Asymmetric filtering-based dense convolutional neural network for person re-identification combined with Joint Bayesian and re-ranking ☆

Shengke Wang [a,*], Xiaoyan Zhang [a], Long Chen [b], Huiyu Zhou [b], Junyu Dong [a]

[a] *Department of Computer Science and Technology, Ocean University of China, China*
[b] *Department of Informatics, University of Leicester, United Kingdom*

ARTICLE INFO

Keywords:
Person re-identification
Joint Bayesian
Deep convolutional neural networks
Multimodal features

ABSTRACT

Person re-identification aims at matching individuals across multiple camera views under surveillance systems. The major challenges lie in the lack of spatial and temporal cues, which makes it difficult to cope with large variations of lighting conditions, viewing angles, body poses and occlusions. How to extract multimodal features including facial features, physical features, behavioral features, color features, etc is still a fundamental problem in person re-identification. In this paper, we propose a novel Convolutional Neural Network, called Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN) to learn powerful features, which can extract different levels' features and take advantage of identity information. Moreover, instead of using typical metric learning methods, we obtain the ranking lists by merging Joint Bayesian and re-ranking techniques which do not need dimensionality reduction. Finally, extensive experiments show that our proposed architecture performs well on four popular benchmark datasets (CUHK01, CUHK03, Market-1501, DukeMTMC-reID).

1. Introduction

The person re-identification (ReID) involves addressing the problem of matching people across disjoint camera views in a multi-camera system, and is prevalent in both computer vision and multimedia analysis communities. Specially, given one or more query images (probe), we need to find the correct images in the gallery that have different sizes of candidate person images obtained at different locations. Based on this, person re-identification also equates to a special image retrieval task. Similar to many visual recognition problems [1–7], the person's facial information, physical information, behavioral information, color information, texture information are used effectively, but variations in pose, viewpoints illumination and occlusion may make ReID non-trivial.

Recently, deep learning-based methods [8–15] are popular in visual tasks and have proved to be effective to delineate features for person re-identification. Even so, person ReID is still a challenging task. On the one hand, the methods based on Convolutional Neural Networks (CNNs) to extract person identity information are mainly divided into two categories: CNNs based on classification tasks [16] which train models using person IDs or attributes as training labels; or CNNs based on verification tasks [17] which input a pair of (two) person images for the networks to learn whether the two pictures refer to the same person or not. Some frameworks employed the Siamese model based verification task that needs extra steps to train CNN and process data. Although this model can increase the number of the training samples, it does not make good use of person identity information in the images. On the other hand, most exiting methods using square filters cannot handle the problem efficiently due to pose variations and perspective in cross-views. For example, as shown in Fig. 1, the two images refer to the same person. Because of the pose variations of one person across different views, the local features appearing in one view may not exactly be at the same position in the other view while it is very similar to some extent. The methods extracted the features from square filters [8–13] which lose more valid information, as shown in the red boxes of Fig. 1.

To overcome these issues, the Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN) based classification task is proposed that does not need to construct positive and negative sample pairs, which can greatly simplify data processing operations and make full use of different identity information to improve the performance. Specially, the asymmetric filters are designed in some critical convolution layers to preserve the horizontal features. For example, if we use the non-square filter, the

E-mail address: neverme@ouc.edu.cn (S. Wang).

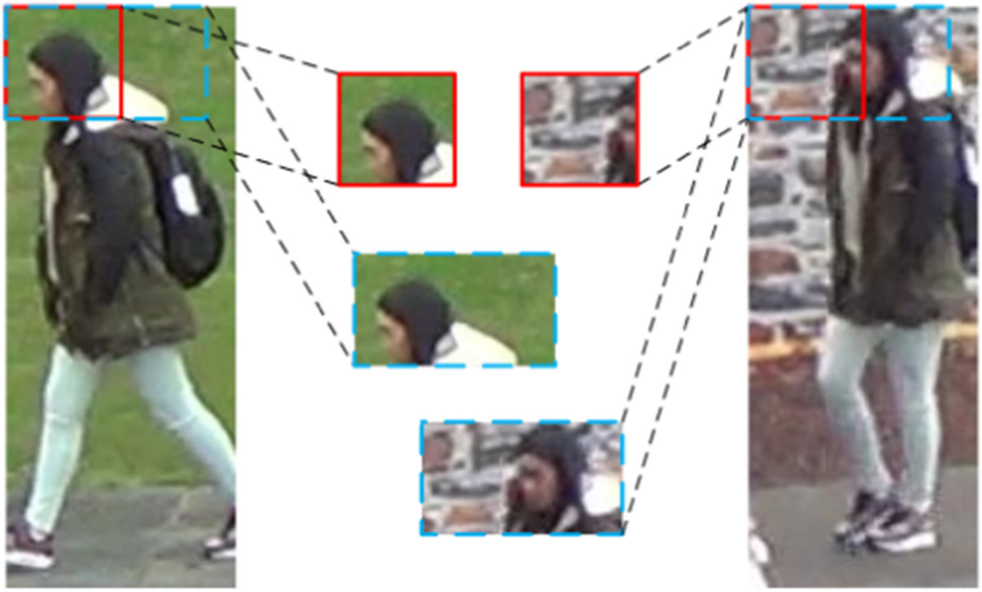


Fig. 1. The problem of a pair images. While the red box of left image can learn the head information, the red box of right image at same location only extracts several valuable information. If the filter changes to rectangle, the two images can gain similar information, as shown as the blue boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

patch corresponding to the head in two images (indicated by blue boxes shown in Fig. 1) has better features. Motivated by [18–20], we pay more attention to the field of blue boxes. What's more, AF D-CNN preserves the aspect ratio of person images, identity information of persons, which are significant for this task.

In this paper, we tackle person re-identification task by proposing a novel and robust network namely Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN). Multimodal features are extracted by a few of dense connections, and features of different semantic levels are combined to generate distinguishing feature maps. Not only this deep model exploits deep feature information to improve the quality of features, but also reduces the risk of vanishing gradients effectively. In order to make similar identities look closer and dissimilar pairs to be more distant in the representation, we go beyond the classic metric learning methods and apply a Joint Bayesian and re-ranking method to improving accuracy. Based on the learned Joint Bayesian model, the distributions of probe and gallery images can be obtained, and the derived closed-form expression of the log likelihood ratio, i.e. similarity, can be performed. The re-ranking method generates an optimized re-ranking list on the basis of the initial list. The overall framework of the proposed method namely ADJR is shown in Fig. 2.

In summary, the main contributions of this paper are as follows:

- We present a powerful architecture called Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN) that can obtain the multimodal features including color features, facial features, behavioral features, etc to address person re-identification problem and outperform some of the existing deep learning methods for person re-identification.

