Unsupervised Person Re-ID in Surveillance Feed Using Re-ranking

Mohit Kumar Singh(✉), Vijay Laxmi, and Neeta Nain

M.N.I.T Jaipur, Jaipur, India
{2017rcp9005,vlaxmi,nnain.cse}@mnit.ac.in

Abstract. With the increase of video feeds from the network of surveillance cameras and available sophisticated detection and bounding box techniques we have seen a jump in the use of deep learning models in the past years. These Deep Neural Network models are Supervised, in the sense that they require large labeled data samples. This hunger of more and more labeled data can be removed by moving on to Unsupervised learning. Despite of significant progress, very less attention is paid to unsupervised techniques. In our approach we tries to improve the accuracy of the Unsupervised Neural Network model by using re-ranking. K-mean algorithm is used to obtain the initial cluster as the samples are unlabeled initially and k-reciprocal nearest neighbor method is used to re-rank the output to remove false matches. Experiments are performed on DUKE and CUHK01 datasets.

Keywords: Supervised learning · Unsupervised learning · K nearest neighbour · CNN

1 Introduction

Person re-identification (re-ID) [1–6,36] has become very popular in the recent years due to its research significance and is primarily used in application areas such as video surveillance system, path detection, Activity-based human recognition and many other. Surveillance is the most crucial one as it includes object detection, object tracking, and classification of moving object or group of objects. The effectiveness of surveillance systems is judged by first how accurately (in shape and size) the system can detect an object or any suspicious behavior, and second by finding how reliable the system is with the change in environmental conditions such as lighting and background conditions.

Person re-id aims in identifying an object of interest in another viewing area or in the same when the object re-enters the frame after leaving it first. It is the need of today's world to identify terrorist attacks or other suspicious activities in the place of mass gatherings. Recently with the increase of video feeds available from the network of cameras installed in several smart cities for surveillance, we have significantly noticed the use of Deep Learning Systems for re-ID.

S. Nandi et al. (Eds.): ISEA-ISAP 2018, CCIS 939, pp. 179–192, 2019.
https://doi.org/10.1007/978-981-13-7561-3_14

Person re-ID is primarily divided into Image-based and Video-based systems [1]. And the complete process of re-ID in a video-based system can be divided into Person Identification or Detection, Person Tracking and Person re-identification. Person identification and tracking in the video scene is a challenging task. We human masters this trick even if we losses the track of a person by checking for their appearance. This process becomes much more complicated in the case of a crowded scene. It will be cherry on the cake when we can master the capability of real-time re-identification in a crowded scene from surveillance video feeds with low-resolution availability.

However, real-time re-identification remains a challenging task as it is the process of matching correctly the two image of the same person having some commonness and uniqueness score [7] under substantial appearance, illumination, pose and viewing area changes. One of the reasons for using Deeply Trained Networks is that it is challenging to identify some of the low-level visible features for person identification such as skin color and other facial image feature due to very low-resolution availability.

Despite several impressive research on re-ID using Supervised methods, all these methods require a significant amount of training data, and less attention is given to unsupervised methods. In this paper, we provide an unsupervised approach to re-ID. This improved version of [38], make use of the trained model to extract deep features. Using clustering approach we obtained some original labels for our unlabelled data, we improve the clusters using the K-reciprocal method which is then used as input features to train the network. The advantage of this method is that it is a self-paced learning method as well as it provides the capability of refining the initial cluster which improves the training process significantly. The organization of the paper is as follow. Section 2 provides a brief insight into recent works, dataset and evaluation methods used. Section 3 provides details of the proposed algorithm and re-ranking approach. Section 4 includes implementation and results. Section 5 is the conclusion.

2 Recent Work

2.1 Image Based Re-ID Methods

Person re-ID is primarily divided into image, video and deep learning based systems.

