Global Contrastive Person Re-identification

Shengyu Pei1, Xiaoping Fan1,2,*

1School of Automation, Central South University, No. 22 Shaoshan South Road, Changsha Hunan 410075 P.R. China
2School of Information Technology and Management, Hunan University of Finance and Economics, No. 139 Fenlin 2nd Road, Changsha Hunan 410205 P.R. China
*xpfan@csu.edu.cn

Abstract: Solving the problem of pedestrians being occluded by objects is extremely challenging. Using part-level features to describe pedestrian images can provide fine-grained information. However, only paying attention to the local features of body will lack global pedestrian information. And the network consumes time and memory. To solve these problems, we propose a new person re-identification network. The network uses a global contrastive module to obtain the features of pedestrians. Through effective use of the pedestrian's global features, as well as the pedestrian's personal information and global contrastive information, the pedestrian can be found in the object occlusion to provide a reliable feature embedding. Our model is tested on Market1501, DukeMTMC-reID, CUHK03 and MSMT17 datasets. The experimental results show that our method is effective in occluded person re-identification.

Keywords: Occlusion Problem, Global Feature, Global ContrastiveModule, Person Re-Identification

1. Introduction
Given a pedestrian that needs to be matched, determine that this person appears at different times on different cameras, which is the goal of person re-identification (ReID). It has been widely studied as a cross-camera specific pedestrian retrieval problem and is widely used in video surveillance, security, and smart cities. Because pedestrians have different attributes, such as gender, clothing, hair, etc. Besides, pedestrians are shot in different locations and different scenes, which are affected by lighting, cameras, occlusion, and background. In the past ten years, with the remarkable development of convolutional neural networks, people can obtain robust pedestrian representations from these changes. Therefore, scholars have proposed many effective person re-identification methods [1]–[5]. However, these methods mostly focus on the feature extraction of the overall image and ignore the occluded image. Solving the problem of pedestrians being occluded by objects is more practical and more challenging.
Researchers improve the accuracy of pedestrian recognition by combining multiple local features, pose estimation, and segmentation semantic information, but these operations consume memory and space. This paper uses a new global contrastive module to improve the performance of ReID. The specific contributions are as follows:

1) Propose a new global contrastive module, which can focus on global information, special feature information, and global contrastive information.
2) The network model proposed in this paper does not add much extra consumption.
3) By testing four classic person re-identification data and comparing the state-of-the-art method, the method proposed in this paper has achieved constructive results.

2. Related work
Aiming at the pedestrian misalignment problem caused by pose changes, a pose estimator based on CNN is used to locate body joints [6]–[8]. Using part-level features of body can provide more fine-grained information for pedestrian retrieval [4], [9], [10]. To deal with the problem of misalignment and locate the different local features, [11] proposed a new full attention module (FAB), which can create attention information in channel and space to mine useful features for person re-identification.

[12], [13] By using Generative adversarial networks (GAN) to enhance the data, thereby improving the recognition performance. Many scholars have solved the task of person re-identification by obtaining the global features of pedestrians. For example, Luo et al. [3] explored a simple and effective baseline framework in CVPR 2019, and collected many effective training techniques for person re-identification tasks. Ye et al. [14] explored a powerful AGW framework, and at the same time gave a new evaluation metrics, and used it in cross-domain problems. This article will also draw on many effective training techniques. Aiming at the problem of pedestrian occlusion, a new global contrastive module is proposed to mine the global feature information and at the same time, it can also obtain the personal information of pedestrians.

3. My approach

3.1. Global contrastive module

Hao Luo et al. proposed the BNNeck module in CVPR in 2019. BNNeck is very simple and efficient. It just adds a batch normalization (BN) after the feature. This feature refers to the global feature ob-
tained by the network model before the FC layer of the classifier. As for why it is so efficient, the author described in the article as it constrains the classification hyperplane to pass through the origin of coordinates. But in fact, this paper only paid attention to the global feature information, while ignoring the special feature information. For occluded pedestrian images, global feature information will mislead the classifier, resulting in the inability to correctly match the pedestrians in the gallery.

Therefore, this paper proposes a new global contrastive module, as shown in Fig. 2. In Fig. 2, the input image is preprocessed first, and then passed through the ResNet50 framework fine-tune, and finally through three channels to obtain the features and classification results. These three channels obtain feature information with different meanings respectively. For 1-th channel, “A” represents the adaptive mean pooling operation on the features after fine-tuning, as shown in formula 1. What it obtains is the general feature information of the image. For 2-th channels, “M” represents the adaptive maximum pooling operation for the features after fine-tuning, as shown in formula 2. It can pay attention to the individual feature of the image. For occluded pedestrian images, the channel can avoid occlusions, and pay special attention to the pedestrian parts that are not occluded. For 3-th channels, “-” means A-M, as shown in formula 3, this is the global contractive module. This module faces pedestrian images with serious occlusion. This global contractive module can compare the overall situation without losing special attention and find pedestrian features that are more suitable for matching. For each channel, we use BNNeck operation. In other words, before the FC layer of the classifier, batch normalization (BN) operations are performed uniformly. This can also ensure that the constrained classification hyperplane passes through the origin of the coordinate, and the distribution will not be disturbed, so that the data distribution follows a certain rule. These operations help the network to converge, resulting in better performance.

\[
\begin{align*}
    f_c &= \frac{1}{|X_c|} \sum_{x \in X_c} x \\
    f_c &= \max_{x \in X_c} x \\
    f_c &= \max_{x \in X_c} x - \frac{1}{|X_c|} \sum_{x \in X_c} x
\end{align*}
\]

3.2. Training skills

To enable the network to get better training, we added some training techniques to the baseline framework. The experimental results also prove that these training techniques can help person re-identification models to obtain better performance.