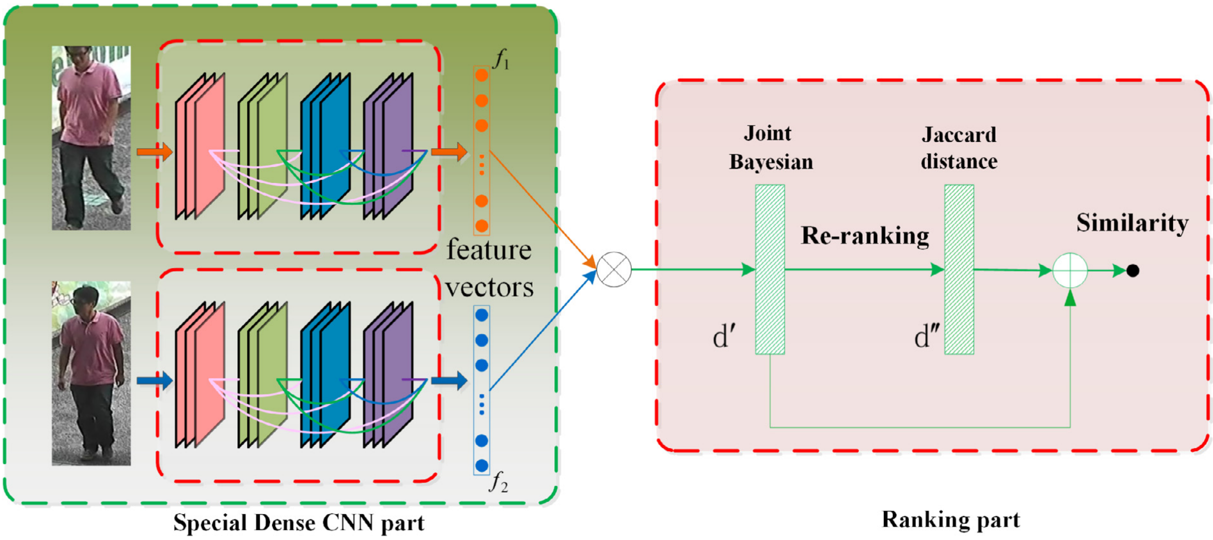


Fig. 2. The overall framework of the proposed method. In the Asymmetric Filtering-based Dense Convolutional Neural Networks, different colors denote different blocks, which are used to learn features extracted from probe and gallery images. Joint Bayesian and Jaccard distance in the Ranking part are employed to compute similarity for each pair. Finally, combining d' and d'' generates the refined ranking result. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- Instead of using traditional metric learning, Joint Bayesian treats all dimensions of the features equally and does not require dimensionality reduction. That is to say, it can explore the distinguishability of the sample features from the original features. To the best of our knowledge, this is the first time to apply Joint Bayesian in person Re-ID. Then re-ranking revises the first round results to produce a more robust list.
- Extensive experiments conducted on several datasets (CUHK01, CUHK03, Market-1501, DukeMTMC-reID) show that the proposed method achieves the state-of-the-art performance.

2. Related works

A considerable amount of papers have been published on Person Re-identification from various perspectives. Here, this section reviews and illustrates previous systems closely related to our work in this paper.

2.1. Traditional person re-identification methods

Numerous methods of person re-identification were proposed, and most of these existing methods typically contain two aspects: (1) Feature extraction: extracting more robust and powerful representations by handling the diversification in a person's appearance. (2) Metric learning: employing active metric learning to compute the similarity between person images. The goal of feature learning is to design more discriminative and stable feature descriptors for representing different persons. All the following stages depend on whether the stage of feature extraction obtain satisfied and efficient information or not. For example, Zhao et al. [21] presented a novel approach using mid-level filters for distinguishing persons and capturing good cross-view invariance. Mid-level filters with patch matching in the RankSVM (SVM-based ranking method) training generate the final scores that represent the similarity between image pairs. Liao et al. [22] proposed the KISSME method by learning a discriminant low dimensional subspace based on the Local Maximal Occurrence (LOMO) features, and generating a stable representation against viewpoint changes. The metric-based methods [23–26] pay more attention to seek an accurate similarity measure to reflect the difference in non-overlapping camera images. For example, Li et al. [25] reported the learning of Locally-Adaptive Decision Functions (LADF) for person verification, which can be considered as a joint model of a distance metric. A large margin nearest neighbor (LMNN) model [26] improved the traditional KNN (K Nearest Neighbor) classification and separated the matched neighbors from the mismatched ones by a large margin.

2.2. Deep learning for person re-identification

With an increasing size of the Re-ID datasets, the capability of DNN are clearly demonstrated. Both feature learning and metric learning [27,16,13,28–33] make use of neural networks and achieve a lot of improvements. Li et al. [16] extracted deep context-aware features of the body and latent parts for re-identification. Then they obtained the fusion of global and local features for powerful person representation. In DeepReID, Li et al. [13] employed a unified deep architecture with special layers to optimize feature learning, misalignment, occlusions and classification as a whole. Zhou et al. [28] achieved the task of learning features via Point to Set (P2S) metric and combined with part-based CNN for feature learning. Spindle Net merging global and local information [29] learnt feature representation. Results on a variety of challenging benchmarks with a rather diverse nature demonstrated the power of the mentioned method. Lin et al. [30] proposed a consistence-aware deep learning (CADL) approach to handle person re-identification with the main objective to search the global optimum in the whole camera network. To connect contextual information, Varior et al. [31] presented a novel Long Short-Term Memory (LSTM) architecture to get a more discriminative local feature representation. The dual-regularized KISS (DR-KISS) Metric Learning proposed by Tao et al. [27] exploits regularization to reduce the bias of two covariance matrices. This method showed that the regularization was necessary for generalization.

Although a deeper network can lead to better results, it may cause vanishing gradients. To cope with the aforementioned problem, Gao et al. [12] put forward Densely Connected Convolutional Networks (DenseNet), which shorten the paths from the early layers to the latter layers. However, most ReID methods use the classic deep networks directly and do not give specific optimization for the person re-identification problem. We here apply Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN) as the feature extractor and learn the Gaussian distribution of the samples from a large number of training data through Joint Bayesian and then calculate the distance (similarity) of the two images by employing a log-likelihood ratio. To enhance the accuracy of Re-ID, we revise the original ranking list with the proposed re-ranking method.

2.3. Joint Bayesian and re-ranking methods

Joint Bayesian has been highly successful for face verification, verifying whether two faces belong to the same person or not by employing a log-likelihood ratio [34]. Each face image is the fusion of inter-personal variations and intra-personal variations which were initialized by two independent Gaussians. The distributions of faces can be obtained by an appropriate prior on the face representation. Sun et al. [35–37] built a Joint Bayesian model to track face verification. Then they compared the ROC of Joint Bayesian with that of the other face verification methods and Joint Bayesian produce competitive performance.

In recent years, a number of previous approaches [38–45] exploit the similarity relationships between top ranked images (such as the k-nearest neighbors) in the initial ranking list. For example, Leng et al. [38] argued that the same person's images not only have similar visual contents but also neighboring contexts. Based on that, they employed content and context similarity to improve the initial ranking result, namely a bidirectional person re-identification technique. Garcia et al. [39] introduced an unsupervised post-ranking framework. To remove the visual ambiguities, discriminant context information is used to refine the initial ranking, leading to good performance. In [46], Ye et al. presented an approach to optimize the original ranks with two-view based ranking optimization. The main idea is that the two identical person images can share the similar reciprocal neighbors through cross KNN rank aggregation and graph-based re-ranking. Motivated by [47], Zhong et al. [48], who regard ReID as a retrieval problem, applied a k-reciprocal nearest neighbor technique to revise the original list which was produced by the Mahalanobis metric. In this paper, the local expansion query is carried out to obtain more robust k-reciprocal features. The final distance is calculated with the combination of the first distance and Jaccard distance.

3. Proposed method

Visual context is an important component to assist the ReID task. However, the real world situation is complex due to the variations in pose, illumination and background clutter. Directly using classification networks may not be effective to cope with these variations. To better learn discriminative features and release the mismatching problem, we will introduce the ADJR architecture which consists of the AF D-CNN, joint Bayesian and re-ranking. The architecture is discussed in the following several aspects.

