Single Image Based Metric Learning via Overlapping Blocks Model for Person Re-Identification

Yipeng Chen, Cairong Zhao*, Tianli Sun
Department of Computer Science and Technology, Tongji University, Shanghai, China
E-mail:zhaocairong@tongji.edu.cn

Abstract

Considering the pedestrian structure characteristics, the first step of many person re-identification algorithms is to divide the pedestrian images or feature map into several blocks, and then the blocks in the same location are used to calculate the special loss functions that measure the differences between different images, to reduce the distance between intra-samples and to increase the distance between inter-samples. However, most of those blocks based deep metric learning methods only measure the difference between different images, but ignore the metrics between different blocks in a single image. In this paper, we propose a novel blocks based method for person re-identification called Overlapping Blocks Model (OBM), in which an innovative strategy of overlapping partition on convolutional features is used to construct multiple overlapping blocks structure and a novel overlapping blocks loss function is utilized to measure the difference between different blocks in a single image, to ensure more blocks can bring more discriminate information and higher performance. We conduct thorough validation experiments on the Market-1501, CUHK03, and DukeMTMC-reID datasets, which demonstrate that our proposed Overlapping Blocks Model can effectively improve the recognition performance of networks by adding the multiple overlapping blocks structure and the overlapping blocks loss.

1. Introduction

Person re-identification (Person re-ID) has attracted more and more researchers in recent years since it is a very important technology in the field of video surveillance. However, in the practical application scenarios, there are various changes in the pedestrian images, such as illumination changes and posture changes, making it difficult to determine whether two pedestrian images belong to the same pedestrian. Therefore, it is important to accurately measure the similarity between two pedestrian images. As a key technique for measuring the similarity between two pedestrian images in person re-identification, metric learning aims to learn a discriminant function to measure the distance between a pair of pedestrian images [1]. In person re-identification, most of classical metric learning methods aim to learn a discriminative metric matrix in a Mahalanobis distance calculation function from the training samples, thus outputting smaller distances for sample pairs belonging to the same person, and outputting larger distances for sample pairs belonging to different pedestrians [2][3][4]. However, it is difficult to calculate the metric matrix in the Mahalanobis distance calculation function due to the SSS (Small Sample Size) problem in person re-identification.

Recently, people have proposed a new research direction called deep metric learning, which effectively combines the deep learning techniques and the metric learning idea to form an end-to-end person re-identification framework. It has been successfully applied in many visual understanding tasks such as face recognition [5][11][14], image classification [6][44], image retrieval [45], visual tracking [7] and person re-identification [24][26][27], etc. According to the difference in the number of inputs, the network frameworks of the deep metric learning methods can be divided into three categories: the traditional single-input classification networks [29][30][34][35], the siamese networks [13-19] and the triplet networks [20-26]. In addition, there are other network frameworks for deep metric learning, such as quadruplet networks [8], N-pair networks [28], etc. The core of those deep metric learning frameworks is that they combine the deep feature representation and the metric learning based loss function in one network, so that they are interdependent and can benefit each other to learn more discriminative feature mapping network parameters. However, most of deep metric learning methods only measure the difference between different images, without measuring the difference in one single image. For example, in person re-identification, considering the body structure in the pedestrian images, it is common to split a pedestrian image or a feature map into multiple blocks vertically and then extract features from each block separately [9][16][40]. The divided blocks of one image are independent and the blocks divided from different input images are used to form the metric loss function. The above-described blocks based deep metric learning methods have a common disadvantage, that is the correlation between adjacent

[*] Corresponding author: Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.

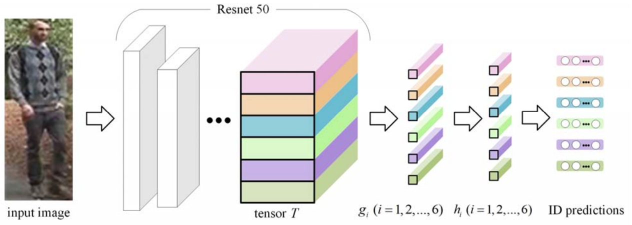


Figure 1: Example network structure of PCB method

blocks from the one image is deficient and the metric between multiple blocks in one image is lacking. For a more specific example, Fig 1 shows the network structure of the Part-based Convolutional Baseline (PCB) method [9], which employs a simple strategy of uniform partition on convolutional features and has been proved as a strong convolutional baseline for person re-identification. The PCB method takes ResNet50 [30] network without hidden fully-connected layers as backbone network. When an image undergoes all the layers inherited from the backbone network, it becomes a 3D tensor T of activations. Afterwards, the 3D tensor T is partitioned into six horizontal blocks and each horizontal block is averaged into a single column vector $g_i (i = 1, 2, ..., 6)$. Then the six column vectors $g_i (i = 1, 2, ..., 6)$ are reduced to column vectors h_i respectively. Finally, each h_i is inputted into a commonly classifier, to predict the identity (ID) of the input. Due to its simple strategy of uniform partition on convolutional features and its efficient network structure based on multiple blocks, the PCB method achieves superior performance in person re-identification. However, in the PCB method, each of the partitioned horizontal blocks in tensor T is independent and there is no correlation between any adjacent blocks, resulting in loss of information of a larger image area.