Image based re-ID assumes that given a probe or query image q_j from query database having K images and the database of gallery $\mathcal{G}$ having N images, denoted by $(g_i)_{i=1}^{N}$ (Eq. 1). They belongs to M different identities.

$$\mathcal{G} = (g_i | i = 1, 2, \cdots N) \tag{1}$$

The identity of the query image q_j is obtained by checking the similarity between gallery images and the query image as in Eq. 2

$$q_j^{id} = arg\, min_{i \in 1,2,\cdots N} DIS(q_j, g_i), \quad id \in 1, 2, \cdots M \tag{2}$$

The distance function used (DIS), can be any of the distance metric [8]. Most commonly used metric is Mahalanobis distance (Eq. 3). Where M is a +ve semi definite matrix.

$$d(q_j, g_i) = (x_{q_j} - x_{g_i})^T M (x_{q_j} - x_{g_i}) \tag{3}$$

Image-based person re-ID is divided into Low-level feature based [9–17], and Attribute based approaches [18–23]. Colour and texture features are used as low-level features. Instead of using the low features it is better to use some mid-level descriptors that are more robust to image translation. Layne et al. [18] make use of 15 descriptors based on attire and soft biometric. Liu et al. [19] use improved Latent Dirichlet Allocation (LDA) method. Su et al. [21] use binary person semantic attributes of individuals with the same id from the different viewing area.

2.2 Deeply Learned Re-ID

The foundation stone for CNN based deep learning was laid down by LeCun in 1998 [24] by using LeNet-5, a 5-level CNN that can classify digits. Later on, Krizhevsky et al. [25] designed AlexNet and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC – 2012) by a large margin, and the race begins. Figure 1 depicts major milestone models for Deep Neural Network.

Broadly the CNN methods can be divided into two category Similarity Learning and Representation Learning [36]. Siamese model is used for similarity learning and is preferred when we have less number of training samples. The input to the Siamese model can be single image [26], the pair [27], triplet [28], or quadruplet [29] of images. Some of the initial works on CNN and Siamese model include [26–28,30–33]. In [26] the input to the 2-CNN layers is horizontal overlapping parts of an image, and the outputs are combined to form a vector which is used to obtain the similarity between images using cosine similarity. Li et al. [30] make use of patch matching layer in different horizontal stripes. Wu et al. [31] use an even more extensive network called "PersonNet". Varior et al. [32] use LSTM (Long Short-Term Memory) module into a Siamese model which works as a memory. [27] use gating functions to enhance discriminative ability. Cheng et al. [28] proposed triplet loss function that takes three input images. A detailed justification of using triplet loss function can be found in [28].

Siamese model does not make use of the re-ID labels, and hence another model which seems promising is Classification model or identification model. Representation learning uses the classification model which is preferable when the training dataset is large enough. Softmax loss is mostly used in the classification model.

2.3 Dataset and Evaluation

Dataset. The bottleneck for deep learning model is the lack of training data. Several dataset for image based re-ID are available. Table 1 summarizes commonly used datasets. The smallest one is VIPeR, which consists of 632 ids with

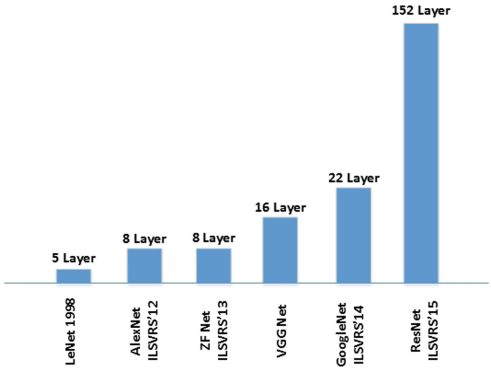

Fig. 1. Milestones models in the field of Deep Neural Network

2 images per ids. In order to reflect various scenarios GRID dataset was collected on a station, CUHK01, CUHK02, CUHK03 and Market-1501 were collected in a university campus. The datasets are improving by adding more and more ids and bounding box in order to provide ample amount of data to train deep CNN models.

As the size of the dataset increases, the bounding box are produced using detectors (DPM [28,35], ACF [34]) rather than hand-drawn. This can cause drop in re-ID accuracy, as the bounding box may deviate from the ideal position while using detectors. More detailed description of the dataset can be found in [6].