3.1. AF D-CNN structure for feature extraction

Given a probe person image t and the gallery set $S = \{s_i | i = 1, 2, \ldots, n\}$ (where n is the number of the gallery images) captured by disjoint camera views, we aim to design a deep feature representation model for person re-identification matching under non-overlapping views. Hence, we formulate the AF D-CNN which inherits the advantages of DenseNet such as Dense Blocks and dense connections. Since the person ReID is not just a simple classification task, the original DenseNet has been improved in this paper. The structure of AF D-CNN is shown in Fig. 3. In order to be more effective to person ReID, both Dense Blocks and transition layers have been improved. The concrete details are shown in Table 1.

Unlike traditional classification networks, we adjust the size of the input images to non-square to keep the morphology of the person. All the inputs are resized to a resolution of 144 × 56. First, the low-level features are captured by the first convolution layer with the kernel size of 3 × 7. As shown in Fig. 1, we introduce the local features appearing in one view may differ from the features of the same location in another same pedestrian image due to the difference between the angle of view and pose variations. The AF D-CNN in this paper uses asymmetric convolution kernels at the critical convolutional layers to increase the range of receptive fields and then extract more horizontal feature information under each convolution calculation. Then we implement a pooling layer so as to reduce the resolution of the feature maps and apply four Dense Blocks (DB) to the multi-layers for acquiring the mixed features. The red box shows the connection of a Dense Block. Here the dense connections maximize the feature reuse and ensure feature discrimination. And the 1 × 1 convolution kernel here is used for integrating the features and reducing the dimensionality, which can learn the optimal features of global and local levels from the entire image and reduce the number of the parameters. The transition layer (Trans) consisting of Batch Normalization layers, two convolutional layers and an average pooling layer aims to reduce the number of the channels. In order to extract the characteristics of the person effectively, Trans2 and Trans3 employ a 1 × 1 convolution kernel and a 1 × 3 convolution kernel, as shown in Table 1. The deeper semantic information of person can be generated via several DBs and Trans blocks. Also, different levels specialise in processing different types of factors, for example, the bottom-layers represent low-level semantic attributes such as clothing color, and top-layers represent high-level semantic attributes such as object carrying and gender. The network employs dense connectivity proposed by Gao et al. [12] to cope with vanishing gradient, which all layers are connected to each other directly, as shown in the red box of Fig. 3. That each layer obtains additional inputs from all preceding layers and passes on its own feature maps to the following layers.

Table 1
The detailed network architecture of AF D-CNN.

Name	Output Size	Type of layers
Input	3 × 144 × 56	Data
Conv1	64 × 71 × 25	C_{3_7}, stride:2
Pool1	64 × 36 × 13	MaxPooling
DB1	448 × 36 × 13	$(N + C_{1\_1} + ReLU + C_{3\_3} + Contact) * 12$
Trans1	448 × 18 × 7	$N + C_{1\_1} + ReLU + C_{1\_1} + AvePooling$
DB2	832 × 18 × 7	$(N + C_{1\_1} + ReLU + C_{3\_3} + Contact) * 12$
Trans2	832 × 9 × 4	$N + C_{1\_1} + ReLU + C_{1\_3} + AvePooling$
DB3	1216 × 9 × 4	$(N + C_{1\_1} + ReLU + C_{3\_3} + Contact) * 12$
Trans3	1216 × 5 × 2	$N + C_{1\_1} + ReLU + C_{1\_3} + AvePooling$
DB4	1600 × 5 × 2	$(N + C_{1\_1} + ReLU + C_{3\_3} + Contact) * 12$
Pool5	1600 × 1 × 1	N + Global Avepooling
Feature	64	FC + Dropout
Classify	m	FC + Softmax with loss

Note that N represents a connection sequence of batch normalization, Scale and ReLU, C_{i_j} indicates the kernel size is $i \times j$ and the convolution layers of C_{1_3} and C_{3_3} default to one pixel padding to adapt the fixed size of the feature maps while other convolution layers have no padding.

As seen in Table 1, we pay attention to the amount of the parameters and the network width. The objective of the overall

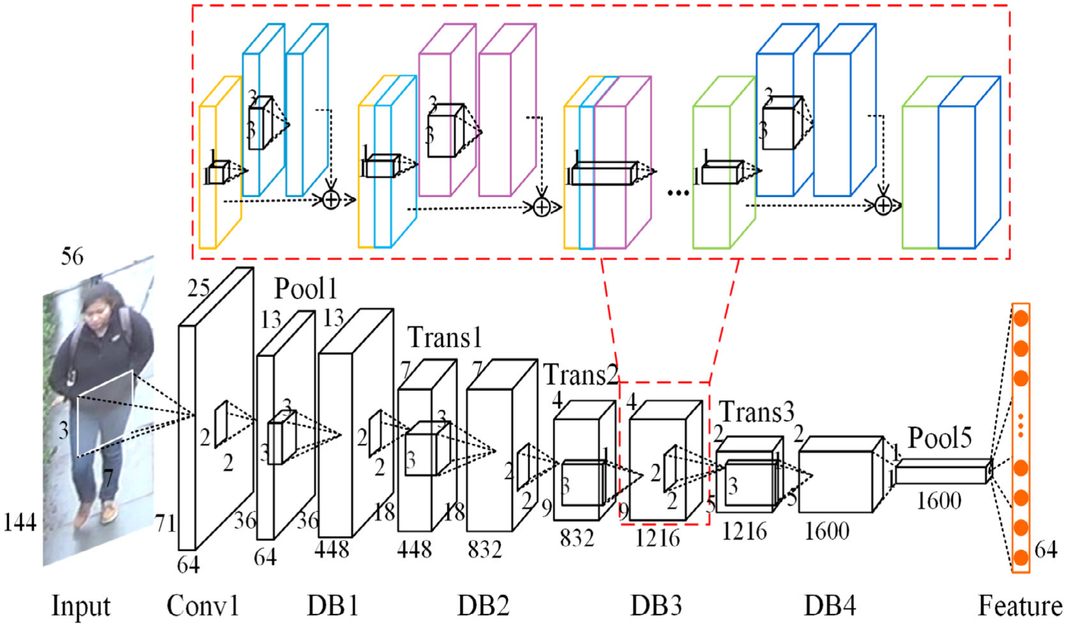


Fig. 3. The architecture of AF D-CNN. The process of feature extraction is shown as above, which takes 144 × 56 inputs and output a $64 - d$ feature vector. Asymmetric filters are used in some critical convolution layers to preserve the horizontal features. The details of a Dense Block (DB) are shown as the red box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

architecture composition is to minimize the model complexity and therefore reduce the network parameter size whilst maintaining the optimal network depth. We employ the batch normalization, Scale and ReLU layers to accelerating the training procedure with standardizing the distribution of inputs of each layer. ReLU (Rectified Linear Unit) is used to decrease the risk of overfitting and reduces the computation between two convolution layers. Here we use a different connection from DenseNet, yielding good performance. To integrate different context information, the DBs, Trans blocks and FC operations play an important role in feature learning.

Finally, a softmax loss function is used to encourage discriminative features. Different from the other networks which always add an activation layer after the feature layer, we here argue that the activation layer reduces the strengths of softmax and cannot enhance the distinguishability between the samples by many experiments. To better enforce this constraint, we remove the activation function between the first fully connected and the second fully connected layers. As observed, one $64-d$ feature vector which simplifies the calculation of the subsequent similarity measures can be extracted from each image by our AF D-CNN.