In order to solve this problem, this paper proposes a novel person re-identification method called Overlapping Blocks Model (OBM), in which an innovative strategy of overlapping partition on convolutional features is used to construct multiple overlapping blocks structure, and then each of overlapping blocks is used to calculate the classification loss. Meanwhile, in the OBM method, a novel overlapping blocks loss function based on a single image is proposed and calculated to combine with the classical classification loss, which can ensure that more blocks with more image information in one image can be used to calculate lower classification loss and obtain higher re-identification performance. Similar to the PCB method mentioned above, our proposed OBM method takes ResNet50 network without hidden fully-connected layers as backbone network and the 3D tensor T is partitioned into six horizontal blocks uniformly. The difference from the PCB method is that the OBM method averaged all two adjacent blocks into a single column vector and averaged all three adjacent blocks into a single column vector, thus the number of column vectors g_i would achieve 15 (6+5+4), and each of the partitioned six horizontal blocks are no longer independent. Meanwhile, in terms of the loss function, the OBM method utilizes a novel overlapping blocks loss to measure the difference between different blocks in one single image. It is consisted of the classification loss differences between overlapping blocks, and it can ensure that the recognition ability of any combination of the adjacent blocks is larger than either of them. We tested these improvements on three public datasets (Market1501 [10], CUHK03 [12] and DukeMTMC-reID [11]) and the experimental results show that the strategy of overlapping partition on convolutional features is effective and the combination of adjacent blocks is a useful complement to single independent block. Moreover, the experimental results also show that the proposed single image based overlapping blocks loss function can effectively improve the re-identification performance in most cases.

The main contributions of this paper are listed as follows:

(1). We propose a novel person re-identification method called Overlapping Blocks Model (OBM), in which an innovative strategy of overlapping partition on convolutional features is used to obtain multiple overlapping blocks structure of feature map.

(2). A novel single image based overlapping blocks loss function is proposed to combine with classification loss to form the loss function of the OBM method, which can ensure that more blocks can bring more discriminate information and higher performance.

(3). We tested our proposed method on three public datasets (Market1501, CUHK03, and DukeMTMC-reID) and the experimental results show that the strategy of overlapping partition on convolutional features in the OBM method is effective and robust, and the single image based overlapping blocks loss can effectively improve the re-identification performance in most cases.

The rest of the paper is organized as follows. Section 2 introduces some related works about deep metric learning methods. Section 3 introduces our proposed OBM method in detail. Experiments on three public datasets are described in Section 4. Finally, the conclusion is draw in Section 5.

2. Related Work

The goal of deep metric learning is to learn a deep metric function to map the original images to a discriminate feature space in which the distance between samples of the same person is smaller than samples of different people. In general, the traditional single-input classification networks, such as AlexNet [34], VGGNet [35], GoogLeNet[29], ResNet [30], can be seen as one of the many network frameworks in the field of deep metric learning. The cross entropy loss calculation function or logistic loss calculation function try to focus the samples of the same person to a unit point and keep the samples of the different people at a certain distance. Sun et al. [9] employed a traditional single-input classification network (ResNet50 [30]) and a simple strategy of uniform partition on convolutional features to reconstruct the network which has been proved as a strong convolutional baseline for person re-identification. However, the correlation between adjacent blocks from the one feature map is deficient and the metric between multiple blocks in one image is lacking.

In addition to the traditional single-input classification networks, there are two main types of network frameworks that are being used in deep metric learning methods: the siamese networks and the triplet networks.

2.1. Siamese Networks

The most common problem when applying single-input classification networks to person re-identification in the early days was lack of training data. In order to deal with this problem, Yi et al. [16] used the siamese networks [39] to determine whether a pair of pedestrian images inputted was the same person. The input pedestrian image pairs can be the same person or different person, so it can greatly alleviate the problem of lack of training data. In [16], the siamese networks consisted of a symmetry structure with two sub-networks, and the two sub-networks were connected by a cosine function. The networks were trained by minimizing a metric learning based contrastive loss function. Varior et al. [36] proposed a gating function to selectively emphasize the common local patterns by comparing the mid-level features across pairs of images, which produces flexible representations for the same image according to the images they are paired with. Chung et al. [37] presented a two stream convolutional neural network where each stream is a siamese network. They also proposed a weighted two stream training objective function which combines the siamese cost of the spatial and temporal streams with the objective of predicting a person's identity.

2.2. Triplet Networks

Compared with the single-input classification networks and the siamese networks, the triplet networks have more choices in the selection of training samples. The key to improving the performance of the triplet networks is to design a suitable metric learning based triplet loss function and select efficient training samples to train the networks. Cheng et al. [38] proposed a multi-channel Parts-Based CNN model, which successfully applied the triplet networks to person re-identification. Shi et al. [24] proposed a novel moderate positive sample mining method to train triplet CNN for person re-identification, dealing with the problem of large variation. In [25], a novel deep metric learning method was proposed that combines the triplet model and the global structure of the embedding space. Moreover, a smart mining procedure was proposed that produces effective training samples for a low computational cost. Hermans et al. [26] showed that, for models trained from scratch as well as pretrained ones, using a variant of the triplet loss to perform end-to-end deep metric learning outperforms most other published methods by a large margin.