Table 1. Datasets used for image based re-ID

Dataset	Year	Individuals	No. of images
VIPeR	2007	632	1,264
iLIDS	2009	119	476
GRID	2009	250	1,275
CAVIAR	2011	72	610
PRID2011	2011	200	1,134
WARD	2012	70	4,786
CUHK01	2012	971	3,884
CUHK02	2013	1,816	7,264
CUHK03	2014	1,467	13,164
RAiD	2014	43	1,264
PRID 450S	2014	450	900
Market-1501	2015	1,501	32,668
DukeMTMC	2016	1,812	36,411

Evaluation. The evaluation of a re-ID system can be performed using Cumulative Match Curve (CMC) and Receiver Operating Characteristic Curve (ROC). CMC [39] is used as a measure to evaluate [1:m] identification system, where as ROC is used for [1:1] identification system. CMC give the performance of the query matching system that returns the ranked list of matching candidates from gallery $\mathcal{G}$. To obtain CMC, the query image is matched with every gallery image and a total K × M scores are obtained.

$$Score_j = [s(q_j, g_1), s(q_j, g_2), \cdots, s(q_j, g_M)], \quad j = 1, 2, \cdots, K \tag{4}$$

The scores are sorted as in Eq. 5 and a rank is provided to every query image based on the position of the matched image in the sorted list.

$$s(q_j, g_1) \geqslant s(q_j, g_2) \geqslant \cdots \geqslant s(q_j, g_M) \tag{5}$$

Now we have a rank associated with each query $Rank_{q_j}$ where $1 \leqslant Rank_{q_j} \leqslant M$. The CMC curve estimates the distribution of these ranks. Such that higher CMC(1) indicates better re-ID system.

$$CMC(1) = \frac{1}{n}(\text{Number of query image with rank} \leqslant 1) \tag{6}$$

The disadvantage of the CMC is that this method can not be used in the scenario when we have multiple ground truth for each query. Hence when multiple ground truth exist in the gallery [17] proposed a method Mean Average Precision (mAP) for the evaluation of the identification system.

3 Proposed Approach

3.1 Problem Definition

From the previous section, it is clear that intensive work has been done on deeply learned person re-ID using supervised learning method. The hunger of more and more data for training is the disadvantage of the supervised methods, and hence more effort has to be taken toward unsupervised learning methods.

In this section, we provide and improved unsupervised learning which is based on the progressive learning method [38]. This method use, an initial trained model (ResNet-50 trained on ImageNet). This initial model is fine tuned on some arbitrary unlabelled data other than the unlabelled dataset which is used for validation and testing. In step second, this fine tuned model is used for unsupervised learning, to extract some features from the unlabelled data. As the data are unlabelled, in step third we have to use some clustering algorithms to obtain initial clusters. Using some threshold on the distance matric use by the clustering algorithm, few samples are selected known as Reliable samples (Samples having significant similarity with cluster center). In step four, only Reliable samples are used to fine tune the original model to generate a new trained model. Step second, third and fourth are repeated until the model stabilizes. And at last

the trained model obtained through subsequent refining is used to extract the feature from the samples used for testing.

The effectiveness of this process depends on the clustering approach used and the number of Reliable features generated which are used to train the model. This improved Unsupervised Learning approach increases the number of Reliable samples generated in each step and in turn improves the training. Less Reliable sample means more iterations of step second, third and fourth. Higher Reliable sample generation means fast and improved training.

3.2 Formulation

Let us consider that we are provided with N unlabelled dataset or images denoted by x_i. These images may belongs to K different individuals. Our task is to assign labels or ids to every unlabelled sample as in Eq. 7, using some model $M_{CNN+Classification}(.;pram)$, which takes as input the samples x_i and initial parameter $pram$.

$$x_i^{id} = M_{CNN+Classification}(x_i; pram), \text{Where } i = 1, 2, \cdots, N \tag{7}$$

The label vector X_ID will contains all the assigned labels to the input samples (Eq. 8)

$$X\_ID = \left[(x_1^{id}, x_2^{id}, \cdots, x_N^{id}), |(1 \leqslant x_i^{id} \leqslant K)\right] \tag{8}$$

The model $M_{CNN+Classification}$ used here consists of CNN module with initial learned parameter θ and a classification module with parameter w at the end. The CNN module takes x_i as input and generates corresponding features f_i using parameter θ.

