CUTMIX DUAL BRANCH NETWORK FOR PERSON RE-IDENTIFICATION

Zengming Tang1,2 and Jun Huang1,2*

1Shanghai Advanced Research Institute, Chinese Academy of Sciences
2University of Chinese Academy of Sciences
{tangzengming2019, huangj}@sari.ac.cn

ABSTRACT
The performance of deep learning methods for person re-identification (Re-ID) is influenced by overfitting problem. To improve the generalization ability, most methods pay attention to generating and utilizing new samples. However, generated samples focus on object occlusion but neglect pedestrian occlusion, while triplet loss fails to preserve all identity information on samples with multiple pedestrians. In this paper, we propose the CutMix dual branch network (CDBN) to relieve these problems. CutMix is introduced to this model, and it is responsible for generating new samples with pedestrian occlusion. CutMix is verified complementary to image erasing strategy for Re-ID. Besides, a generalized triplet loss called CutMix triplet loss (CTP) is employed on samples augmented by CutMix with consideration of identity information from multiple pedestrians, making CDBN robust to two kinds of occlusions. Extensive experiments on two benchmarks demonstrate the strengths of CDBN, which is superior to state-of-the-art methods.

Index Terms— Person re-identification, data augmentation, triplet loss, deep learning

1. INTRODUCTION

Person re-identification (Re-ID) is a fundamental task in computer vision field, aiming to identify the same person across non-overlapping camera views. Deep learning based Re-ID methods have dominated this community, and push the performance of Re-ID to a new level. However, many large-scale Re-ID datasets are still small compared to datasets in image classification, face recognition tasks for the small number of images for each class. As mentioned in [1], each identity involves 17.2 and 23.5 images in Market-1501 [2] and DukeMTMC-reID [3] datasets. As a result, overfitting problem is more severe in Re-ID task. Although many data augmentation methods have been proposed to improve the generalization ability of convolutional neural networks (CNNs), there is still room in generating and utilizing new samples.

* Corresponding author.

This paper was supported by National Key R&D Program of China (2020YFC1523202, 2017YFE0192600) and NSFC (61972007).

Fig. 1. Room in generating and utilizing new samples. Top: the generalization ability is influenced by object occlusion and pedestrian occlusion. For object occlusion, person of interest is occluded by tailing and bicycle. For pedestrian occlusion, person of interest is occluded by another person with red border. Person of interest and person with red border contain similar visual parts, which confuses Re-ID models in feature extraction. Bottom: triplet loss reduces the intra-class variations and enlarges the inter-class variations for anchors with one and two pedestrians, but ignores the identity information of person with red border.

Room in generating new samples. The generalization ability of CNNs is influenced by occlusion [4]. We observe that occlusion can be categorized into two types: 1) Object occlusion which means the person of interest is occluded by objects; 2) Pedestrian occlusion which means the person of interest is occluded by other pedestrians. Image erasing strategies [4, 5] randomly select and mask out one rectangle region for each input image, and generate new images with various occlusion rates. In this process, image erasing strategies improve generalization ability of CNNs and robustness to object occlusion. However, they can not handle pedestrian occlusion problem. Pedestrian occlusion not only hides informative regions like object occlusion, but also introduces heavy noise from other pedestrians, as illustrated in Fig. 1.

Room in utilizing new samples. Metric learning is one of the fundamental steps for learning discriminative features with CNNs. Triplet loss is frequently applied to recent Re-ID
works for its great power in optimizing the similarity among samples. Most previous works focus on mining scheme [6, 7, 8] and improving the gradient back-propagation [9, 10, 11]. They perform well on samples with single pedestrian, but can not preserve all identity information on samples with multiple pedestrians, as shown in Fig. 1.

To overcome the above limitations, we propose a novel CutMix dual branch network (CDBN) in this paper. Our CDBN consists of two branches namely standard branch (SB) and CutMix branch (CB). Both branches are designed for Re-ID task, but they differ in data augmentation strategy. SB coupled with an additional image erasing strategy while CB performs Re-ID task on input images with an additional CutMix strategy [12]. During training, we replace the rectangle portion for each input image with a region from another image in CB. Various pedestrian occlusion status can be produced to enlarge the diversity of dataset. Further more, we introduce a generalized triplet loss called CutMix triplet loss (CTP). In CTP, anchor is generated with CutMix in CB while positive and negative are sampled in SB for preserving original identity information. In this manner, CutMix is combined with image erasing strategy. Specifically, anchor is synthesized by mixing two positives corresponding to two labels, and intra-class distance is the linear interpolation of the distances between anchor and two positives. In a nutshell, we achieve training a robust CNN for two kinds of occlusion problems and performing metric learning by CTP with consideration of multiple pedestrians in our proposed CDBN simultaneously.