2.3. Others

Existing network frameworks of deep metric learning methods based on the contrastive loss or the triplet loss often suffer from slow convergence, in part because they use only one negative sample and do not interact with the other negative samples in each update. To solve this problem, Song et al. [30] described a deep metric learning method for taking full advantage of the training batches in the neural network training by lifting the vector of pairwise distances within the batch to the matrix of pairwise distances. In [31], Sohn presented a scalable novel objective, multi-class N-pair loss, for deep metric learning, which significantly improves upon the triplet loss by pushing away multiple negative examples jointly at each update. In addition, Chen et al. [11] proposed a quadruplet deep network using a margin-based online hard negative mining based on the quadruplet loss.

3. Our approach

In the above-mentioned deep metric learning methods, there are some local feature extraction models based on the block images and some global feature extraction models based on the entire pedestrian image. All of them measure the differences between different images, to reduce the distance between intra-samples and to increase the distance between inter-samples by designing specific loss functions for training networks. However, it's also important to measure the differences in a single image, such as the difference between two different blocks in one image, which may help improve the recognition performance of the networks, and that is the intention of our proposed method. In this paper, we propose a novel deep metric learning framework for person re-identification called Overlapping Blocks Model (OBM) which can utilize the discerning information between overlapping blocks effectively. Comparing to the PCB method [9], the improvements of OBM method lie in the following two points:

(1) After backbone network, OBM averaged each block, each two neighboring blocks and each three adjacent blocks into a single column vector, thus the number of column vector g_i would achieve 15 (6+5+4) and the adjacent blocks are no longer independent.

(2) A novel overlapping blocks loss function was proposed and combined with the classification loss function to train networks which can ensure that more blocks of one image can bring more discriminate information and higher

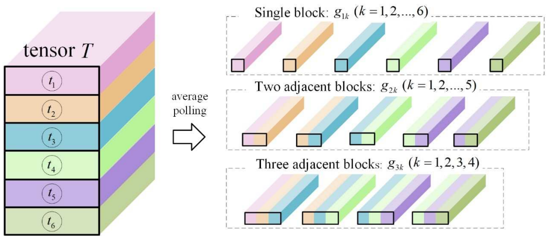

tensor T
t_1
t_2
t_3
t_4
t_5
t_6
average polling
Single block: g_{1k} $(k=1,2,...,6)$
Two adjacent blocks: g_{2k} $(k=1,2,...,5)$
Three adjacent blocks: g_{3k} $(k=1,2,3,4)$

Figure 2: The schematic diagram of the multiple overlapping blocks structure

performance. In this section, we will introduce the OBM method in more detail.

3.1. Multiple Overlapping Blocks Structure

In PCB method, as shown in Fig 1, when an image undergoes all the layers inherited from the backbone network, it becomes a 3D tensor T of activations, and then the tensor T was partitioned into six horizontal blocks equally. After that, each horizontal block was averaged into a single column vector and then each column vector was used as an independent ID classifier. However, each horizontal block in tensor T is independent and there are no correlations between neighboring blocks in PCB method, causing the information between blocks losing.

Thus, we propose a multiple overlapping blocks structure that based on the six-blocks structure in PCB method, we adding some bigger blocks composed of two adjacent blocks or three adjacent blocks. More specifically, as shown in Fig 2, the tensor T was partitioned into six horizontal blocks t_k $(k=1,2,...,6)$ equally . Not only we averaged each horizontal block into a single column vector g_{1k} $(k=1,2,...,6)$, but also we averaged all two adjacent blocks and three adjacent blocks into column vectors g_{2k} $(k=1,2,...,5)$ and g_{3k} $(k=1,2,3,4)$, where the column vector g_{2k} was obtained by averaging two adjacent blocks t_k and t_{k+1} , and the column vector g_{3k} was obtained by averaging three adjacent blocks t_k , t_{k+1} and t_{k+2} .

All the two adjacent blocks and three adjacent blocks are averaged into single column vectors and thus the total number of column vectors would achieve 15(6+5+4). After that, all the column vectors were reduced into lower-dimension column vectors by a 1×1 convolution layer respectively and then each column vector was input into a commonly ID classifier.

3.2. Overlapping Blocks Loss

As mentioned above, we partition the tensor T into six independent horizontal blocks and utilize those six independent blocks to build up fifteen interrelated blocks. Each column vector was reduced into a lower-dimension column vector and then inputted into a commonly ID classifier. The final loss of networks is calculated by adding those fifteen different ID classification loss. In a sense, the ID classification loss can be seen as a metric learning loss that pulling the different images of same person to be closer and pushing the images of different person to be farther. However, those metric learning loss functions only consider the metric loss between different images but ignore the metric loss in the interior of one image, such as the metric loss between the different blocks in one image. For this problem, we propose a novel overlapping blocks loss function which can metric the loss between different blocks of one image and improve the performance effectively.