$$f_i = M_{CNN}(x_i; \theta), \text{Where, } i = 1, 2, \cdots, N \tag{9}$$

Few Reliable features out of the N features generated, are selected and before training these feature are refined using automatic Re-ranking to get final refined reliable features which are then used for training. All the features are assigned a selection indicator value $v_i \in \{0, 1\}$, such that if $v_i = 1$ means the feature is selected else rejected ($v_i = 0$). A selection indicator vector V is used which contains all the values as $V = [v_1, v_2, \cdots v_N]$. After repeated t training's, the final learned model $M_{CNN}(.;\theta_t)$ is used to extract features from the testing samples which are then fed to the classification module $M_{Classification}(.;w)$ to obtain the final class labels. θ_t is the improved parameter obtained by the subsequent training of the CNN model on the training set. The process is depicted in Fig. 2.

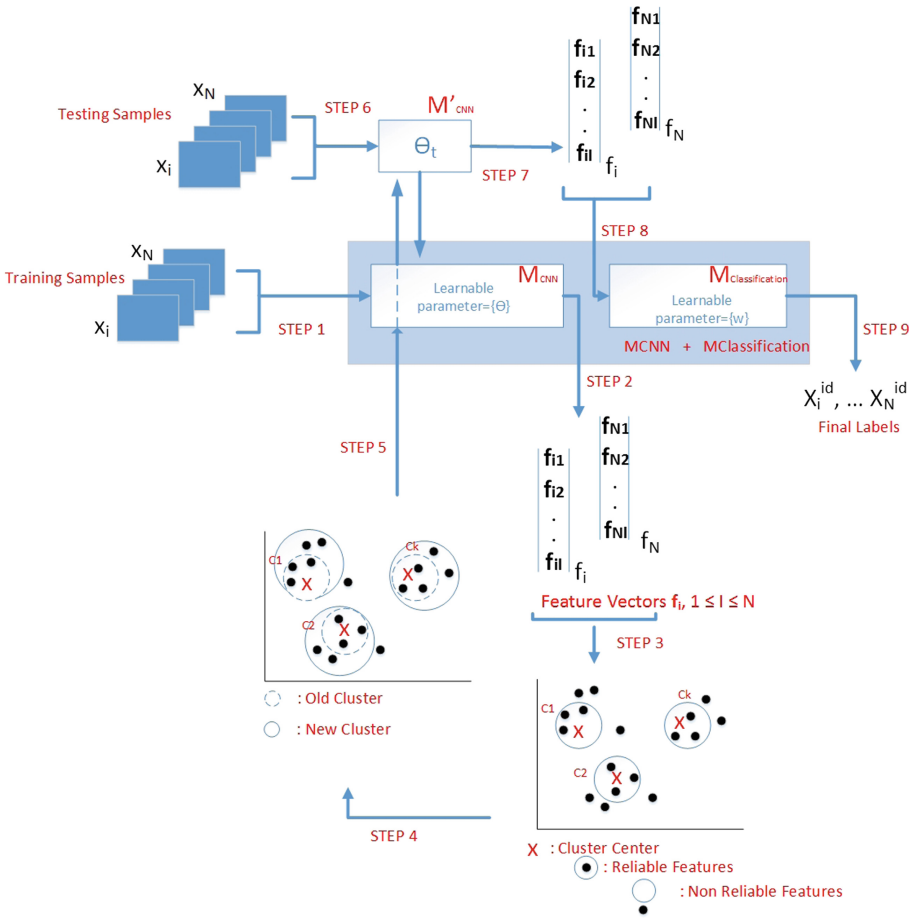

Fig. 2. Illustration of the complete process. The model $M_{CNN} + M_{Classification}$ is ResNet-50 pretrained on ImageNet. STEP 1: Unlabelled training sample are used as input. STEP 2: M_{CNN} submodule of the trained model is used to extract feature vector from each training samples. STEP 3: K-means clustering is used to obtain initial clusters and using distance metric the features are grouped into individual clusters. STEP 4: Features selected to form cluster are reliable features and those left are Non-reliable. Using K-reciprocal neighbourhood some of the non reliable features are selected to refine old clusters and their labels are defined using K-means approach. STEP 5: Clustered features and labels are used to generate a fine-tune submodule M_{CNN}. After $'t'$ iteration of the STEP 2, 3, 4 and 5 a new fine-tuned module with new parameter θ_t is obtained. STEP 6: This fine-tuned model of STEP 5 is used to extract features from testing samples in STEP 7. STEP 8: The extracted features are classified using classification submodule $M_{Classification}$ to generate labels for the testing samples.