In summary, our main contributions are outlined as follows:

- We introduce a novel CutMix dual branch network (CDBN) to perform Re-ID task with individual data augmentation strategy in standard branch (SB) and CutMix branch (CB).

- We incorporate CutMix in CDBN for Re-ID task, which turns out efficient in boosting the performance. Besides, CutMix operates in conjunction with image erasing strategy gains further improvements.

- We propose a new CutMix triplet loss (CTP) with the consideration of multiple pedestrians from samples, and it is superior to the classic triplet loss.

- We improve the robustness of CDBN against object occlusion and pedestrian occlusion by employing CTP on samples augmented with CutMix.

2. RELATED WORKS

In this section, we only discuss recent works in terms of synthetic data based Re-ID and metric learning based Re-ID.

2.1. Synthetic data based Re-ID

Overfitting problem is severe in Re-ID as mentioned in Sec. 1. We divide synthetic data based methods into two types, i.e., data augmentation based and GAN based. In the first type, image erasing strategy, e.g., Random Erasing [4], Cutout [5], randomly select and erase one rectangle region for each input image. In the second type, GAN is applied in [13] to generate more training data for the first time. To relieve pose variations in Re-ID, PT [14], PN-GAN [15] and FD-GAN [16] synthesize realistic person images with pose guidance.

CDBN generates new training samples by CutMix, which belongs to the first type. However, CutMix has not been verified its effectiveness in triplet loss and Re-ID task in previous works. On the contrary, CDBN with CutMix and CTP enhances the robustness against occlusion and boosts the performance of Re-ID.

2.2. Metric learning based Re-ID

Metric learning focuses on learning a similarity, which reduces intra-class variations and enlarges inter-class variations. Metric learning approaches can be categorized into two groups, i.e., classification learning based and pairwise learning based. In the first group, a siamese network [1] is trained with cross-entropy loss and a proposed verification loss simultaneously. Specifically, verification loss is a binary cross-entropy loss indicating the image pair belongs to the same person or not. In the second group, a variant of triplet loss [6] with mining the hardest positive and hardest negative is applied to Re-ID. HAP2S [7] with a soft hard-mining scheme has the properties of accuracy, robustness, flexibility and generality. Apart from mining scheme, most previous works [9, 10, 11] focus on improving the gradient back-propagation.

The above works learn a similarity among samples by taking only one main pedestrian into account. In contrast, our proposed CTP is introduced with the consideration of identity information from multiple pedestrians.

3. CUTMIX DUAL BRANCH NETWORK (CDBN)

In this section, we first introduce the overall structure of CutMix dual branch network (CDBN). Then we discuss the details about the standard branch (SB) and the CutMix branch (CB). Finally, we describe the inference phase for the CDBN.

3.1. Overall Structure

As is shown in Fig. 2, the CDBN takes images augmented with base data augmentation strategy, i.e., random horizontal flip, normalization, as input. The mini-batch of input images is denoted as \( X = \{ x_i \}_{i=1}^N \) with labels \( \{ y_i \}_{i=1}^N \), where
$N$ is the batchsize and $x_i \in \mathbb{R}^{W \times H \times C}$. The CDBN consists of two branches, namely standard branch (SB) and Cutmix branch (CB). In SB, the backbone network for feature extraction is ResNet-50 [17]. We modify the backbone network slightly by removing the last spatial down-sampling operation to enrich the granularity of pedestrian representation. With the global average pooling (GAP) after the backbone network, global feature representation with 2048-dim is generated. The CB keeps the same structure setting and shares all the weights. GAP is short for global average pooling.