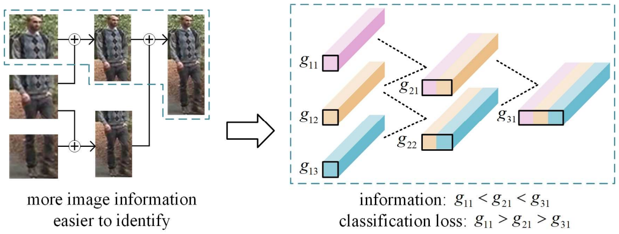


Figure 3: The schematic diagram of the overlapping blocks loss

As we can see in the left side of Fig 3, the first column shows three block images split from a pedestrian image, the second column shows the images composed of adjacent block images in the first column, and the third column shows the complete pedestrian image which can be combined by two images in the second column. In the blue dotted box, with the increase of the number of block images, there are more pixels and more information in the images, thus it is easier to recognize the identity of pedestrians. In a similar way, for the tensor T in the multiple overlapping blocks structure, the information of different blocks is different and any two adjacent blocks must have more information than either of them, since two adjacent blocks have double neurons than single block. Taking this into consideration, we believe that the ID classification loss of any two adjacent blocks must less than the ID classification loss of either of them. For example, as shown in the right of Fig 3, the column vector g_{11} has less information than g_{21} , and g_{21} has less information than g_{31} , thus the classification loss of g_{31} should lower than g_{21} , and the classification loss of g_{21} should lower than g_{11} .

Thus we utilize the difference between single block classification loss $Loss_{g_{11}}$ and two adjacent blocks classification loss $Loss_{g_{21}}$ as the overlapping blocks loss (OBLoss) between g_{11} and g_{21} in the interior of one image, as shown in Formula (1).

$$OBLoss(g_{21}, g_{11}) = [Loss_{g_{21}} - Loss_{g_{11}} + \alpha]_+ \qquad (1)$$

where the operation of $[\cdot]_+$ refers to the hinge function $\max(0,\cdot)$, and the parameter α is a enforced margin between two different loss. In the experiments, the parameter α is set to 0.

Similar to this, we can obtain the overlapping blocks loss

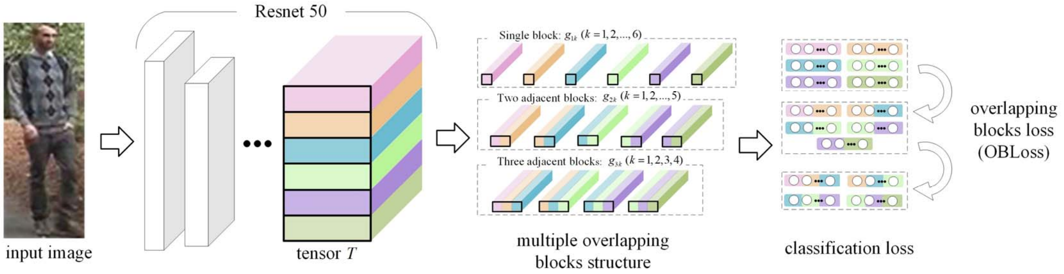

Resnet 50
input image
tensor T
Single block: g_{1k} $(k=1,2,...,6)$
Two adjacent blocks: g_{2k} $(k=1,2,...,5)$
Three adjacent blocks: g_{3k} $(k=1,2,3,4)$
multiple overlapping blocks structure
classification loss
overlapping blocks loss (OBLoss)

Figure 4: The framework of the Overlapping Blocks Model

between two adjacent blocks g_{21} and three adjacent blocks g_{31} such as

$$OBLoss(g_{31}, g_{21}) = [Loss_{g_{31}} - Loss_{g_{21}} + \alpha]_+ \qquad (2)$$

In the train procedure, to boost the speed of convergence, the smaller block loss is not partitioned and the formula (1) and formula (2) can be rewritten as

$$OBLoss(g_{21}, g_{11}) = [Loss_{g_{21}} - ZG(Loss_{g_{11}}) + \alpha]_+ \qquad (3)$$

$$OBLoss(g_{31}, g_{21}) = [Loss_{g_{31}} - ZG(Loss_{g_{21}}) + \alpha]_+ \qquad (4)$$

where $ZG(\cdot)$ represents the zero gradient function, which treats the variable as constant when calculating gradients, stopping the backpropagation in the learning stage.

Combining all the overlapping blocks losses between single blocks, two adjacent blocks and three adjacent blocks, we can obtain the final overlapping blocks loss as shown in Formula (5)

$$\begin{aligned} OBLoss = &\sum_{k=1}^{5}\left[OBLoss(g_{2k}, g_{1k}) + OBLoss(g_{2k}, g_{1(k+1)})\right] \\ &+\sum_{k=1}^{4}\left[OBLoss(g_{3k}, g_{2k}) + OBLoss(g_{3k}, g_{2(k+1)})\right] \end{aligned} \qquad (5)$$

The advantages of overlapping blocks loss are listed as follows:

(1). Overlapping blocks loss ensure that more blocks would bring more useful information for person Re-ID.

(2). The learned deep feature would concentrate on global feature more and focus less on the interference point of local blocks.

3.3. Overlapping Blocks Model

Based on the PCB method, our proposed Overlapping Blocks Model (OBM) combines the multiple overlapping blocks structure (section 3.1) and the overlapping blocks loss (section 3.2) to increase the correlation between different blocks and measure the loss between two overlapping blocks in one image. Fig 4 shows the framework of the Overlapping Blocks Model.