1) Learning rate (LR): The learning rate has a great influence on the performance of the person re-identification. The original network model was trained with a large and constant learning rate. Here, we refer to the literature [3], [5] and choose to use a warmup strategy to guide the network to improve performance.

2) Random erasing: In the occluded person re-identification, especially when the pedestrian image is occluded by other objects. In order to solve the occlusion problem and improve the generalization ability of the ReID model, Zhong et al. [15] proposed a new data enhancement method called random erasure enhancement.

3) Strides(1, 1): The higher spatial resolution will always enrich the granularity of the elements. Sun et al. [4] deleted the last spatial down-sampling operation in the backbone network to increase the size of the feature map. Therefore, the strides of the last convolutional layer of our ResNet50 is set to (1, 1). This operation only adds a very light computational cost and does not involve additional training parameters.

4. Experiments

4.1. Datasets and settings
Datasets: This article uses 4 data sets to evaluate the algorithm. For example, Market-1501[16], DukeMTMC-reID[17], [18], CUKH03[19] and MSMT17[20].

4.2. Implementation Details
The image undergoes preprocessing operations such as fixed size, random horizontal flip, filling, random cropping, and random erasing. The fixed size is set to (256,128). The probability of random horizontal flip and random erasing are both set to 0.5. The padding parameter is set to 10.

The features obtained by the network are back-propagated through triplet loss and cross entropy loss. The margin of triplet loss is set to 0.3. In sampling, \( p = 16 \) means the number of different persons each batch, and \( k = 4 \) means the number of images per person. Therefore, a total of 64 images per batch. We use Adam optimizer to train the network. Using warmup strategy, the initial learning rate is 0.00035, and the total number of iteration steps is 120. We use the same configuration to test on 4 classic data sets.

4.3. Comparison with the state of the art

| Method         | Market1501 R1 | Market1501 mAP | DukeMTMC-ReID R1 | DukeMTMC-ReID mAP |
|----------------|--------------|----------------|------------------|-------------------|
| PIE[6]         | 87.7         | 69.0           | 79.8             | 62.0              |
| SPReID[7]      | 92.5         | 81.3           | 84.4             | 71.0              |
| MaskReID[8]    | 90.0         | 75.3           | 78.8             | 61.9              |
| AlignedReID[9] | 90.6         | 77.7           | 81.2             | 67.4              |
| SCPNet[21]     | 91.2         | 75.2           | 80.3             | 62.6              |
| PCB[4]         | 93.8         | 81.6           | 83.3             | 69.2              |
| Pyramid[10]    | 95.7         | 88.2           | 89.0             | 79.0              |
| Mancs[11]      | 93.1         | 82.3           | 84.9             | 71.8              |
| Camstyle[12]   | 88.1         | 68.7           | 75.3             | 53.5              |
| PN-GAN[13]     | 89.4         | 72.6           | 73.6             | 53.2              |
| SVDNet[1]      | 82.3         | 62.1           | 76.7             | 56.8              |
| AWTL[2]        | 89.5         | 75.7           | 79.8             | 63.4              |
| BagTricks[3]   | 94.5         | 85.9           | 86.4             | 76.4              |
| AGW (Full)[14] | 95.1         | 87.8           | 89.0             | 79.6              |
| Ours (ResNet50)| 94.7         | 87.2           | 87.7             | 76.7              |
| Ours (ResNet101-IBN-A) | **95.3** | **89.3** | **90.4** | **80.6** |

Tab.1 Comparison of state-of-the-arts methods (Market1501, DukeMTMC-ReID)

We compare with state-of-the-art. These networks use different strategies and are representative. For example, AlignedReID, SCPNet, Pyramid and PCB are based on component features. For the Market1501 data set, the results obtained in this paper are better than PCB by 1.5% R1 and 7.7% mAP.

| Method         | CUHK03 R1 | CUHK03 mAP | MSMT17 R1 | MSMT17 mAP |
|----------------|----------|-----------|----------|-----------|
| BagTricks[3]   | 58.0     | 56.6      | 63.4     | 45.1      |
| AGW (Full)[14] | 63.6     | 62.0      | 68.3     | 49.3      |
| Circle loss[22]| -        | -         | 76.9     | 52.1      |
| Ours (ResNet50)| **72.8** | **69.7** | **77.7** | **52.5** |

Tab.2 Comparison of state-of-the-arts methods (CUHK03, MSMT17)
For the CUHK03 and MSMT17 data sets, compared to BagTricks, AGW and Circle loss, the algorithm in this paper achieves the best results on both R1 and mAP metrics.