Afterwards, we output a probability distribution and each probability represents the probability of a certain category. The network parameters are optimized by minimizing the cross-entropy loss, which is defined as follows:

$$L(f,y,\theta_{id}) = -\sum_{i=1}^{n} p_i log q_i \tag{1}$$

where f is the extracted feature vector by AF D-CNN, y is the target class, and θ_{id} is the parameter of the softmax layer, i is the $i-th$ category, p_i is the true probability distribution, and q_i is the predicted probability of the proposal. The ground truth label $p_i = 0$ if $i \neq y$, and $p_i = 1$ if $i = y$ for the target class. This loss function is optimized by using back-propagation (BP) and stochastic gradient descent (SGD) jointly.

3.2. Joint Bayesian and re-ranking

This study adopts a Joint Bayesian technique for person Re-ID to compute the similarity on the basis of the extracted AF D-CNN features. According to [34–37], the features of a person image are influenced by two factors: intra- and inter-personal variations. A pedestrian image is described by the sum of two independent Gaussian variables μ and ϵ , where $\mu \sim N(0, S_\mu)$ denotes the identity information, $\epsilon \sim N(0, S_\epsilon)$ denotes the variation information (i.e. pose and illumination). Based on [29], we regard the above representation and assumptions as a prior. Given two images $\{t, s_i\}$, Joint Bayesian generates a similarity that measures the distance between feature vectors t and s_i. Joint Bayesian models the joint probability of two faces given the intra or extra-personal variation hypothesis, $p(t, s_i|H_I)$ and $p(t, s_i|H_\epsilon)$. Typically, the space of the pairwise differences $\Delta = t - s_i$. In order to address the problem of reducing the separability and degrading the accuracy, the distance of the two images is defined by testing a log-likelihood ratio:

$$d'(t, s_i) = log\frac{P(t, s_i|H_I)}{P(t, s_i|H_\epsilon)} = (t - s_i)^T A(t - s_i) + 2t^T(A - G)s_i \tag{2}$$

where A and G are semi-definite matrices and can be estimated by Table 2. More importantly, this formulation preserves the separability.

Through computing the distances $d(t, s_i)$ by the Joint Bayesian model, the first ranking list can be obtained as $d'_t(S) = \{s_1^0, s_2^0, \ldots, s_i^0\}$, where t indicates the probe image (target image), s_i^0 denotes the $i-th$ image of the original list and the smaller subscripts are corresponding to the higher similarities.

Table 2
The Joint Bayesian algorithm.

```
Input: Training data $x_i$, initialized $S_\mu, S_\epsilon$ by positive definite matrix,
$t$    0
While not converge do
    $t \leftarrow t + 1$
    $F = S_\epsilon^{-1}, G = -(x_i S_\mu + S_\epsilon)^{-1} S_\mu S_\epsilon^{-1}$
    $\mu_i = \sum_{j=1}^{m_i} S_\mu (F + m_i G) x_j, \epsilon_{ij} = x_j + \sum_{j=1}^{m_i} S_\mu G x_j$
    Update the parameters $S_\mu$ by $S_\mu = \frac{1}{n}\sum_i \mu_i \mu_i^T$
    Update the parameters $S_\epsilon$ by $S_\epsilon = \frac{1}{n}\sum_i \sum_j \epsilon_{ij} \mu_{ij}^T$
end while
$F = S_\epsilon^{-1}, G = -(2S_\mu + S_\epsilon)^{-1} S_\mu S_\epsilon^{-1}$
$A = (S_\mu + S_\epsilon)^{-1} - (F + G)$
Output: A, G
```

Aiming to generate better matching results, we apply the re-ranking technique reported in [48] which utilizes k-reciprocal Nearest Neighbors and Jaccard distance. We query half of the k-reciprocal nearest neighbors of every candidate and further add this set to a larger set $R''_t(k)$. Next, we need to compute the distance between $R_t(k)$ (the k-reciprocal Nearest Neighbors) and $R''_t(k)$ that some common images. Then, the Jaccard distance between t and s_i is defined as:

$$d''(t, s_i) = 1 - \frac{\sum_N^{j=1} min(f_{t,s_i}, f_{s_i,s_i})}{\sum_N^{j=1} min(f_{t,s_i}, f_{s_i,s_i})} \tag{3}$$

where $f_{t,s_i} = \begin{cases} e^{-d'(t,s_i)} & if\ s_i \in R''_t(k) \\ 0 & otherwise \end{cases}$, f is the feature vector.

Hence, this paper takes the initial ranking result and Jaccard distance into account. The final distance d is depicted as

$$d(t, s_i) = \varphi d'(t, s_i) + (1 - \varphi) d''(t, s_i)(0 \leqslant \varphi \leqslant 1) \tag{4}$$

where $d'(t, s_i)$ is the original distance, $d''(t, s_i)$ is the re-ranking distance. φ is the penalty parameters that represent the proportion of the first distance and Jaccard distance.

4. Experiments

To evaluate the performance of our ADJR, we conduct experiments on various benchmarks to show the generalization of our method and compare our method with other methods for person re-identification.

4.1. Datasets and protocols

Recently, some datasets experience certain changes: (1) The size of datasets are expanding. For example, the CUHK03, Market-1501 and DukeMTMC-reID possess 1000+ identities. (2) The method of obtaining bounding boxes tends to make use of person detectors rather than human labors. (3) More cameras will be used when collecting datasets. The dataset, DukeMTMC-reID, has up to eight cameras, which requires the method to have a good generalization ability. We choose four datasets (i.e. CUHK01 [49], CUHK03 [13], Market-1501 [50], DukeMTMC-reID [51]) to evaluate our method, since their low image resolution, illuminations, and viewpoint variations are close to practical surveillance scenes (see Fig. 4). Each of them has at least one image for each person from each camera view.

The CUHK01 dataset is captured with two disjoint cameras in a campus environment. It contains 971 persons with 3,884 images, and there are two images for each person under every camera view. The first camera obtains the side view of persons and the second camera captures the front and back views. All the images are