The input of model is a pedestrian image and after the input layer, connect one backbone network. The backbone network can be Google Inception [29], ResNet [30], or any other popular architectures of convolutional networks. Here we choose ResNet50 as the backbone network in Overlapping Blocks Model. When the image undergoes all the layers inherited from the backbone network, it becomes a 3D tensor T of activations. After that, the 3D tensor T was partitioned into six horizontal blocks and each horizontal blocks was averaged into a single column vectors $g_{1k}(k=1,2,...,6)$. Moreover, all the two adjacent blocks were averaged into column vectors $g_{2k}(k=1,2,...,5)$ and all the three adjacent blocks were averaged into column vectors $g_{3k}(k=1,2,3,4)$. Then each column vector was reduce to a lower-dimension column vector by a 1×1 convolution layer respectively. After that, each column vector was input into a commonly ID classifier to calculate the classification loss. Then we can obtain the overlapping blocks loss between all the overlapping blocks according to the formula 5.

The final loss function of Overlapping Blocks Model is composed of the overlapping blocks loss mentioned above and the original classification loss,

$$ModelLoss = W_{cl} * ClassficationLoss + W_{obl} * OBLoss \quad (6)$$

where W_{cl} and W_{obl} are the weighted parameters. In the next section, we would discuss the impact of different weighted parameters through experiments.

Comparing with the PCB method, our proposed Overlapping Blocks Model (OBM) not only consider the independent single blocks but also the overlapping blocks. By adding overlapping blocks loss between single block and adjacent blocks, the relevance between them has been more effectively utilized.

4. Experiments

4.1. Datasets and Settings

4.1.1 Datasets

Three widely used challenging datasets, Market-1501 [10], CUHK03 [12], and DukeMTMC-reID [11], were used for experiments. The Market-1501 dataset is a challenging person re-identification dataset which contains 1501 pedestrians observed under 6 camera viewpoints, 19732 gallery images and 12936 training images detected by DPM [31]. The CUHK03 dataset is another challenging person re-identification dataset which contains 13164 images of 1467 pedestrian captured from two different camera views in a campus environment. The DukeMTMC-reID dataset manifests itself as one of the most challenging re-ID datasets up to now which consists 1404 identities, 16522 training images, 2228 queries, and 17661 gallery images captured by 8 cameras.

4.1.2 Training

The training images are augmented with horizontal flip and normalization. We set batch size to 64 and train the model for 80 epochs with base learning rate initialized at 0.1 and decayed to 0.01 after 40 epochs. The backbone model is pre-trained on ImageNet [33]. With one NVIDIA GeForce GTX 1080Ti GPU and Pytorch as the platform, training a standard Overlapping Blocks Model on Market-1501 (12,936 training images) consumes about 40 and 50 minutes, which is little larger than the standard PCB method.

4.1.3 Evaluation metrics

We used Cumulated Matching Characteristics (CMC) curve to evaluate the performance of person re-identification methods for all datasets in this paper. Because the complexity of the re-identification problem, the top-ranked matching rate was considered. In this paper, 1-ranked, 5-ranked and 10-ranked matching rates were selected for compared. Moreover, we also regard person re-identification as a retrieval task and thus mean Average Precision (mAP) is also used as the evaluation metric.

4.2. Experimental Results and Analysis

4.2.1 The experimental results on Market-1501 dataset

To show the improvement of our proposed OBM method, we conducted thorough validation experiments on market-1501 dataset.

Firstly, to evaluate the necessity of multiple overlapping blocks structure, we construct one-layer zero lap structure, two-layers overlapping blocks structure, three-layers overlapping blocks structure and remove the overlapping blocks loss (set W_{obl} to 0) in the training stage, where one-layer zero lap structure is consisted of six single horizontal blocks $g_{1k}(k=1,2,...,6)$ (same as PCB method), two-layers overlapping blocks structure is consisted of six single horizontal blocks $g_{1k}(k=1,2,...,6)$ and five two adjacent blocks $g_{2k}(k=1,2,...,5)$, three-layers overlapping blocks structure is consisted of $g_{1k}(k=1,2,...,6)$, $g_{2k}(k=1,2,...,5)$ and $g_{3k}(k=1,2,3,4)$. Table 1 shows the comparison of three different structures. Two-layers overlapping blocks structure achieves better performance than one-layer zero lap structure. Meanwhile, three-layers overlapping blocks structure is better than two-layers overlapping blocks structure. The gap between the results in Table 1 indicate the multiple overlapping blocks structure are indispensable.

Next, we conducted experiments with different parameters (W_{obl}) to show the impact of the overlapping blocks loss. Three-layers overlapping blocks structure are used in all experiments and the weight parameters W_{cl} is set to 1. Table 2 shows the comparison of different parameter values (0, 0.1, 0.2, 0.5, 1, 2). As we can see in the Table 2, the results of the parameter values (0.1, 0.2, 0.5, and 1) are better than the parameter values (0, 2) and the parameter value (0.2) achieves the highest performance, which indicates the overlapping blocks loss is indispensable in the training stage and can enable neural networks to learn more useful and discriminate features from adjacent blocks. However, when the parameter value is set too high (2), the recognition performance will decrease, which indicates that we should not pay too much attention to the loss of adjacent blocks and ignore the loss between different images. Therefore, setting the parameter W_{obl} between 0.1 and 1 is a good choice.