3.2. Standard Branch (SB)

The input data for the SB is augmented with base data augmentation strategy and image erasing augmentation strategy, i.e., Cutout, while input for CB is augmented with an additional CutMix strategy. The SB is trained with triplet loss [6] and CutMix strategy. The SB is trained with triplet loss [6] and Cutout, while input for CB is augmented with an additional image erasing augmentation strategy, i.e., Cutout.

Specifically, input for SB is augmented with an additional CutMix strategy, and it is trained with our proposed CutMix triplet loss (CTP). Specifically, these two branches share all the weights.

The input data for the SB is augmented with base data augmentation strategy, and it is trained with triplet loss and classification loss. 2) The CB takes input data with an additional CutMix strategy, and it is trained with our proposed CutMix triplet loss (CTP). Specifically, these two branches share all the weights.

Fig. 2. The overall architecture of our CDBN. The CDBN contains two branches: 1) The SB takes input data with an additional image erasing augmentation strategy, and it is trained with triplet loss and classification loss. 2) The CB takes input data with an additional CutMix strategy, and it is trained with our proposed CutMix triplet loss (CTP). Specifically, these two branches share all the weights. GAP is short for global average pooling.

loss on $f_s$ can be formulated as:

$$L_{cls}^s = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp((W y_i)^T f_s)}{\sum_{j=1}^{C} \exp((W_j)^T f_s)}$$

where $C$ is the number of identities, $W_j$ represents the weight matrix in the fully connected (FC) layer whose input is $f_s$ for $i$-th identity and $y_i$ is the label for $i$-th input image.

3.3. CutMix Branch (CB)

The input data for the CB is augmented with base data augmentation strategy and CutMix strategy. The aim of CutMix is to create a mixed training image $(\tilde{x}, \tilde{y})$ given two training images $(x_m, y_m)$ and $(x_n, y_n)$. Mixing operation is defined as following:

$$\tilde{x} = M \odot x_m + (1 - M) \odot x_n$$
$$\tilde{y} = \lambda y_m + (1 - \lambda) y_n$$

where $\odot$ is point-wise multiplication, $\lambda \in [0, 1]$ which is sampled from the beta distribution Beta($\alpha, \alpha$), controls the area ratio of patches from $x_m$ to generated image $\tilde{x}$. $\alpha$ is a constant value equals 1 for our experiments. Binary mask $M \in \{0, 1\}^{W \times H}$ indicates which pixel is preserved and removed from two images while binary mask $1$ is filled with ones. Given $\lambda$, the binary mask $M$ is constructed by filling with $0$ within the sampled bounding box $B = (b_x, b_y, b_w, b_h)$ and $1$ without $B$. Coordinates of bounding box $B$ are uniformly sampled as following:

$$b_x \sim U(0, W), \quad b_y \sim U(0, H), \quad b_w = W \sqrt{1 - \lambda}, \quad b_h = H \sqrt{1 - \lambda}$$

letting the preserved area ratio $1 - \frac{b_\lambda b_\alpha}{WTP} = \lambda$. The augmented input data is fed into the CB to acquire the 512-dim global feature $f_c$. The CB is trained with our proposed CTP.
Given two original training samples \( x_m, x_n \) and a mixed sample \( \bar{x} \), we sample two triplet units \((x_m, x_{m+}, x_{m-})\) and \((x_n, x_{n+}, x_{n-})\), and a negative sample \( x_{mn-} \) whose label satisfies \( y_{mn-} \neq y_m \) and \( y_{mn-} \neq y_n \). In this fashion, \( \bar{x} \) can be regarded as the anchor. Then the CTP is defined as:

\[
L^c_{ctri} = \frac{1}{N} \sum_{i=1}^{N} [m + d_{a,p} - d_{a,n}]_+ \tag{5}
\]

where \( m \) is the margin parameter, \( d_{a,p} \) and \( d_{a,n} \) are the intra-class and inter-class distance. As mentioned in [6], mining of hard triplets is a significant part of learning with triplet loss, mining the hardest positive and negative in a triplet is a good choice. To this end, intra-class and inter-class distance are formulated as:

\[
d_{a,p} = \lambda \max_{y_{m+} = y_m} d(f^m_{mix}, f^{m+}_{s}) + (1 - \lambda) \max_{y_{n+} = y_n} d(f^m_{mix}, f^{n+}_{s}) \tag{6}
\]

where \( f^m_{mix} \) denotes the feature \( f_c \) extracted from mix sample \( \bar{x} \), \( f^{m+}_{s} \) is the feature \( f_s \) extracted from \( x_{m+} \), and so forth. The complete formula of CTP includes Eq. 5 and Eq. 6. CTP aims to minimize the linear interpolation of distances between mixed sample \( \bar{x} \) and two positive \( x_{m+}, x_{n+} \), while maximizing the distance between mixed sample \( \bar{x} \) and negative \( x_{mn-} \). In this manner, CTP optimizes the network by taking original and mixed data into consideration.