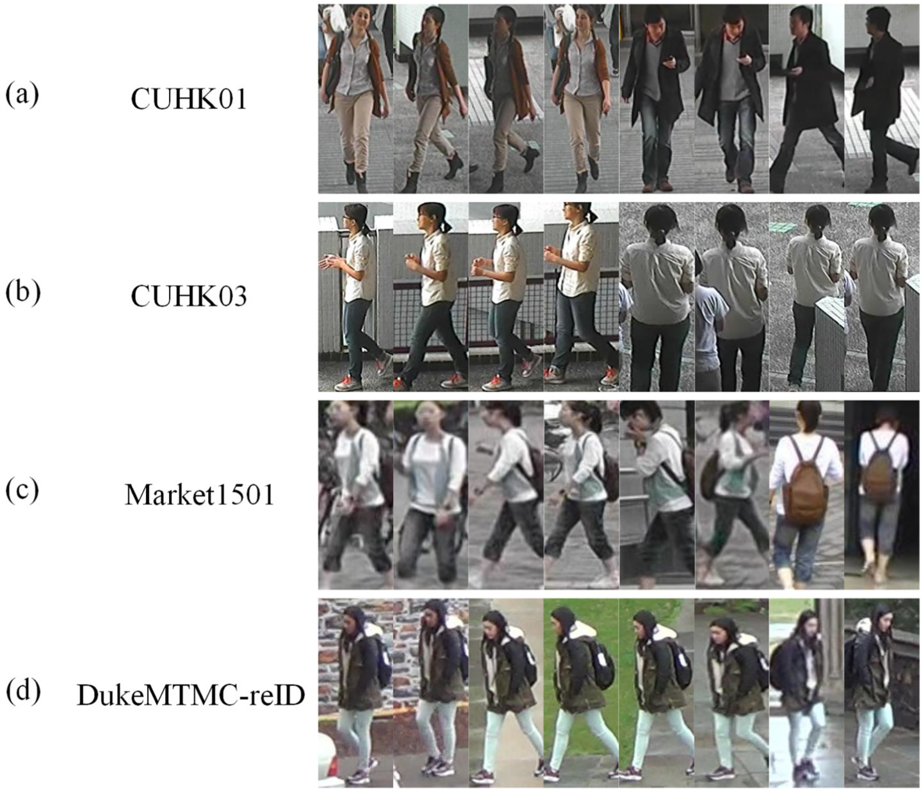


Fig. 4. Sample images from four datasets. (a) CUHK01: There are two cameras and each person has two images under every camera view. It shows eight images from two persons. (b) CUHK03: The images are captured from five pairs of cameras. (c) Market1501: The dataset is captured by five high resolution cameras and one low resolution camera. (d) DukeMTMC-reID: The images are obtained from different views and the number of each person pictures vary greatly.

normalized to 160 × 60 for experiments. Not much data available for training is the main challenge of the CUHK01 dataset.

The CUHK03 dataset is one of the largest published person re-identification datasets. It contains 1467 individuals obtained from five pairs of cameras on campus without overlapping of person identities. There are 13,164 person images in total. Different from the CUHK01 dataset, this dataset provided two versions of annotations: CUHK03-labeled and CUHK03-detected with manually labeled bounding boxes and bounding boxes detected using Deformable Part based Model (DPM). In this paper, we use the mixture of the two datasets.

The Market-1501 has 32,668 images of 1501 identities captured by six cameras at Tsinghua University. This dataset employs DPM as the person detector to obtain the bounding boxes, which is not as ideal as human annotation ones but is close to the real world. Besides, this dataset also contains false alarm results, which makes the dataset more challenging.

DukeMTMC-reID has the same format as that of the Market-1501 dataset published by Zheng et al. in 2017. The new dataset has 36,411 images of 1812 identities captured by 8 different cameras. There are 1404 identities appearing in more than two cameras and 408 identities appearing in one camera. This dataset has manually defined bounding boxes. Some examples of the DukeMTMC-reID dataset are shown in Fig. 4(d). It is obvious that this dataset is challenging due to the cluttered backgrounds, severe occlusions, illumination changes and pose variations.

Average performance will be shown using Cumulative Matching Characteristic (CMC) curves, which describes the expectation of the true match within the first r ranks. Additionally, the mean average precision (mAP) is also reported along with the Rank-1 (the query identity occurs in the first candidate lists) accuracy on the Market-1501 and DukeMTMC-reID datasets. The mAP considers both precision and recall rates, which are complementary to CMC.

4.2. Implementation details

4.2.1. The implementation details of AF D-CNN

We select Caffe [52] deep learning framework to implement the experiments. When operating the Asymmetric Filtering-based Dense Convolutional Neural Network (AF D-CNN) structure to learn valid identity information of persons, we make use of mini-batch stochastic gradient descent (SGD) for faster back propagation and smoother convergence [53]. In the training stage, each mini-batch consists of 50 images. The learning rate is initialized to 0.01 and decreased to 10% after every 10,000 iterations. Alternatively, to overcome data imbalance and overfitting, we enlarge the dataset by introducing random cropping and mirror operations which is commonly used in [26,54].

We firstly train the CUHK03 dataset and then use the other datasets for fine-tuning. In the CUHK03 dataset, we randomly divide 1467 persons into two parts. One part takes 21,003 images of 1367 identities as the train set for training AF D-CNN and Joint Bayesian, and takes 5250 images as the validation set to verify the accuracy of the classification network. Another part namely testing set contains 100 persons with 5250 images, and we evaluate the capability of person re-identification methods. Through 35,000 iterations of training, the classification accuracy of our AF D-CNN reaches 99.8%. After the training steps, powerful features are extracted from the train and test sets by using the gained network model and each person is represented by a feature vector.

For Market-1501, CUHK01 and DukeMTMC-reID datasets, we also randomly split them into the train and test parts. The partition is described in Table 3, where $*$ denotes the quantity of persons, ID represents different identities, AF D-CNN is our proposed network, probe is the target image, and gallery is a retrieval dataset.

Other datasets generate their models after 35,000 iterations based on the pre-trained model using the CUHK03 dataset. Next,

Table 3
The details of four datasets evaluated in our experiment. The number of train/validation images, together with the number if query/gallery identities and images are listed.

Dataset	*ID	*AF D-CNN			*Test	
		*Train ID	*Train images	*Val images	*Probe ID	*Gallery ID
CUHK03	1467	1367	21,003	5250	100	100
Market-1501	1501	751	10,394	2542	750	750
CUHK01	971	486	1552	388	485	485
DukeMTMC-reID	1812	702	16,522	0	702	1110

good features are learned in the feature layer by propagating forward once.

4.2.2. The implementation details of Joint Bayesian and re-ranking

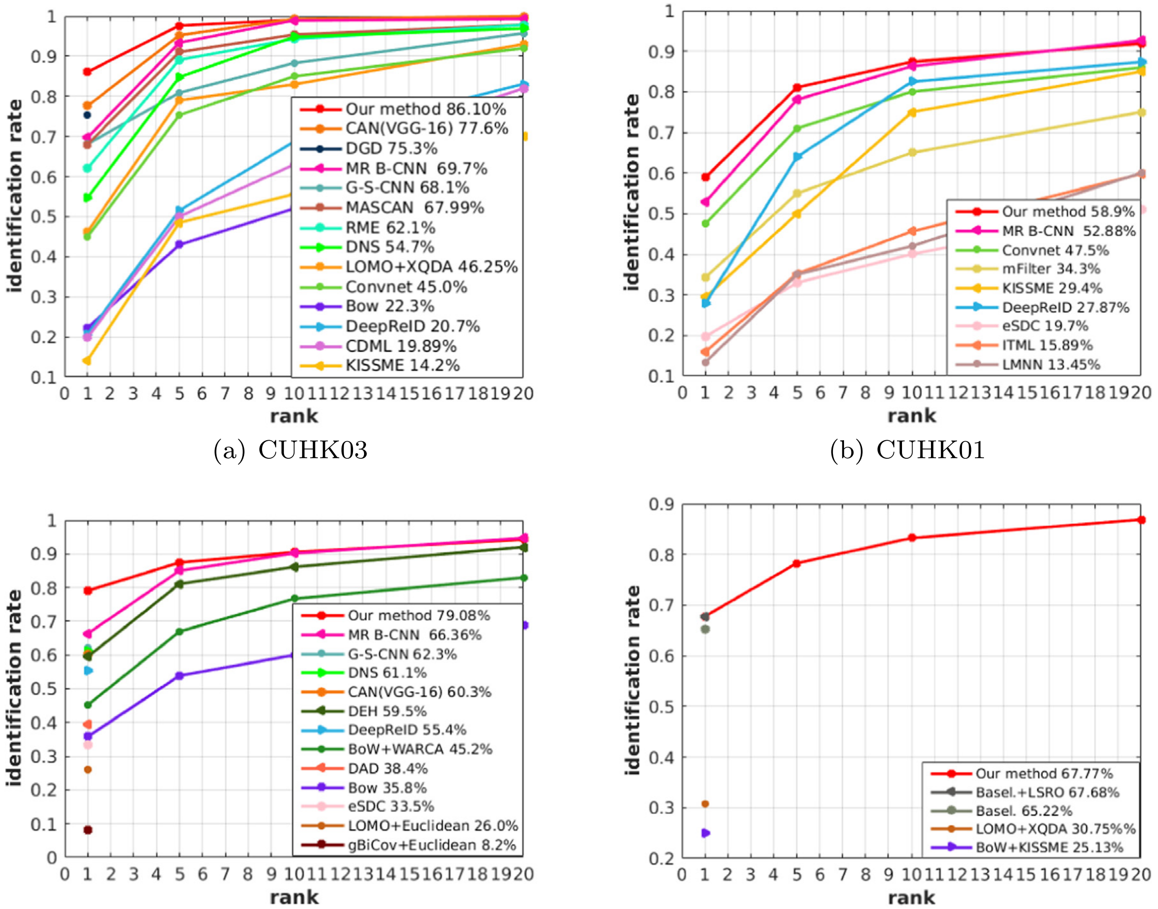
In this section, we assume every image can be expressed by $x = \mu + \epsilon$, which follows Guassian with zero mean. In the training of Joint Bayesian, we compute the mean value of the overall samples and present a person using the whole sample minus the mean value. Similarly, that each person's features minus the own mean value produces two independent parts μ and ϵ. We can obtain a $64 - d$ vector to represent one person after applying AF D-CNN. Finally, the parameters of the Gaussian distribution can be estimated by the standard Expectation Maximization Algorithm (EM algorithm). The similarity between two images is obtained using the log-likelihood ratio.