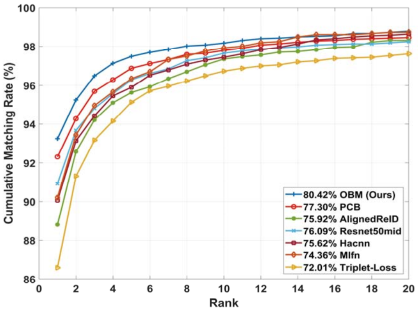
TABLE I
THE COMPARISON OF THREE DIFFERENT STRUCTURES ON MARKET-1501 DATASET.

Number of Layers	Rank=1	Rank=5	Rank=10	mAP
One	92.31	96.88	97.80	77.30
Two	92.43	97.00	97.92	79.14
Three	**92.61**	**97.05**	**98.06**	**79.50**

TABLE II
THE COMPARISON OF SIX DIFFERENT PARAMETER ON MARKET-1501 DATASET.

Parameters	Rank=1	Rank=5	Rank=10	mAP
0	92.61	97.05	98.06	79.50
0.1	93.08	97.42	97.98	80.18
0.2	**93.24**	**97.49**	98.16	**80.42**
0.5	92.70	96.91	97.95	80.24
1	92.96	97.39	**98.25**	79.67
2	92.58	97.27	98.22	79.45

Compare to the PCB method, our proposed OBM method focus more on adjacent blocks and utilizing a novel overlapping blocks loss to measure the difference between the single block and adjacent blocks, learning more useful and discriminate features from adjacent blocks. The Table 3 and Fig 5 shows the experimental results of the OBM method, the PCB method and other state-of-the-art methods on market-1501 dataset. As shown in Table 3, the OBM method can consistently improve the PCB method, where the gain of rank-1 accuracy and mAP can achieve 0.93%, 3.12% respectively. Meanwhile, we use reranking method [32] to further boost the performance especially mAP. When the OBM is combined with the reranking method, the rank-1 accuracy can achieve 94.06% and the mAP can achieve 91.21%, which are better than “PCB+Reranking”. Moreover, to shown the robustness of multiple overlapping blocks structure and overlapping blocks loss, we replace the backbone network from ResNet50 to VGG 16. The experimental results are shown in the last two lines in the Table 3. Our proposed method is also effective when using the VGG16 network as the backbone network.



x

Figure 5: The CMC curves and mAP on the Market-1501 dataset.

TABLE III
THE RECOGNITION RESULTS OF OUR MODEL AND OTHER THE STATE-OF-THE-ART METHODS ON MARKET-1501 DATASET.

Method	Rank=1	Rank=5	Rank=10	mAP
OBM (Ours)	**93.24**	**97.49**	**98.16**	**80.42**
PCB [9]	92.31	96.88	97.80	77.30
AlignedReID [40]	88.81	95.64	97.36	75.92
Hacnn [41]	90.93	96.35	97.77	75.62
Mlfn [42]	90.18	95.93	97.44	74.36
Resnet50mid [43]	90.23	96.37	97.94	76.09
Triplet-Loss [26]	86.61	95.04	97.73	72.01
OBM+Reranking	**94.06**	**96.70**	**97.80**	**91.21**
PCB+Reranking	93.57	96.62	97.25	89.42
OBM (VGG16)	**91.12**	**96.44**	**97.77**	**75.22**
PCB (VGG16)	90.44	96.29	97.39	74.08

4.2.2 *The experimental results on CUHK03 dataset*

Similar to the experiments on market-1501 dataset, we conducted thorough validation experiments on CUHK03 dataset.

Firstly, we construct one-layer zero lap structure, three-layers overlapping blocks structure without the overlapping blocks loss (set W_{obl} to 0), and three-layers overlapping blocks structure with the overlapping blocks loss (set W_{obl} to 0.1) in the training stage to evaluate the necessity of multiple overlapping blocks structure and overlapping blocks loss. Table 4 shows the comparison results. We can see the three-layers overlapping blocks structure is better than the one-layer zero lap structure, which indicates the importance of multiple overlapping blocks structure. Meanwhile, three-layers overlapping blocks structure with the overlapping blocks loss achieve higher rank-1 accuracy and mAP than three-layers overlapping blocks structure without the overlapping blocks loss. The results verify the view that the overlapping blocks loss is beneficial to performance.

We also compare the OBM method to the PCB method and other state-of-the-art methods on CUHK03 dataset and the experimental results are shown in Table 5 and Fig 6. As we can see in the Table 5, the OBM method is 3.87% and 5.15% better than PCB method on the rank-1 accuracy and mAP respectively. Compared with other state-of-the-art methods, the OBM method also has advantages in performance. Moreover, combined with the reranking method [32], the OBM method is still better than PCB method with significantly improved performance.

TABLE IV
THE COMPARISON OF THREE DIFFERENT STRUCTURES ON CUHK03 DATASET.