To highlight the difference between triplet loss and our proposed CTP, they are visualized in Fig. 3. Triplet loss minimizes the distance between anchor and hardest positive while maximizes the distance between anchor and hardest negative. CTP aims to learn a similarity between anchor, two corresponding hardest positive and hardest negative. CTP captures a more complex interaction among more samples (4 samples) compared to the original triplet loss (3 samples). Specifically, triplet loss can be seen the special case of CTP if \( x_m = x_n \).

Finally, the overall loss function can be written as:

\[
L = \alpha (L^c_{ctri} + L^c_{tri}) + L_{cls} \tag{7}
\]

where \( \alpha \) is a trade-off parameter. In our experiments, \( \alpha \) is set to 0.5.

### 3.4. Inference phase

During inference, the input images are resized to \( 288 \times 144 \) and normalized before fed into the SB without image erasing strategy. The 512-dim global feature \( f_s \) is extracted as the final pedestrian representation.

### 4. EXPERIMENTS

Experiments are conducted on two Re-ID benchmarks, Market-1501 [2] and DukeMTMC-reID [3] datasets. First, we make comparisons between our proposed CDBN and state-of-the-art methods, showing the effectiveness of our method. Then a number of ablation experiments on Market-1501 are carried out to verify that each component boosts the performance of CDBN. Last, we analyse the robustness to object occlusion and pedestrian occlusion of our method.

#### 4.1. Implementation Details

The CDBN is initialized with the weights of ResNet-50 [17] pretrained on ImageNet. During training, the input images are resized to \( 288 \times 144 \). A mini-batch is randomly sampled with 8 identities, and 4 images of each identity. The margin \( m \) in TP , CTP all equals 0.3 to make a fair comparison. The model is trained with 120 epochs. The CDBN is optimized using stochastic gradient descent (SGD) with weight decay \( 5 \times 10^{-4} \). The initial learning rate is \( 3 \times 10^{-2} \), with linear warm-up in the first 10 epochs, and decays by \( 10 \times \) after 30, 70 epochs.

#### 4.2. Datasets and Evaluation Metrics

**Market-1501** contains 32,668 images from 1501 pedestrians captured by six cameras. The whole dataset is split into training set with 12,936 images of 751 persons, and testing set with 19,732 gallery images and 3,368 query images of 750 persons. Specifically, 2,793 images are regarded as distractors. **DukeMTMC-reID** is a subset of DukeMTMC for Re-ID specifically. This dataset includes 36,411 images of 1,812 pedestrians captured by eight cameras. However, only 1,404 persons appear in more than two cameras while 408 persons are distractors. Training set is selected with 16,522 images of 702 persons. The remaining 702 persons are divided into 17,661 gallery images and 2,228 query images.

Cumulative Matching Characteristics (CMC) at Rank-1 and the mean Average Precision (mAP) are applied for Re-ID evaluation metrics. All experiments are evaluated under single-query mode.
4.3. Comparison with State-of-the-art Methods

In this section, we compare our CDBN with 18 recently proposed state-of-the-art methods in Tab. 1. These compared methods are separated into 3 groups: synthetic samples based (S), part feature based methods (P), global feature based (G).