Re-ranking needs to be based on searching k nearest neighbors of the probe images. So we set $k = 20$. It is proved that re-ranking is benefit for the ADJR structure, so we incorporate the Joint Bayesian distance and the Jaccard distance to generate the final distance and the weight of two distances depend on hyper-parameter φ. Conducting a number of experiments, different datasets show various results with different values (0.3, 0.75, 0.85). For example, the hyper-parameter φ of the CUHK03 and Market-1501 datasets are set to be 0.3, while φ of the CUHK01 dataset is set to be 0.85 and φ of DukeMTMC-reID is set to be 0.75.

4.3. Comparison with state-of-the-art methods

We have conducted a number of experiments with our ADJR architecture on the four datasets. To illustrate the effectiveness of our model, we evaluate our method by using Rank-1 accuracy for the four datasets. For the Market-1501 and the DukeMTMC-reID, we also show the mean average precision. The CMC of our ADJR is shown in Fig. 5.

CUHK03: For the CUHK03 dataset, we compare our method with several existing methods including CAN [55], DGD [22], MR B-CNN [9], G-S-CNN [17], MASCAN [16], RME [56], DNS [57], LOMO + XQDA [22], Convnet [11], Bow [50], DeepReID [13], CDML



(a) CUHK03

(b) CUHK01

(c) Market-1501

(d) DukeMTMC-reID

Fig. 5. Experimental results of our ADJR methods and other comparisons on four datasets. The CMC Top-1-5-10-20 accuracies are listed. The Top-1 accuracy of the best performance is reported.

[10], and KISSME [58]. On this dataset, we conduct experiments on both the detected and the labeled datasets. Fig. 5(a) plots the Rank-1 identification rates on the CUHK03 dataset. Our ADJR method outperforms the previous methods including deep learning methods.

From Fig. 5(a), we achieve the state-of-the-art result with Rank-1 recognition accuracy of 86.1%. Given a probe image, the probability that we search top 20 similar person images in the gallery is 99%. In fact, our ADJR improves over the CAN by a margin of 8.5%, which uses End-to-End Comparative Attention Networks for Person Re-Identification. Compared with the other standard methods, our approach obtains state-of-the-art performance.

CUHK01: Since the number of the CUHK01 dataset does not suit a deep architecture, we pre-train a network on the CUHK03 and adapt it for the CUHK01 by fine-tuning the parameters. Also we compare our methods with several existing methods, like MR B-CNN [9], Convnet [11], mFilter [21], KISSME [58], DeepReID [13], eSDC [59], ITML [60], LMNN [61]. The CMC is shown in Fig. 5(b).

Compared with MR B-CNN, Convert, KISSME, DeepReID, eSDC, LMNN, our ADJR framework does obtain the best result. It is shown that our final method reaches 58.9% accuracy, better than MR B-CNN that uses Multi-Region Bilinear Convolutional Neural Networks. Because of lack of training samples in the CUHK01, the ADJR framework has much potential to explore.

Market-1501: For the Market1501 dataset, several classical methods are compared, including MR B-CNN [9], G-S-CNN [17], DNS [57], CAN [55], DEH [62], Bow + WARCA [63], DAD [64], Bow [50], eSDC [59], LOMO + Euclidean [50], gBiCov + Euclidean [13]. The CMC results on the Market-1501 are shown in Fig. 5(c). It is noticed that the proposed framework effectively handles the person images detected by the DPM algorithm. Importantly, the method proposed in this paper beats the other approaches, such as some deep learning methods MR B-CNN, G-S-CNN, CAN and so on. Specially, the ADJR framework gains 79.08% Rank-1 accuracy, improving 43.28% than Bow. Also the mAP improves about 22.7% compared with the 2^{nd} best result produced by MR B-CNN. Compared with MR B-CNN and G-S-CNN, the proposed approach improves Rank-1 identification rate about 12.72% and 16.78%. Additionally, the ADJR framework is superior to other typical approaches. Experiments suggest that AF D-CNN is more suitable for person re-identification, and Joint Bayesian combined with re-ranking can enhance the recognition capability. Table 4 gives the mean average precision value additionally. From that, it can be observed that our method significantly outperforms the existing person re-identification approaches. For instance, ADJR boosts the mAP by 22.7% than MR B-CNN which achieved the best performance compared with G-S-CNN, DNS, CAN(VGG-16) and other methods (DAD, Bow, eSDC, LOMO + Euclidean, gBiCov + Euclidean).

DukeMTMC-reID: This dataset is a large dataset, captured from high-resolution videos. We evaluate our method with Basel.+LSRO [51], Basel. [65], LOMO + XQDA [22], BoW + KISSME [50]. The results from the DukeMTMC-reID dataset are summarized in Fig. 5(d).

Table 4
Comparison of various methods performance on Market-1501 dataset.

Methods	mAP
gBiCov + Euclidean	2.23%
LOMO + Euclidean	7.75%
eSDC	13.5%
Bow	14.75%
DAD	19.6%
CAN(VGG-16)	35.9%
DNS	35.7%
G-S-CNN	39.55%
MR B-CNN	41.17%
ADJR	63.87%

From the results, we can see that our framework outperforms the previous methods, performing 67.77% in Rank-1 accuracy. In particular, our framework outperforms the Basel. and Basel.+LSRO with 2.55% and 0.09% in top 1 accuracy, respectively. Similarity, the mAP (see as Table 5) improves by 5.4% compared against the two methods. Compared with LOMO + XQDA and BoW + KISSME, the ADJR framework exceeds them greatly.