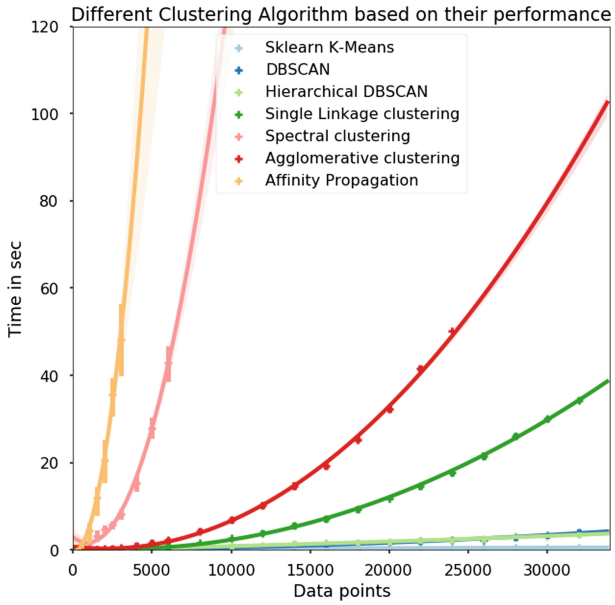

Fig. 3. Performance comparision of different clustering approaches

The idea of this approach is to minimize the equations below:

$$\arg\min_{x_i^{id}, C_k} \Sigma_{k=1}^{K} \Sigma_{x_i^{id}=k} (DIS(x_i^k, C_k)) \tag{10}$$

$$\text{Where, } DIS(x_i^k, C_k) = ||(M_{CNN}(x_i; \theta) - C_k)||^2$$

$$\arg\min_{V} \Sigma_{k=1}^{K} \Sigma_{x_i^{id}=k} v_i DIS(x_i^k, C_k) - \lambda ||V||_1 \tag{11}$$

$$\text{Such that, } v_i \in \{0, 1\}, \text{ and } \Sigma_{x_i^{id}=k} v_i \geqslant 1, \forall k$$

The step in Eq. 10 infers to the label generation process using the clustering method. After feature generation in Eq. 9, these features are clustered using K-means clustering approaches. The clusters should be such that it minimizes the distance between the feature and the cluster center. There are several available approaches such as K-means, Affinity Propagation, Mean Shift, Agglomerative, DBSCAN, and HDBSCAN. The performance analysis for the clustering algorithms is illustrated in Fig. 3. HDBSCAN is a density-based clustering approach and is faster compared to K-means, but K-mean is best suited for data samples where no. of clusters are known in advanced.

After generating initial clusters and labels using K-means, Eq. 11 is used to select reliable features that are close enough to the cluster center. A sample is reliable, if the distance of its feature from the cluster center is smaller than λ.

λ is the threshold that defines which sample or feature to consider reliable, the value should be small as possible to remove the false matches from considering as reliable samples. Specially $\lambda = 0.85$ yields superior accuracy [38]. The condition $\Sigma_{x_i^{id}=k} v_i \geqslant 1, \forall k$ makes sure that the clusters are not blank, and at least one sample (Cluster center) must be there.

As we are dealing with unlabelled data, clustering and reliable features selection is very crucial for this algorithm. First of all, the feature selection must be pure and must not contain any False Matches. We are completely dependent on K-means and distance metric for reliable feature generation, but quality of cluster is dependent on the selection of initial cluster centroid which are generated using k-mean++ algorithm. Second, we are only checking for the distance between the cluster centers and feature. What if a feature which is not close enough to the cluster centers, but bears enough resemblance with the neighbouring reliable features. Now this is where we can increase the number of reliable features generated, having a check that no false matches should be included.

