

# Better Person Re-identification Using ResNet Model and Re-ranking Strategy

**Nengchao Pan\***

Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China

\*pannengchao@126.com

**Abstract.** In this paper, we propose a novel person re-identification algorithm based on deep learning. To start with, deep features are extracted from images by ResNet model. Then a set of related images are selected under the low similarity constraint. In the end, we re-rank the retrieval result with related images. Our algorithm has three advantages. First, the transfer learning method reduces the cost needed to train the model compared with the others. Second, we employ the genetic algorithm under adaptive sparsity constraint to detect related images, which reduces the computation complexity. Third, the detected related images can be applied to promote the initial retrieval result. Experiments on Market1501 have demonstrated the effectiveness of our algorithm and our algorithm is robust against variations in background, pose, illumination and viewpoint.

## 1. Introduction

Person re-identification(ReID) aims to search an expected pedestrian's image(probe image) in the gallery for images containing the same person and those images were taken under non-overlapping environment. Person re-identification is an important task in the surveillance system and has various applications and re-identification was developed rapidly during the last decades for its great potential.

Although the person ReID problem has been studied for years, there are still several challenges. Pedestrians' images vary largely in background, pose, illuminance and viewpoint because these images were taken under non-overlapped environment. To tackle these problems, traditional methods are mainly studied in two aspects: designing a robust feature descriptor[1] and learning the distance metrics. Some representative methods, like symmetry-driven features[2], probabilistic colour histogram[3] and PSD constrained asymmetric metric learning[4], have promoted the experimental results. Deep learning is adopted for it can reduce the impact of human intervention on models. Ahmed[5] proposed a binary classification network to determine whether two images containing the same pedestrian. Xiao[6] mixed all databases to pre-train the model and finetune on the target database. State-of-the-art mainly studies the distance between probe image and each image in the gallery to obtain an initial ranking list. A good practice consists of adding a re-ranking step can improve the searching result.

We propose a novel re-ranking algorithm using ResNet model and related images. Our main contributions are summarized as follows:1)We finetune the pre-trained ResNet model on target person ReID database, and extract features from images with the trained model to obtain the initial ranking list.2)Use the genetic algorithm under adaptive sparsity constraint to detect related images in which each image is relevant to each other. Aggregate the original distance and the distance between related



images and gallery images to generate the final distance. Experimental results on Market1501[7] show the effectiveness of our algorithm.

## 2. Our Approach

### 2.1. Problem Definition

Given a probe image and the gallery set with  $N$  images  $G = \{g_i | i = 1, 2, \dots, N\}$ , we obtain an initial ranking list  $\bar{L}(p, G) = \{g_i^0 | i = 1, 2, \dots, N\}$ , simplified as  $\bar{L}$ , by metric the distance between two images  $p$  and  $g_i$ .  $d(p, g_i) = (x_p - x_{g_i})^T M (x_p - x_{g_i})$ , where  $x_p$  and  $x_{g_i}$  respectively represents the feature of probe image  $p$  and gallery  $g_i$ ,  $M$  is a positive semidefinite matrix. Our goal is to detect a set of images, called related images, in  $\bar{L}$ , and use the related images to re-rank  $\bar{L}$  with the expectation of improving the results that more positive samples rank higher in the list than previous.

### 2.2. Initial Ranking List Generation

To ensure the features can represent the images and reduce the cost of computing, we finetune the pre-trained neural network model on target person ReID database. Transfer learning is very common in practical use, with the benefits of saving time and calculation. We choose the ResNet model to extract image features. ResNet model solves the problem of vanishing gradient and exploding gradient by shortcut connection. After finetuning the ResNet model on the train data of the target dataset, we take the outputs of the model as the features extracted from the images. The distance between probe image  $p$  and gallery image  $g_i$  is defined as follows:

$$d(x_p, x_{g_i}) = x_p \odot x_p + x_{g_i} \odot x_{g_i} - 2 * x_p \odot x_{g_i} \quad (1)$$

The initial ranking list  $\bar{L}$  is acquired by sorting the gallery images according to their distance to the probe image in ascending order. Let the images of  $\bar{L}$  be  $\bar{G} = \{\bar{g}_i | i = 1, 2, \dots, N\}$ .