4.4. Experiment analysis

4.4.1. Comparison with deep learning network

The proposed AF D-CNN has certain advantages, such as simpler training, lower feature dimension, suitable for ReID and so on. This paper shows some results with AlexNet [54], ResNet [66] and AF D-CNN. In the experiments, we select the Market-1501 dataset to implement experiments and select Euclidean distance to calculate similarity. The performance is shown in Fig. 6.

Obviously, the AF D-CNN is superior to AlexNet and ResNet-50. In our network, we maintain the distinguished features via integrating asymmetric convolution kernel with different channels and eliminating the activation layer, which adapts the network to the task of ReID. The AF D-CNN boosts the Rank-1 matching rate by 3.18%. For the mAP evaluation (see as Table 6), the same conclusion can be made.

4.4.2. Analysis of Joint Bayesian and re-ranking results

We perform a series of experiments on the CUHK03 dataset to quantify the effect of Joint Bayesian and re-ranking components. Note that this section needs to use the same network model to extract features. Different combinations of Euclidean distance, Joint Bayesian and re-ranking generate different results which show Rank-1 recognition rates in Table 7.

Table 5
Comparison of various methods performance on Market-1501 dataset.

Methods	mAP
BoW + KISSME	12.17%
LOMO + XQDA	17.04%
Basel.	44.99%
Basel. + LSRO	47.13%
ADJR	52.53%

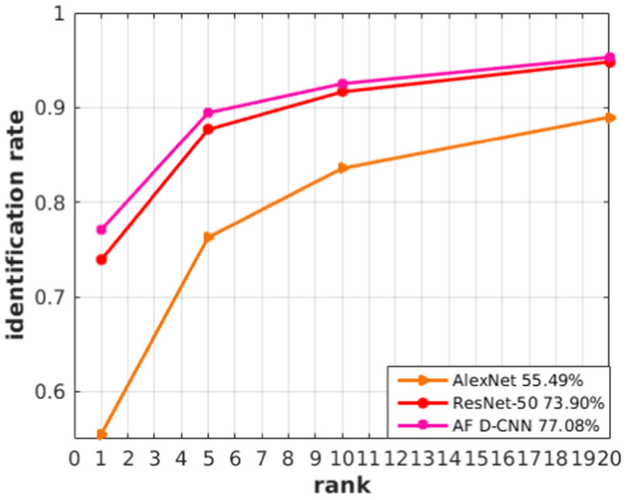


Fig. 6. The CMC curves of three deep learning network on the Market-1501 dataset.

Table 6
Comparison with the mAP of three deep network structures.

Deep learning framework	mAP
AlexNet	32.36%
ResNet-50	47.78%
AF D-CNN	63.87%

Table 7
The results of different combinations in CUHK03 dataset.

	Euclidean distance	Joint Bayesian	Re-ranking	Rank-1
1	✓			77.30%
2		✓		77.80%
3	✓		✓	85.90%
4		✓	✓	86.10%

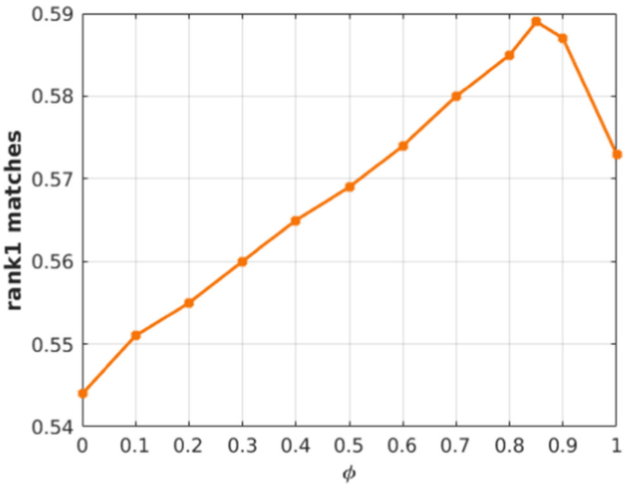


Fig. 7. The influence of different φ on CUHK01.

From the first row of Table 7, we can see that the proposed AF D-CNN in this paper can extract features well and yield good matching rates, even though it only uses Euclidean distance. Although Joint Bayesian improves a little (see the second row), it is valuable for ReID. It clearly demonstrates that re-ranking is beneficial to exploit the local and whole information via comparing rows 2 with 4 or rows 1 with 3. The last row integrates new components, leading to better performance.

4.4.3. *Hyper-parameter φ of final distance*

As observed in our experiments, the weight parameter φ has major effects on our method. In the following, we give an empirical analysis on the method on the CUHK01. The influence of parameter φ is shown in Fig. 7, in which we analyze the influence by changing different values.

When φ is 0, the final result only takes Jaccard distance into account. Similarity, when φ is 1, we only consider feature representations and metric learning. From Fig. 7, we argue that if φ increases, the Rank-1 matching rate also improves significantly. The best result is obtained when φ is 0.85, that is, 0.85 is an optimal value to exploit their function.

5. Conclusion

In this work, we have presented a novel framework namely ADJR for person re-identification. Specifically, our Asymmetric Filtering-based Dense Convolutional Neural Network preserves the aspect ratio of person images, identity information of persons, and distinguishable information between persons, which is important for learning multimodal features. Compared with the thousands-dimension feature extracted by traditional deep networks, the features extracted by AF D-CNN are only 64-dimension. It removes the redundant information and protects the core information such as identity. Joint Bayesian that retains the dimensionality combined with k-reciprocal nearest neighbors which reduces mismatching produces satisfactory performance. Extensive evaluation has been conducted using four public benchmark datasets (CUHK01, CUHK03, Market-1501, DukeMTMC-reID). Results demonstrated that for three dataset including CUHK03, Market-1501, DukeMTMC-reID, baseline models performance obtains State-of-the-art performance in Rank-1 accuracy or the mAP. In the future, we use this framework to process the video to get multimodal features directly, which has great significance for catching fugitives.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) Grants U1706218, 61602229, 41606198, 61501417 and 41706010, Natural Science Foundation of Shandong Provincial ZR2016FM13, ZR2016FB02. H. Zhou was supported in part by the European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska-Curie grant agreement No 720325 FoodSmartphone, the UK EPSRC under Grant EP/N011074/1 and the Royal Society-Newton Advanced Fellowship under Grant NA160342.

References

[1] X. Song, F. Feng, X. Han, X. Yang, W. Liu, L. Nie, Neural compatibility modeling with attentive knowledge distillation, 2018.
[2] J.L.Z.L.L.N.J.M. Xuemeng Song, Fuli Feng, Neurostylist: Neural compatibility modeling for clothing matching, in: ACM International Conference on Multimedia, 2017.
[3] L. Nie, L. Zhang, Y. Yan, X. Chang, M. Liu, L. Shao, Multiview physician-specific attributes fusion for health seeking, IEEE Trans. Cybern. PP (99) (2017) 1–12.
[4] P. Jing, Y. Su, L. Nie, X. Bai, J. Liu, M. Wang, Low-rank multi-view embedding learning for micro-video popularity prediction, IEEE Trans. Knowledge Data Eng. PP (99) (2018) 1519–1532.
[5] P. Jing, Y. Su, L. Nie, H. Gu, J. Liu, M. Wang, A framework of joint low-rank and sparse regression for image memorability prediction, IEEE Trans. Circ. Syst. Video Technol. PP (99) (2018) 1.
[6] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, X. Li, Modeling disease progression via multisource multitask learners: A case study with alzheimer's disease, IEEE Trans. Neural Networks Learn. Syst. 28 (7) (2016) 1508–1519.
[7] L. Nie, L. Zhang, Y. Yan, X. Chang, M. Liu, L. Shao, Multiview physician-specific attributes fusion for health seeking, IEEE Trans. Cybern. PP (99) (2017) 1–12.
[8] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based cnn with improved triplet loss function, in: Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.
[9] E. Ustinova, Y. Ganin, V. Lempitsky, Multi-region bilinear convolutional neural networks for person re-identification, Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on. IEEE, 2017 pp. 1–6.
[10] H. Shi, X. Zhu, S. Liao, Z. Lei, Y. Yang, S.Z. Li, Constrained deep metric learning for person re-identification, Comput. Sci. (2015) 34–39.
[11] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, in: Computer Vision and Pattern Recognition, 2015, pp. 3908–3916.
[12] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: CVPR, 2017.
[13] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: Deep filter pairing neural network for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159.
[14] Y. Yang, S. Liao, Z. Lei, S.Z. Li, Large scale similarity learning using similar pairs for person verification, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 3655–3661.