Table 1. Comparison of state-of-the-art methods on two Re-ID benchmarks. Red and Blue indicate the best and the second best results respectively. The bold numbers refer to the best performance in each group. "-" means not available.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Market-1501 Rank-1</th>
<th>mAP</th>
<th>DukeMTMC-reID Rank-1</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSRO [13] (ICCV17)</td>
<td>84.0</td>
<td>66.1</td>
<td>67.7</td>
<td>47.1</td>
</tr>
<tr>
<td>PT [14] (CVPR18)</td>
<td>87.7</td>
<td>68.9</td>
<td>78.5</td>
<td>56.9</td>
</tr>
<tr>
<td>Camstyle [18] (CVPR18)</td>
<td>88.1</td>
<td>68.7</td>
<td>75.3</td>
<td>53.3</td>
</tr>
<tr>
<td>PN-GAN [15] (ECCV18)</td>
<td>89.4</td>
<td>72.6</td>
<td>73.6</td>
<td>53.2</td>
</tr>
<tr>
<td>FD-GAN [16] (NeurIPS18)</td>
<td>90.5</td>
<td>77.7</td>
<td>80.0</td>
<td>64.5</td>
</tr>
<tr>
<td>DG-Net [19] (CVPR19)</td>
<td>94.8</td>
<td>86.0</td>
<td>86.8</td>
<td>74.8</td>
</tr>
<tr>
<td>P2S [10] (CVPR17)</td>
<td>70.7</td>
<td>44.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PBC-RP [20] (ECCV18)</td>
<td>93.8</td>
<td>81.2</td>
<td>83.3</td>
<td>69.2</td>
</tr>
<tr>
<td>HPM [21] (AAAI19)</td>
<td>94.2</td>
<td>82.7</td>
<td>86.6</td>
<td>74.3</td>
</tr>
<tr>
<td>HORiE [22] (CVPR20)</td>
<td>94.2</td>
<td>84.9</td>
<td>86.9</td>
<td>75.6</td>
</tr>
<tr>
<td>BDB [33] (ICCV19)</td>
<td>94.5</td>
<td>85.0</td>
<td>88.7</td>
<td>75.8</td>
</tr>
<tr>
<td>IDE [1] (CVPR18)</td>
<td>79.5</td>
<td>59.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SVDNet [24] (CVPR18)</td>
<td>82.3</td>
<td>62.1</td>
<td>76.7</td>
<td>56.8</td>
</tr>
<tr>
<td>TriNet [6] (arXiv17)</td>
<td>84.9</td>
<td>69.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HAP2S_E [7] (CVPR19)</td>
<td>84.2</td>
<td>69.8</td>
<td>76.1</td>
<td>59.6</td>
</tr>
<tr>
<td>HAP2S_P [7] (CVPR19)</td>
<td>84.6</td>
<td>69.4</td>
<td>75.9</td>
<td>60.6</td>
</tr>
<tr>
<td>MVP Loss [25] (ICCV19)</td>
<td>91.4</td>
<td>80.5</td>
<td>83.4</td>
<td>70.0</td>
</tr>
<tr>
<td>BoT [26] (CVPRW19)</td>
<td>94.5</td>
<td>85.9</td>
<td>86.9</td>
<td>76.4</td>
</tr>
<tr>
<td>CDBN (Ours)</td>
<td>95.2</td>
<td>86.6</td>
<td>86.9</td>
<td>75.2</td>
</tr>
</tbody>
</table>

Comparison with Synthetic Samples based Methods (S methods). Our proposed CDBN achieves Rank-1/mAP=95.2%/86.6% on Market-1501 and Rank-1/mAP=86.9%/75.2% on DukeMTMC-reID respectively. Although S methods utilize generated images, CDBN surpasses all of them. Besides, CDBN generates synthetic images by CutMix augmentation strategy, which is more lightweight than S methods incorporating generative modules.

Comparison with Part Feature based Methods (P methods). P methods focus on fine-grained details, which have been verified efficient in boosting Re-ID performance. However, they aggregate features of multi-branch or multi-scale. CDBN achieves competitive results only using global feature from SB.

Comparison with Global Feature based Methods (G methods). CDBN can be included in G methods for SB sharing all weights with CB. The BoT combines many training tricks not applied by CDBN, e.g., center loss, large batch size. With simpler training tricks, CDBN outperforms BoT in Rank-1/mAP by 0.7%/0.7% on Market-1501.

Qualitative Results. Some retrieval results on Market-1501 dataset are illustrated in Fig. 4. Even though the query image is occluded by object and another pedestrian, CDBN can still make correct decision.

4.4. Ablation Studies

Each component of CDBN is evaluated incrementally on Market-1501 in Tab. 2.