3.3 Re-ranking

After an initial ranking list is obtained it is a good practice to include the re ranking step to improve the ranks. No further requirement of training samples is the main advantage of this step and can be directly applied to the previous step. The underlying assumption is that if an image is returned as the true match of the query image, than it can subsequently be used to find other true matches in its neighbourhood. Inclusion of False Matches may be a problem in K nearest neighbour searching approach and to deal with it we applied a check using distance threshold (λ). In literature, the K- reciprocal nearest neighbour [37] is the effective solution to the problem of false match.

3.4 K-Reciprocal Nearest Neighbour

We can define $N(q, k)$ as the k-nearest neighbour of the query image as:

$$N(q, k) = (g_1, g_2, \cdots, g_k), |\# \text{ of samples in } N(q, k) = k \tag{12}$$

The K-reciprocal nearest neighbour can be defined as,

$$\mathcal{R}(q, k) = \{g_i | (g_i \in N(q, k)) \wedge (q \in N(g_i, k))\} \tag{13}$$

Equation 13 states that the if g_i is selected as the true match for query q, we can find other true matches in the neighbourhood of g_i with the condition that q must also be there. K-reciprocal nearest neighbour can be easily implemented in our approach to improve the generation of reliable features, by considering our cluster center as query q and selected reliable features as $g_i \in N(q, k)$. The new distance between q and g_i can be calculated using Jaccards metric of K-reciprocal

approach as given in Eq. 14. Here $\mathcal{R}^*$ is the improved version of Eq. 13, so that no False match is included due to variation in illumination, pose and view.

$$d_J(q, g_i) = 1 - \frac{|\mathcal{R}^*(q,k) \cap \mathcal{R}^*(g_i,k)|}{|\mathcal{R}^*(q,k) \cup \mathcal{R}^*(g_i,k)|} \tag{14}$$

In order to consider the importance of Euclidean distance in re-ranking the final distance is changed to the one in Eq. 15 and is used to refine the reliable feature selection step.

$$d^*(q, g_i) = (1 - \lambda_J) d_J(q, g_i) + \lambda_J d(q, g_i) \tag{15}$$

Where, $\lambda_J \in [0, 1]$ is different than the KNN distance threshold (λ) one used in Eq. 11, and is used to penalize the features g_i that are far away from the query q. When $\lambda_J = 0$, K-reciprocal distance $d_J(q, g_i)$ is used and when $\lambda_J = 1$ Euclidean distance $d(q, g_i)$ is used. The K-reciprocal approach is depicted in Fig. 4. Figure 5 shows the improvement in the no. of reliable feature generation using the re-ranking method.

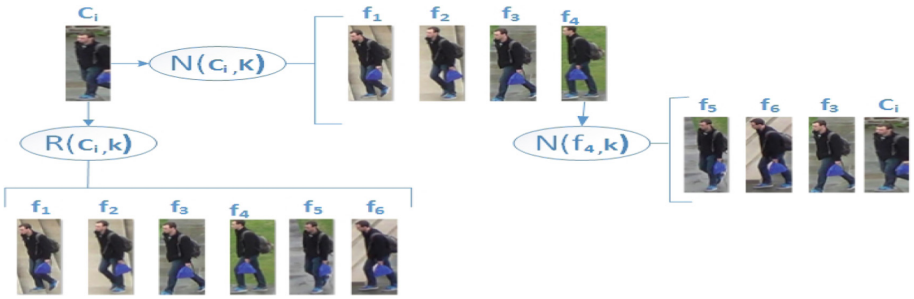

Fig. 4. Illustration of K-reciprocal neighbours approach. C_i is the cluster center and $N(C_i, K)$ is K-nearest neighbour to the cluster center. $\mathcal{R}(C_i, K)$ is the K-reciprocal neighbour to the cluster center. Feature f_4 lies in the K-nearest neighbour of the cluster center C_i. This illustration shows that we can add more reliable features in the cluster of C_i from $N(f_i, K)$ using $\mathcal{R}(C_i, K)$ based on Eq. 13

