv preprint arXiv:1803.09786*, 2018.
[47] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
[48] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, "Scalable person re-identification: A benchmark", *IEEE International Conference on Computer Vision*, pp. 1116-1124, 2015.
[49] Z. Zheng, L. Zheng, Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro", *IEEE International Conference on Computer Vision*, PP. 3774-3782, 2017.
[50] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, "Object detection with discriminatively trained part-based models", *IEEE T. Pattern Anal.*, vol. 32, no. 9, pp. 1627-1645, Sep. 2010.
[51] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking", *European Conference on Computer Vision*, pp. 17-35, 2016.
[52] Z. Zhong, L. Zheng, D. Cao, S. Li, "Re-ranking person re-identification with k-reciprocal encoding", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3652-3661, 2017.
[53] Y. Chen, X. Zhu, G. Shao, "Person re-identification by deep learning multi-scale representations", *IEEE International Conference on Computer Vision*, pp. 2590-2600, 2017.
[54] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, "Random erasing data augmentation", *arXiv preprint arXiv:1708.04896*, 2017.
[55] L. Wei, S. Zhang, W. Gao, Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 79-88, 2018.
[56] M. P. Kumar, B. Packer, D. Koller, "Self-paced learning for latent variable models", *Neural Information Processing Systems*, pp. 1189-1197, 2010.
[57] L. Jiang, D. Meng, S. Yu, Z. Lan, S. Shan, A. G. Hauptmann, "Self-paced learning with diversity", *Neural Information Processing Systems*, pp. 2078-2086, 2014.
[58] M. Fan, M. Deyu, X. Qi, Z. Li, X. Dong, "Self-paced co-training", *International Conference on Machine Learning*, pp. 2275-2284, 2017.
[59] L. V. D. Maaten, G. Hinton, "Visualizing data using t-SNE", *J. Mach. Learn. Res.*, vol. 9, pp. 2579-2605, Nov. 2008.

Zhong Zhang (M' 14) is an Associate Professor at Tianjin Normal University, Tianjin, China. He received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published about 90 papers in international journals and conferences such as the Pattern Recognition, IEEE Transactions on Circuits Systems Video Technology, IEEE Transactions on Information Forensics and Security, Signal Processing (Elsevier), CVPR, ICPR and ICIP. His research interests include computer vision, pattern recognition and deep learning. He is a member of IEEE.

Meiyan Huang is a master student at Tianjin Normal University, Tianjin, China. Her research interests include person re-identification and deep learning.

Shuang Liu (M' 18) is an Associate Professor at Tianjin Normal University, Tianjin, China. She received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her research interests include computer vision and deep learning. She is a member of IEEE.

Baihua Xiao received the B.S. degree in Electronic Engineering from Northwestern Polytechnical University, Xian, China and the Ph.D. degree in Pattern Recognition and Artificial Intelligence from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995 and 2000, respectively. From 2005, he has been a Professor at the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include pattern recognition, computer vision, image processing and machine learning.

Tariq Durrani is Research Professor at University of Strathclyde, Glasgow Scotland. His research covers AI, Signal Processing and Technology Management. He has authored 350 publications; supervised 45 PhDs.

He is a Fellow of the: IEEE, UK Royal Academy of Engineering, Royal Society of Edinburgh, IET, and the Third World Academy of Sciences. In 2018 he was elected Foreign Member of the US National Academy of Engineering.