Table 2. Ablation studies of CDBN on Market-1501 dataset. SB and CB: standard branch and CutMix branch; TP and CLS: triplet loss and classification loss; CTP: CutMix triplet loss; RE and Cut: Random Erasing and Cutout.

<table>
<thead>
<tr>
<th>Model</th>
<th>Market-1501</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank-1</td>
</tr>
<tr>
<td>1</td>
<td>SB (TP)</td>
</tr>
<tr>
<td>2</td>
<td>CB (CTP)</td>
</tr>
<tr>
<td>3</td>
<td>CB (CTP) + RE</td>
</tr>
<tr>
<td>4</td>
<td>CB (CTP) + Cut</td>
</tr>
<tr>
<td>5</td>
<td>SB (TP) + CB (CTP)</td>
</tr>
<tr>
<td>6</td>
<td>SB (TP) + CB (CTP) + RE</td>
</tr>
<tr>
<td>7</td>
<td>SB (TP) + CB (CTP) + Cut</td>
</tr>
<tr>
<td>8</td>
<td>SB (TP/CLS) + CB (CTP) + RE</td>
</tr>
<tr>
<td>9</td>
<td>SB (TP/CLS) + CB (CTP) + Cut (CDBN)</td>
</tr>
</tbody>
</table>

Benefit of CTP. Triplet loss and CTP are compared in Tab. 2 with model 1, 2, 5. When CTP applied alone, CTP outperforms triplet loss in Rank-1/mAP by 0.6%/3.3%. CutMix augmentation strategy benefits metric learning process with variation of triplet loss for the first time. Furthermore, TP and CTP are complementary. The combination of TP and CTP (model 5) exceeds TP (model 1) and CTP (model 2) by a large margin.

Relationship with Image Erasing Strategy. Although CutMix and image erasing strategy contain the similar operation i.e., removing a patch from the sample, the comparison of model 2, 3, 4 demonstrates that image erasing strategy can benefit CB. The combination of CutMix and image erasing strategy with SB leads to further gains by the comparison of 3, 4, 6, 7. To the best of our knowledge, this is the first time that CutMix strategy and image erasing strategy boost the performance simultaneously.

Benefit of Classification Loss. The effectiveness of classification loss is shown in the comparison of model 6, 7, 8, 9 in Tab. 2. Our CDBN is trained by using CTP, triplet loss.
Fig. 5. Robustness experiments on the Market-1501. SB (TP) (O) and CB (CTP) (O) are short for model SB (TP) and CB (CTP) under object occlusion. SB (TP) (P) and CB (CTP) (P) are short for model SB (TP) and CB (CTP) under pedestrian occlusion.

and classification loss. We can see that incorporating classification loss on SB improves the performance.

4.5. Robustness to Occlusion

For occlusion experiments, we divide occlusion into two types: object occlusion and pedestrian occlusion. Specifically, we add these two occlusions to the query set of Market-1501 during testing. As for generating samples occluded by objects, we randomly select a bounding box whose coordinates $B = (b_x, b_y, b_w, b_h)$ satisfy Eq. 4 with occlusion rate $1 - \lambda$, then fill the region $B$ with random values. As for generating samples occluded by pedestrians, we replace the region $B$ with a patch of the same coordinates from another image in a mini-batch. Results are shown in Fig. 5. Firstly, the retrieval accuracies i.e., Rank-1 and mAP, of all models drop quickly with the increase of occlusion rate, which indicates occlusion is a key factor in degrading the performance of Re-ID. Secondly, the retrieval accuracies under pedestrian occlusion is much lower than under object occlusion, which shows pedestrian occlusion is a more pressing problem than object occlusion. Finally, CTP models drop slowly compared with TP models, which demonstrates that CTP improves the robustness against two kinds of occlusion.

5. CONCLUSIONS

In this paper, we propose the CDBN, a siamese network-like multi-branch network. CutMix is introduced to increase the diversity of training data. Specifically, CutMix incorporates with our proposed CTP enhances the generalization ability of CDBN for Re-ID task, and improves the robustness against two kinds of occlusions. Extensive ablation studies verify that each component of CDBN is beneficial for Re-ID task. Besides, CDBN achieves state-of-the-art performance only with global features. Re-ID is always viewed as an image retrieval task, we will apply our method to general image retrieval tasks to verify its generality in the future.

6. REFERENCES