After repeated clustering and feature selection process of Eqs. 10, 11 and 15 which are used to fine tune the model, Eq. 16 is used to calculate the loss for the classification module.

$$\Sigma_{i=1}^{N}\, v_i \mathcal{L} M_{Classification}(x_i^{id}, M_{CNN}(x_i; \theta_t); w) \tag{16}$$

Here, $\mathcal{L}$ denotes the loss from the classification model. The inputs to the classification model are the labeled samples x_i^{id} (which is obtained using the fine tuned CNN model $M_{CNN}(x_i; \theta_t)$), classification parameter (w) and the selection indicator v_i.

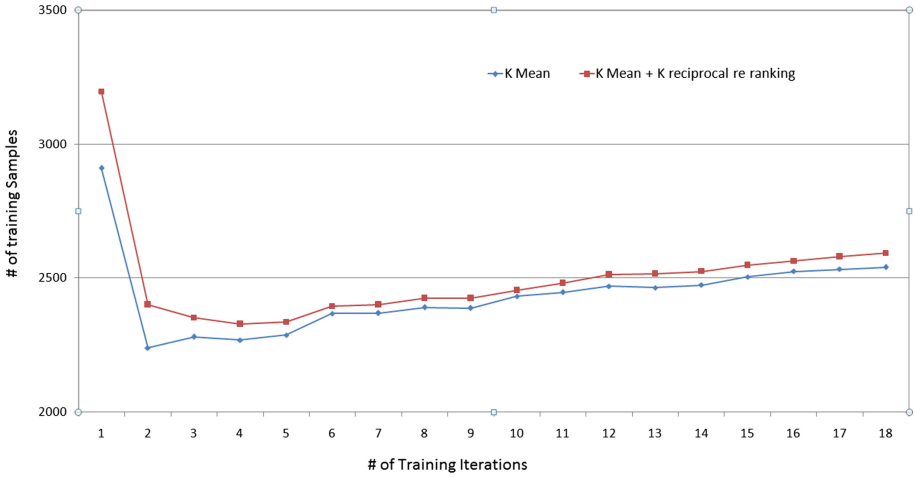

Fig. 5. Impact of using K-reciprocal re-ranking to improve the numbers of reliable feature generation. During iteration 1 only 2910 samples out of the total 3500 are selected using the threshold λ, in order to remove some false match cases. The drop in the reliable sample generation from iteration 1 to 2 is because initially the CNN model was trained on the dataset other than the training dataset and hence classifies more samples as reliable during iteration 1. After first training iteration on actual training set number of reliable features decreases. After subsequent training the model becomes more and more trained and hence from iteration 2 onwards the number of reliable features increases.

Table 2. Improved re-ID tested on Duke dataset

Method	CUHK01				
	rank-1	rank-5	rank-10	rank-20	mAP
Duke	0.375673	0.525135	0.600987	0.6669666	0.205698

4 Experiments

A ResNet-50 model pre-trained on ImageNet is used as our initial CNN model. A dropout of 0.5 is set just after the CNN layer. Every image is resized to 224×224. Parameter λ, which is used to select reliable features is set to 0.85. Parameter λ_J used in jaccards distance is set to 0.3. We have trained our model for 18 iterations with 10 epochs/iteration. The maximum saturation limit achieved was 2,593 reliable image out of 3,884 total images used for training.

The trained model was fine-tuned on CUHK01 dataset having only 3,884 images and is tested on Duke dataset with 17,661 gallery images and 2,228 query images. The result is summarized in Table 2 with Mean Average Precision (mAP) and rank outputs.

5 Conclusion

As discussed earlier that the hunger of more and more labeled samples is the disadvantage of Supervised learning algorithm. It is very tedious to gather labeled sample and have various problems when using automated tools.

In this paper we discussed an unsupervised approach of person re-id and tried to improve the reliable feature generation step using K-reciprocal nearest neighbour method. The reliable features are used to train the deep neural network model, and improving feature generation helps improving the training. The experimental results are provided using small datasets and can be further improved by providing training on larger datasets available such as Market and DukeMTMC.