5. Conclusion

In this article, a new global contrastive network is proposed for the extremely challenging occlusion problem in person re-identification. The network provides 3 channels with different functions. The network model can not only pay attention to the global feature information, but also the personalized information of pedestrians and the global contrastive information. This method provides an effective and feasible idea for solving the problem of pedestrian occlusion. In order to verify the effectiveness of the method proposed in this paper, we have done a lot of experiments on 4 data sets. In the future, we will use more visualization methods to present the working principles behind the network.

References

[1] Sun Y, Zheng L, Deng W and Wang S 2017 SVDNet for Pedestrian Retrieval Proceedings of the IEEE International Conference on Computer Vision pp 3800–8

[2] Ristani E and Tomasi C 2018 Features for Multi-Target Multi-Camera Tracking and Re-Identification Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 6036–46

[3] Luo H, Gu Y, Liao X, Lai S and Jiang W 2019 Bag of Tricks and A Strong Baseline for Deep Person Re-identification arXiv:1903.07071

[4] Sun Y, Zheng L, Yang Y, Tian Q and Wang S 2018 Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline) Computer Vision – ECCV 2018 Lecture Notes in Computer Science ed V Ferrari, M Hebert, C Sminchisescu and Y Weiss pp 501–18

[5] Fan X, Jiang W, Luo H and Fei M 2019 SphereReID: Deep hypersphere manifold embedding for person re-identification J. Vis. Commun. Image Represent 60 pp51–58

[6] Zheng L, Huang Y, Lu H and Yang Y 2019 Pose-Invariant Embedding for Deep Person Re-identification IEEE Trans. Image Process 28 pp4500–4509

[7] Sarfraz M S, Schumann A, Eberle A and Stiefelhagen R 2018 A Pose-Sensitive Embedding for Person Re-identification with Expanded Cross Neighborhood Re-ranking IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 420–9

[8] Qi L, Huo J, Wang L, Shi Y and Gao Y 2019 MaskReID: A Mask Based Deep Ranking Neural Network for Person Re-identification arXiv:1804.03864

[9] Luo H, Jiang W, Zhang X, Fan X, Qian J and Zhang C 2019 AlignedReID++: Dynamically matching local information for person re-identification Pattern Recognition 94 pp53–61

[10] Zheng F, Deng C, Sun X, Jiang X, Guo X, Yu Z, Huang F and Ji R 2019 Pyramidal Person Re-Identification via Multi-Loss Dynamic Training arXiv:1810.12193

[11] Wang C, Zhang Q, Huang C, Liu W and Wang X 2018 Manes: A Multi-task Attentional Network with Curriculum Sampling for Person Re-Identification Computer Vision – ECCV 2018 Lecture Notes in Computer Science ed V Ferrari, M Hebert, C Sminchisescu and Y Weiss (Cham: Springer International Publishing) pp 384–400

[12] Zhong Z, Zheng L, Zheng Z, Li S and Yang Y 2019 CamStyle: A Novel Data Augmentation Method for Person Re-Identification IEEE Trans. Image Process 28 pp1176–1190

[13] Qian X, Fu Y, Xiang T, Wang W, Qiu J, Wu Y, Jiang Y-G and Xue X 2018 Pose-Normalized Image Generation for Person Re-identification Proceedings of the European Conference on Computer Vision (ECCV) pp 650–67

[14] Ye M, Shen J, Lin G, Xiang T, Shao L and Hoi S C H 2020 Deep Learning for Person Re-identification: A Survey and Outlook arXiv:2001.04193

[15] Zhong Z, Zheng L, Kang G, Li S and Yang Y 2017 Random Erasing Data Augmentation arXiv:1708.04896

[16] Zheng L, Shen L, Tian L, Wang S, Wang J and Tian Q 2015 Scalable Person Re-identification:
A Benchmark 2015 IEEE International Conference on Computer Vision pp 1116–24

[17] Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C 2016 Performance Measures and a Data Set for Multi-target, Multi-camera Tracking Computer Vision – ECCV Workshops Lecture Notes in Computer Science 9914 pp 17–35

[18] Zheng Z, Zheng L and Yang Y 2017 Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro arXiv:1701.07717

[19] Li W, Zhao R, Xiao T and Wang X 2014 DeepReID: Deep Filter Pairing Neural Network for Person Re-identification IEEE Conference on Computer Vision and Pattern Recognition pp 152–159

[20] Wei L, Zhang S, Gao W and Tian Q 2018 Person Transfer GAN to Bridge Domain Gap for Person Re-Identification IEEE Conference on Computer Vision and Pattern Recognition pp 79–88

[21] Fan X, Luo H, Zhang X, He L, Zhang C and Jiang W 2019 SCPNet: Spatial-Channel Parallelism Network for Joint Holistic and Partial Person Re-identification Computer Vision 11362 pp 19–34

[22] Sun Y, Cheng C, Zhang Y, Zhang C, Zheng L, Wang Z and Wei Y 2020 Circle Loss: A Unified Perspective of Pair Similarity Optimization CoRR abs/2002.10857