### 2.3. Detect Related Image

In this subsection, we describe how to detect the related images in the initial ranking list. As mentioned earlier[8], those images relevant to probe image should have high similarity to each other because the expected individual's images share the same visual pattern in different cameras. It means that the total distance of a positive sample to all other images in the initial ranking list should be smaller than the total distance of a negative sample because a negative sample does not share a similar visual pattern with other images in the ranking list. We name the set of images in the initial ranking list which is highly similar to each other related images. To detect related images from top  $M$  items in the initial sequence, we name a matrix  $K \in R^{M \times M}$  in which  $K_{ij}$  is the distance between  $x_{\bar{g}_i}$  and  $x_{\bar{g}_j}$ , let  $s = \{s_i | i = 1, 2, \dots, M\}$  be the sum of gallery image  $x_{\bar{g}_i}$  to every image in the top  $M$  items.

$$s_i = \sum_{j=1}^M d(\bar{g}_i, \bar{g}_j) \quad (2)$$

Let  $c \in \{0, 1\}^M$  be the index of our detection of related images in initial ranking list,  $c_i = 1$  means image  $\bar{g}_i$  been chosen as related image, and vice versa for  $c_i = 0$ . If the total distance of an expected image to others is low, the expected image is possible to be similar to other images and could possibly be a relevant image. In order to find the set of images which are most similar to each other, we aim to maximize the result of  $\|s - Kc\|$ . We add a sparsity constraint to  $c$  to ensure the precision of selected images.

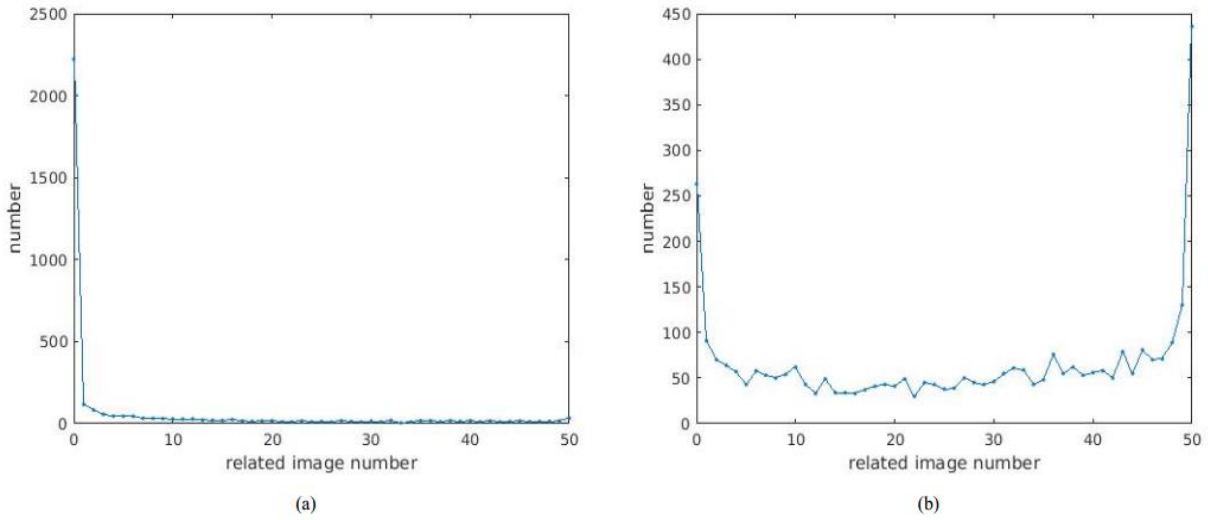
$$\max \|s - Kc\|_p^q + a \|c\|_2 \quad (3)$$

$$s.t. c \in \{0, 1\}^M$$

Where  $\alpha$  is the sparsity constraint on  $\mathbf{c}$ . To simplify the problem, we solve the dual problem of the original problem, as Equation 3.