References

1. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. arXiv preprint arXiv:1610.02984, 10 October 2016
2. Martinel, N., Foresti, G.L., Micheloni, C.: Person reidentification in a distributed camera network framework. IEEE Trans. Cybern. **47**(11), 3530–3541 (2017)
3. Ma, A.J., Yuen, P.C., Li, J.: Domain transfer support vector ranking for person re-identification without target camera label information. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3567–3574 (2013)
4. Ma, A.J., Li, J., Yuen, P.C., Li, P.: Cross-domain person reidentification using domain adaptation ranking SVMs. IEEE Trans. Image Process. **24**(5), 1599–1613 (2015)
5. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
6. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. Image Vis. Comput. **32**(4), 270–86 (2014)
7. Xu, Y., Ma, B., Huang, R., Lin, L.: Person search in a scene by jointly modeling people commonness and person uniqueness. In: Proceedings of the 22nd ACM International Conference on Multimedia, 3 November 2014, pp. 937–940. ACM (2014)
8. Yang, L., Jin, R.: Distance metric learning: a comprehensive survey, vol. 2, no. 2, p. 4. Michigan State Universiy, 19 May 2006
9. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 13 June 2010, pp. 2360–2367. IEEE (2010)
10. Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: Null, 17 June 2006, pp. 1528–1535. IEEE (2006)
11. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_21
12. Mignon, A., Jurie, F.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 16 June 2012, pp. 2666–2672. IEEE (2012)

13. Shen, Y., Lin, W., Yan, J., Xu, M., Wu, J., Wang, J.: Person re-identification with correspondence structure learning. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3200–3208 (2015)
14. Das, A., Chakraborty, A., Roy-Chowdhury, A.K.: Consistent re-identification in a camera network. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 330–345. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_22
15. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3318–3325 (2013)
16. Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3550–3557 (2014)
17. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
18. Layne, R., Hospedales, T.M., Gong, S., Mary, Q.: Person re-identification by attributes. In: BMVC 2012, vol. 2, no. 3, p. 8, September 2012
19. Liu, X., Song, M., Zhao, Q., Tao, D., Chen, C., Bu, J.: Attribute-restricted latent topic model for person re-identification. Pattern Recogn. **45**(12), 4204–4213 (2012)
20. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012. LNCS, vol. 7583, pp. 391–401. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33863-2_39
21. Su, C., Yang, F., Zhang, S., Tian, Q., Davis, L.S., Gao, W.: Multi-task learning with low rank attribute embedding for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3739–3747 (2015)
22. Shi, Z., Hospedales, T.M., Xiang, T.: Transferring a semantic representation for person re-identification and search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4184–4193 (2015)
23. Li, D., Zhang, Z., Chen, X., Ling, H., Huang, K.: A richly annotated dataset for pedestrian attribute recognition. arXiv preprint arXiv:1603.07054, 23 March 2016
24. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
25. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
26. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: 2014 22nd International Conference on Pattern Recognition (ICPR), 24 August 2014, pp. 34–39. IEEE (2014)
27. Varior, R.R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 791–808. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_48
28. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1335–1344 (2016)
29. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1 July 2017, vol. 2, no. 8 (2017)

30. Li, W., Zhao, R., Xiao, T., Wang, X.: DeepReID: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)
31. Wu, L., Shen, C., Hengel, A.V.: PersonNet: person re-identification with deep convolutional neural networks. arXiv preprint arXiv:1601.07255, 27 January 2016
32. Varior, R.R., Shuai, B., Lu, J., Xu, D., Wang, G.: A siamese long short-term memory architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 135–153. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_9
33. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification. arXiv preprint arXiv:1606.04404, 14 June 2016
34. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep attributes driven multi-camera person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 475–491. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_30
35. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2010)
36. Dollár, P., Appel, R., Belongie, S., Perona, P.: Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014)
37. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21 July 2017, pp. 3652–3661. IEEE (2017)
38. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: clustering and fine-tuning. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **14**(4), 83 (2018)
39. Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: The relation between the ROC curve and the CMC. In: Null, 17 October 2005, pp. 15–20. IEEE (2005)