[15] S. Wu, Y.C. Chen, X. Li, A.C. Wu, J.J. You, W.S. Zheng, An enhanced deep feature representation for person re-identification, in: Applications of Computer Vision, 2016, pp. 1–8.

[16] D. Li, X. Chen, Z. Zhang, K. Huang, Learning deep context-aware features over body and latent parts for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 384–393.

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

[18] Z. Cheng, J. Shen, On effective location-aware music recommendation, Acm Trans. Informat. Syst. 34 (2) (2016) 1–32.

[19] Z. Cheng, Y. Ding, L. Zhu, M. Kankanhalli, Aspect-aware latent factor model: Rating prediction with ratings and reviews.

[20] X.H.L.Z.X.S.M.S.K. Zhiyong Cheng, Ying Ding, A∧3ncf: An adaptive aspect attention model for rating prediction., in: Proceedings of the 27-th International Joint Conference on Artificial Intelligence, 2018.

[21] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person re-identification, in: Computer Vision and Pattern Recognition, 2014, pp. 144–151.

[22] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.

[23] A. Globerson, S.T. Roweis, Metric learning by collapsing classes, Adv. Neural Informat. Process. Syst. 18 (2005) 451–458.

[24] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person Re-Identification Using Kernel-Based Metric Learning Methods, Springer International Publishing, 2014.

[25] K.Q. Weinberger, L.K. Saul, Distance Metric Learning for Large Margin Nearest Neighbor Classification, JMLR.org, 2009.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European Conference on Computer Vision, 2016, pp. 21–37.

[27] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, Y.Y. Tang, Person re-identification by dual-regularized kiss metric learning, IEEE Trans. Image Process. 25 (6) (2016) 2726–2738.

[28] S. Zhou, J. Wang, J. Wang, Y. Gong, N. Zheng, Point to set similarity based deep feature learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5028–5037.

[29] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, X. Tang, Spindle net: Person re-identification with human body region guided feature decomposition and fusion, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 907–915.

[30] J. Lin, L. Ren, J. Lu, J. Feng, J. Zhou, Consistent-aware deep learning for person re-identification in a camera network, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[31] R.R. Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A siamese long short-term memory architecture for human re-identification, in: European Conference on Computer Vision, 2016, pp. 135–153.

[32] T. Xiao, H. Li, W. Ouyang, X.Wang, Learning deep feature representations with domain guided dropout for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1249–1258.

[33] L. Ma, X. Yang, D. Tao, Person re-identification over camera networks using multi-task distance metric learning, IEEE Trans. Image Process. 23 (8) (2014) 3656–3670.

[34] D. Chen, X. Cao, L. Wang, F. Wen, J. Sun, Bayesian face revisited: A joint formulation, in: European Conference on Computer Vision, 2012, pp. 566–579.

[35] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.

[36] Y. Sun, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, Adv. Neural Informat. Process. Syst. 27 (2014) 1988–1996.

[37] Y. Sun, X. Wang, X. Tang, Deeply learned face representations are sparse, selective, and robust, in: Computer Vision and Pattern Recognition, 2015, pp. 2892–2900.

[38] Q. Leng, R. Hu, C. Liang, Y. Wang, J. Chen, Person re-identification with content and context re-ranking, Multimedia Tools Appl. 74 (17) (2015) 6989–7014.

[39] J. Garcia, N. Martinel, A. Gardel, I. Bravo, G.L. Foresti, C. Micheloni, Discriminant context information analysis for post-ranking person re-identification, IEEE Trans. Image Process. 26 (4) (2017) 1650.

[40] V.H. Nguyen, T.D. Ngo, K.M.T.T. Nguyen, D.A. Duong, Re-ranking for person re-identification, in: Soft Computing and Pattern Recognition, 2015, pp. 304–308.

[41] A.J. Ma, P. Li, Query based adaptive re-ranking for person re-identification, Lect. Notes Comput. Sci. 9007 (2014) 397–412.

[42] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, Q. Tian, Query-adaptive late fusion for image search and person re-identification, in: Computer Vision and Pattern Recognition, 2015, pp. 1741–1750.

[43] W. Li, Y. Wu, M. Mukunoki, M. Minoh, Common-near-neighbor analysis for person re-identification, in: IEEE International Conference on Image Processing, 2013, pp. 1621–1624.

[44] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person re-identification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Trans. Multimedia PP (99) (2016) 1.

[45] J. Garcia, N. Martinel, C. Micheloni, A. Gardel, Person re-identification ranking optimisation by discriminant context information analysis, in: IEEE International Conference on Computer Vision, 2015, pp. 1305–1313.

[46] M. Ye, J. Chen, Q. Leng, C. Liang, Z. Wang, K. Sun, Coupled-view based ranking optimization for person re-identification, International Conference on Multimedia Modeling, 2015, pp. 105–117.

[47] D. Qin, S. Gammeter, L. Bossard, T. Quack, L.V. Gool, Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors, in: Computer Vision and Pattern Recognition, 2011, pp. 777–784.

[48] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[49] W. Li, X. Wang, Locally aligned feature transforms across views, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3594–3601.

[50] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: IEEE International Conference on Computer Vision, 2016, pp. 1116–1124.

[51] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro.

[52] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Caffe: Convolutional architecture for fast feature embedding, 2014, pp. 675–678.

[53] L. Bottou, Stochastic Gradient Descent Tricks, Springer, Berlin Heidelberg, 2012.

[54] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.

[55] H. Liu, J. Feng, M. Qi, J. Jiang, S. Yan, End-to-end comparative attention networks for person re-identification, IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc. PP (99) (2016) 1.

[56] S. Paisitkriangkrai, C. Shen, A. V. D. Hengel, Learning to rank in person re-identification with metric ensembles, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1846–1855.

[57] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.

[58] M. Kstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.

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

[60] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: ICML '07: Proceedings of the International Conference on Machine Learning, 2007, pp. 209–216.

[61] M. Hirzer, P.M. Roth, H. Bischof, Person re-identification by efficient impostor-based metric learning, in: IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, 2012, pp. 203–208.

[62] E. Ustinova, V. Lempitsky, Learning deep embeddings with histogram loss.

[63] C. Jose, F. Fleuret, Scalable metric learning via weighted approximate rank component analysis, in: European Conference on Computer Vision, 2016, pp. 875–890.

[64] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: European Conference on Computer Vision, 2016, pp. 475–491.

[65] L. Zheng, Y. Yang, A.G. Hauptmann, Person re-identification: Past, present and future.

[66] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.