$$\begin{aligned} \max \quad & \|\mathbf{s} - \mathbf{K}\bar{\mathbf{c}}\|_p^q + b\|\bar{\mathbf{c}}\|_2 \\ \text{s.t.} \quad & \bar{\mathbf{c}} \in \{0,1\}^M \end{aligned} \quad (4)$$

Where  $\bar{\mathbf{c}}$  is the index the unrelated images,  $\mathbf{c}$  equals  $\bar{\mathbf{c}}$  flipped.  $b$  is the constraint to ensure that most of the unrelated images are selected.



**Figure 1.** The impact of  $b$  on detection of related images. We detect related images in top 50 items. The horizontal axis represents the number of related image of an initial ranking list, the longitudinal axis represents the number of initial ranking lists. (a) is the statistical result of  $b=20$ , (b) is the result of  $b = 30$

Both the original problem and the dual problem are NP-hard, and exhaustive search is infeasible to solve as the demand for computation is hard to meet. Instead, we obtain the suboptimal solution by optimizing. It is not doable to detect related images of different initial ranking list under the same  $b$  because it is not guaranteed that all the related images detected from the different initial ranking list meet the sparsity constraint, as shown in Figure 1. We propose a related image detected algorithm based on the adaptive sparse parameter. Set the upper and lower bounds on the sparse parameter for every initial ranking list, the result of related images detection of a particular initial ranking list is iteratively calculated under current  $b$  by binary search the appropriate value of  $b$ , until the number of related images conforms to the sparse constraint. We use the genetic algorithm to detect related images because  $\mathbf{c}$  is not differentiable and the number of related images depends on the value of  $b$ , the process of detecting related images from the expected initial ranking list at different weight may need to be iterated many times until the final result is obtained. The genetic algorithm has a small computational cost and can solve non-differential problems.

#### 2.4. Final Distance

In this subsection, we focus on how to re-rank the initial ranking list with the related images. The importance of original distance cannot be neglected, so we jointly aggregate the original distance and the distance between related images and the gallery images, the final distance  $d(g_i)$  is defined as equation 5.

$$d(g_i) = d(\mathbf{x}_p, \mathbf{x}_{g_i}) + vd(g_i, \mathbf{c}) \quad (5)$$

$$d(g_i, \mathbf{c}) = \sum_{j=1}^N c_j * d(g_i, \bar{L}_j)$$

Where  $v \in [0,1]$  denote the weight factor of the distance between related images and the gallery images,  $d(g_i, \mathbf{c})$  is the total distance of a gallery image to all the related images. The final distance of a relevant image should be lower than the total distance of an uncorrelated image. Only the original distance is considered when  $v=0$ .

### 3. Experiment Result

Experiments are conducted on the Market1501[7]. Market1501 is the largest person re-identification dataset at the moment, consisting of 3368 images with 750 identities for querying, 12936 images with 751 identities for training and 19732 images with 750 identities for testing. We compare the two indices of rank-i and mAP with state-of-the-art algorithms to evaluate the proposed ResNet model and related image based person ReID algorithm. Rank-i measures the proportion of the queries that find the positive samples in top-i items in all the queries. Higher rank-i means the method can find at least one positive sample in top-i items faster. The mAP is the mean average precision, which reflects global performance.

**Table 1.** The impact of a, detect 2~3 related images in top-5 items

	Rank1	Rank5	mAP
NFST[9]	55.4	-	-
PersonNet[11]	48.2	-	-
S-CNN[12]	65.9	-	-
BoW-best[7]	44.4	63.9	-
WARCA[10]	45.16	-	-
CAN[13]	48.24	-	24.43
ResNet	73.9	88.24	51.52
ResNet+re-rank	74.91	87.35	53.44

We demonstrate the effectiveness of our approach by comparing with the following algorithm. Traditional methods, like NFST[9] is a representative metric learning algorithm which projects a high dimension vector to a low dimension vector by a projection matrix. BoW[7] is a method of encoding feature which aims at reducing intra class distance and increase inter class distance. WARCA[10] is a Mahalanobis distance based metric algorithm. Deep learning has promoted the effect. Both PersonNet[11] and S-CNN[12] are binary classification model. CAN[13] proposed a new soft attention based model. The results are exhibited in Table 1.