Method	Rank=1	Rank=5	Rank=10	mAP
OneLayer	61.13	79.29	85.79	57.04
ThreeLayers	63.71	81.33	86.34	59.63
ThreeLayers+OBLoss	**65.00**	**82.14**	**87.29**	**62.19**

TABLE V
THE RECOGNITION RESULTS OF OUR MODEL AND OTHER THE STATE-OF-THE-ART METHODS ON CUHK03 DATASET.

Method	Rank=1	Rank=5	Rank=10	mAP
OBM (Ours)	**65.00**	**82.14**	87.29	62.19
PCB [9]	61.13	79.29	85.79	57.04
AlignedReID [40]	60.36	78.64	85.07	57.34
Hacnn [41]	51.13	79.12	**87.62**	**62.67**
Mlfn [42]	50.12	76.26	84.06	61.83
Resnet50mid [43]	51.05	76.58	83.93	62.06
Triplet-Loss [26]	53.79	73.79	82.36	51.73
OBM+Reranking	**73.93**	**83.50**	**88.79**	**76.13**
PCB+Reranking	72.00	81.86	87.71	74.43

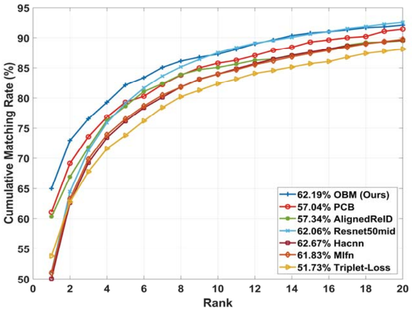


Figure 6: The CMC curves and mAP on the CUHK03 dataset.

4.2.3 *The experimental results on DukeMTMC-reID dataset*

Our proposed OBM method has a significant improvement on the market-1501 and CUHK03 datasets. In order to verify the robustness of the OBM method, we also conducted thorough validation experiments on the DukeMTMC-reID dataset, which is the dataset with the largest number of views in the three tested datasets.

Firstly, we compare the performance of baseline structure (one-layer zero lap structure), three-layers overlapping blocks structure without the overlapping blocks loss (set W_{obl} to 0), and three-layers overlapping blocks structure with the overlapping blocks loss (set W_{obl} to 0.2). The experimental results are shown in Table 6.

From the Table 6 we can see the same results that the performance of three-layers overlapping blocks structure is better than baseline structure (one-layer zero lap structure) and the overlapping blocks loss has a great help to improve performance. The results indicate that the multiple overlapping blocks structure and the overlapping blocks loss are robust that also showing effective improvement on DukeMTMC-reID dataset.

TABLE VI
THE COMPARISON OF THREE DIFFERENT STRUCTURES ON DUKEMTMC-REID DATASET.

Method	Rank=1	Rank=5	Rank=10	mAP
OneLayer	84.47	91.57	93.76	69.94
ThreeLayers	85.05	92.18	94.21	71.04
ThreeLayers+OBLoss	**85.32**	**92.46**	**94.38**	**71.73**

TABLE VII
THE RECOGNITION RESULTS OF OUR MODEL AND OTHER THE STATE-OF-THE-ART METHODS ON DUKEMTMC-REID DATASET.

Method	Rank=1	Rank=5	Rank=10	mAP
OBM (Ours)	**85.32**	**92.46**	**94.38**	**71.73**
PCB [9]	84.47	91.57	93.76	69.94
AlignedReID [40]	81.55	90.44	93.13	66.71
Hacnn [41]	80.13	89.62	92.14	63.21
Mlfn [42]	81.14	90.25	92.56	63.22
Resnet50mid [43]	81.67	89.98	93.04	64.08
Triplet-Loss [26]	78.19	89.41	91.83	62.57
OBM+Reranking	**89.09**	**93.36**	**94.61**	**85.39**
PCB+Reranking	88.31	93.18	94.38	84.12

Finally, we compare our proposed OBM method to the PCB method and other state-of-the-art methods on the DukeMTMC-reID dataset. The experimental results are shown in Table 7 and Fig 7. As we can see, the OBM method can consistently improve the PCB method, where the gain of rank-1 accuracy and mAP achieve 0.85%, 1.79% respectively. Moreover, the performance of OBM method is better than other compared state-of-the-art methods. When the OBM method is combined with the reranking method in [32], the rank-1 accuracy can achieves 89.09% and the mAP can achieves 85.39%, which are better than the “PCB+Reranking” method.

5. Conclusion

In this paper, we proposed a novel deep metric learning approach for person re-identification called Overlapping Blocks Model (OBM), which is consist of the multiple overlapping blocks structure and the overlapping blocks loss. The proposed multiple overlapping blocks structure utilizes the multiple overlapping blocks of feature map to calculate the classification loss, which can increase the correlation between different independent blocks in one image. The overlapping blocks loss can measure the differences between two overlapping blocks in one single image and ensure more blocks can bring more discriminate information. The experiments on three publicly datasets, market-1501, CUHK03, and DukeMTMC-reID, show the effectiveness and the robustness of OBM method. It would be interesting to see that the multiple overlapping blocks structure and the overlapping blocks loss can be applied to other block based methods.

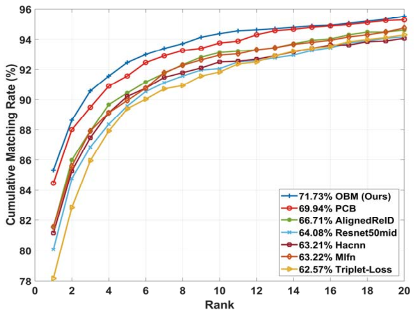


Figure 7: The CMC curves and mAP on the DukeMTMC-reID dataset.

Acknowledges

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work was supported by the China National Natural Science Foundation under Grant No. 61673299, 61203247, 61573259, 61573255.

References

[1] Kulis B. Metric learning: A survey[J]. Foundations and Trends® in Machine Learning, 2013, 5(4): 287-364.

[2] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2288-2295.

[3] Liao S, Hu Y, Zhu X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2197-2206.

[4] Xiong F, Gou M, Camps O, et al. Person re-identification using kernel-based metric learning methods[C]. European conference on computer vision. Springer, Cham, 2014: 1-16.

[5] Lu J, Hu J, Tan Y P. Discriminative deep metric learning for face and kinship verification[J]. IEEE Transactions on Image Processing, 2017, 26(9): 4269-4282.

[6] Cui Y, Zhou F, Lin Y, et al. Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1153-1162.

[7] Hu J, Lu J, Tan Y P. Deep metric learning for visual tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(11): 2056-2068.

[8] Chen W, Chen X, Zhang J, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C]. The

IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, 2(8).

[9] Sun Y, Zheng L, Yang Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling[J]. arXiv preprint arXiv:1711.09349, 2017.

[10] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]. Proceedings of the IEEE International Conference on Computer Vision. 2015: 1116-1124.

[11] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]. European Conference on Computer Vision. Springer, Cham, 2016: 17-35.

[12] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 152-159.

[13] Cai X, Wang C, Xiao B, et al. Deep nonlinear metric learning with independent subspace analysis for face verification[C]. Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012: 749-752.

[14] Hu J, Lu J, Tan Y P. Discriminative deep metric learning for face verification in the wild[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1875-1882.

[15] Lu J, Hu J, Tan Y P. Discriminative deep metric learning for face and kinship verification[J]. IEEE Transactions on Image Processing, 2017, 26(9): 4269-4282.

[16] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]. Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 34-39.

[17] Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]. Advances in neural information processing systems. 2014: 1988-1996.

[18] Hu J, Lu J, Tan Y P. Deep transfer metric learning[C]. Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. IEEE, 2015: 325-333.

[19] Lu J, Wang G, Deng W, et al. Multi-manifold deep metric learning for image set classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1137-1145.

[20] Wang J, Song Y, Leung T, et al. Learning fine-grained image similarity with deep ranking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1386-1393.

[21] Hoffer E, Ailon N. Deep metric learning using triplet network[C]. International Workshop on Similarity-Based Pattern Recognition. Springer, Cham, 2015: 84-92.

[22] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 815-823.

[23] Cui Y, Zhou F, Lin Y, et al. Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1153-1162.

[24] Shi H, Yang Y, Zhu X, et al. Embedding deep metric for person re-identification: A study against large variations[C]. European Conference on Computer Vision. Springer, Cham, 2016: 732-748.

[25] Harwood B, Carneiro G, Reid I, et al. Smart mining for deep metric learning[J]. arXiv preprint arXiv:1704.01285, 2017.

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

[27] Song H O, Xiang Y, Jegelka S, et al. Deep metric learning via lifted structured feature embedding[C]. Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. IEEE, 2016: 4004-4012.

[28] Sohn K. Improved deep metric learning with multi-class n-pair loss objective[C]. Advances in Neural Information Processing Systems. 2016: 1857-1865.

[29] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]. AAAI. 2017, 4: 12.

[30] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[31] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[32] Zhong Z, Zheng L, Cao D, et al. Re-ranking person re-identification with k-reciprocal encoding[C]. Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017: 3652-3661.

[33] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009: 248-255.

[34] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. 2012: 1097-1105.

[35] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[36] Varior R R, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human re-identification[C]. European Conference on Computer Vision. Springer, Cham, 2016: 791-808.

[37] Chung D, Tahboub K, Delp E J. A two stream siamese convolutional neural network for person re-identification[C]. The IEEE international conference on computer vision (ICCV). 2017.

[38] Cheng D, Gong Y, Zhou S, et al. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1335-1344.

[39] Bromley J, Guyon I, LeCun Y, et al. Signature verification using a" siamese" time delay neural network[C]. Advances in neural information processing systems. 1994: 737-744.

[40] Zhang X, Luo H, Fan X, et al. Alignedreid: Surpassing human-level performance in person re-identification[J]. arXiv preprint arXiv:1711.08184, 2017.

[41] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2285-2294.

[42] Chang X, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2109-2118.

[43] Yu Q, Chang X, Song Y Z, et al. The Devil is in the Middle: Exploiting Mid-level Representations for Cross-Domain Instance Matching[J]. arXiv preprint arXiv:1711.08106, 2017.

[44] Shen F, Shen C, Zhou X, et al. Face image classification by pooling raw features[J]. Pattern Recognition. 2016, 54: 94-103.

[45] Li Z, Tang J. Weakly supervised deep metric learning for community-contributed image retrieval[J]. IEEE Transactions on Multimedia. 2015, 17(11): 1989-1999.