It is clear from the results that our algorithm outperforms other existing person ReID algorithms. The initial ranking list generated from ResNet model achieves a better result than the traditional methods and the deep learning methods mentioned above. And the re-ranking stage based on related images improves the ranking result. The main reason is that the feature extracted from images by

ResNet model is robust against variations of background, pose, illuminance and viewpoint, and the features can represent the images well. At the same time, the images detected from the initial ranking list are similar to each other, the related images are more likely to be relevant to the query image, and in the ideal case, the final ranking result is the total distance of the expected individual in multi-cameras, which can mitigate the difference between same class samples caused by taken under non-overlapping environment.

#### 4. Conclusion

We present a novel person ReID method using ResNet model and related images. We use ResNet model to generate the initial ranking list, then detect related images from the list and aggregate the original distance and the distance between gallery images and related images to obtain the final ranking list. The experiments conducted on Market1501 show our algorithm achieves a better ranking result compared with some state-of-the-art algorithms.

#### Reference

- [1] Liu C, Chen C L and Chen C L 2012 *International Conference on Computer Vision* Person re-identification: what features are important? pp.391–401.
- [2] Farenzena M, Bazzani L and Perina A 2010 *Computer Vision and Pattern Recognition* Person re-identification by symmetry-driven accumulation of local features pp. 2360–2367.
- [3] D’Angelo A and Dugelay J L 2011 *Proceedings of SPIE - The International Society for Optical Engineering* People re-identification in camera networks based on probabilistic color histograms pp.2155–2157.
- [4] Liao S and Li S Z 2015 *IEEE International Conference on Computer Vision* Efficient PSD Constrained Asymmetric Metric Learning for Person ReIdentification pp.3685–3693.
- [5] Ahmed E, Jones M and Marks T K 2015 *Computer Vision and Pattern Recognition* An improved deep learning architecture for person re-identification pp.3908–3916.
- [6] Xiao T, Li H and Ouyang W 2016 *Computer Vision and Pattern Recognition* Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification pp.1249–1258.
- [7] Zheng, Liang, Shen Liyue and Tian Lu, Wang Shengjin, Wang Jingdong and Tian Qi 2015 *IEEE International Conference on Computer Vision* Scalable Person Re-identification: A Benchmark pp.1116–1124.
- [8] Morioka N and Wang J 2011 *ACM International Conference on Multimedia* Robust visual reranking via sparsity and ranking constraints pp.533–542, 2011.
- [9] Zhang Li, Xiang Tao and Gong Shaogang 2016 *Computer Vision and Pattern Recognition* Learning a Discriminative Null Space for Person Re-identification pp.1239–1248.
- [10] Jose, Cijo and Fleuret François 2016 *European Conference on Computer Vision* Scalable Metric Learning via Weighted Approximate Rank Component Analysis pp.875–890.
- [11] Wu Lin, Shen Chunhua, Hengel and Anton Van Den 2016 *CoRP* PersonNet: Person Re-identification with Deep Convolutional Neural Networks vol.1601.07255.
- [12] Varior, Rahul Rama, Haloi, Mrinal and Wang Gang 2016 *European Conference on Computer Vision* Gated Siamese Convolutional Neural Network Architecture for Human Re-identification pp.791–808.
- [13] Liu Hao, Feng Jiashi, Qi Meibin, Jiang Jianguo and Yan Shuicheng 2016 *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* End-to-End Comparative Attention Networks for Person Re-identification vol. PP, number.99, pp.1–1.