Police-In-The-Loop Person Re-identification

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Person re-identification (Re-ID) has gained significant interest in the computer vision community. One important application of Re-ID is the analysis of CCTV footages for police investigations. This setting poses a challenge of identifying all the appearances of a given identity on time while also providing an opportunity to utilize feedback from a human investigator. In this paper, we propose a general framework for building human-in-the-loop (HITL) Re-ID systems for criminal investigations. Using this framework, we investigate various HITL re-ranking methods based on relevance feedback and hierarchical clustering. Quantitative evaluations on large scale datasets show significant improvements in Re-ID performance. Code is available at `https://`.

1 Introduction

Person re-identification (Re-ID) is the task of retrieving a specific person within a gallery of images from a network of disjoint cameras, where a query image of the person is given. Due to the rising demand for public safety and the abundance of CCTV cameras, Re-ID has become a crucial component in intelligent surveillance systems [23]. However, many of the evaluation protocols for current Re-ID methods focus mainly on a static ranking of gallery images given by an end-to-end model.

Our work focuses on the specific setting of the police investigation. The use of CCTV in criminal investigations has become prominent in identifying the activity of criminal suspects [8]. In serious crimes, CCTV is often the first focus of an investigation team [5], as tracking the suspect on multiple footages based on the initial crime footage is critical to advancing investigations [1]. The nature of police investigation comes with its challenges that one must tackle in Re-ID systems, as well as opportunities that one may exploit to aid in solving these challenges:

- (Opportunity) In criminal investigations, person re-identification is performed by a human investigator who can provide ground-truth feedback during the re-identification process.
- (Challenge 1) Timely identification is required, as delays in any step of the investigation have a direct impact on arrests [5].
- (Challenge 2) Re-ID systems must identify every occurrence of a suspect on camera to prevent the omission of crucial evidence [1].

These characteristics inspire the use of human-in-the-loop (HITL) methods for Re-ID, where gallery rankings are updated after each round of human annotation based on the feedback provided. Several HITL methods have been explored for traditional Re-ID methods[20, 11, 9, 21]. However, as far as we know, these methods have not been revisited in recent deep Re-ID methods, despite having clear advantages for the challenges above.

Our paper addresses these issues with the following contributions.

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

- We identify the opportunities and challenges of the police investigation task and propose a specialized HITL framework.
- We show that using hierarchical clustering methods with the existing state-of-the-art deep learning models are advantageous.

2 Related Works

2.1 Re-Identification

Traditional person Re-ID methods have used hand-crafted features such as low-level image features[2] and SIFT descriptors[12]. Due to its limitations, previous methods focus on learning the embeddings from the fixed features, including those based on the Mahalanobis distance metric[6, 17] and graph propagation[11]. However, the recent advent of convolution neural networks[7] and deep learning-based approaches in Re-ID research[13, 24, 26, 19] has made it possible to learn better feature embeddings. Classification loss is commonly employed, but additional loss functions have been proposed such as triplet loss[4]. These loss functions allow the model to learn the desired embedding space directly, diminishing the need for further embedding space transformation. Consequently, the distance matrix is often derived directly from the embedding using Euclidean distance or cosine distance [23]. Additional ranking enhancements such as [22] and k-reciprocal encoding [27] have also been adopted in current state-of-the-art methods.

2.2 Human-In-The-Loop Re-ID

Multiple HITL-related methods have been incorporated in Re-ID, such as POP[11] and HVIL[20]. These works focus on improving the embedding of traditional Re-ID models using iterative human feedback. In contrast, the Rocchio algorithm [9, 16] does not modify the gallery embeddings but shift the initial query closer to the positive samples and away from the negative samples. However, HITL methods are relatively unexplored in deep Re-ID models, as far as we know.

We also look into hierarchical clustering methods that can effectively accept positive samples on-the-fly, such as single-linkage clustering (SLC), complete-linkage clustering (CLC), and unweighted average linkage clustering (UPGMA) [18]. Each method is nearly identical except the distance function between embedding clusters; minimum (SLC), maximum (CLC), average (UPGMA) distance is chosen between samples in different clusters. To our knowledge, these methods are also not explored in-depth in deep Re-ID settings.

3 Re-Identification With Police-In-The-Loop

3.1 Considerations for police investigations

We translate the details of police investigations into the deep learning domain to show that pre-existing approaches may induce several complications.

Opportunity The human supervisor behind the Re-ID system is a crucial factor in real-world situations to ensure no wrong decisions are made. In other words, there exists a human oracle as an operator, and the machine should only recommend the targets and not automate the identification process as a whole. Therefore, we have to view this system as a feedback system that employs positive and negative feedback on the fly. Note that these recommendations must be updated in a swift manner to be displayed in real-time, hence restricting the use of time-consuming methods, e.g., HVIL or POP.

Challenge 1 As there is limited time to pursue the evidence [5], we must view the human oracle as a limited resource. Hence, the Re-ID system must aim for high recall given a fixed amount of viewed items. As the resource amount may vary, the Re-ID system must be robust in handling differing scenarios.

Challenge 2 Any occurrence may be the most crucial evidence to the investigation. Therefore, Re-ID systems must aim for extremely high recall. In other words, we must design the system with worst-case queries in mind, where the positive samples may be placed far below the initial gallery list.

3.2 Metrics

To reflect the aforementioned scenarios, we recommend the following metrics to evaluate the Re-ID systems in police investigations.

Precision-recall curve and mean average precision (mAP) mAP has been used in the deep learning domain for quite some time, as it can depict the overall trade-off between precision and recall, which is directly related to both challenges. We use mAP to show the efficacy of various methods in a general sense.

Mean inverse negative penalty (mINP) mINP is a relatively new metric introduced in [23], which considers the rank position of the hardest correct match. This metric is useful in evaluating HITL methods concerning the second challenge – finding all matching identities.

Last-match rankings In addition to mINP, we also look into the last-match rankings of gallery images to evaluate our methods in the worst case. For each given query, we identify the number of human feedback rounds required to find the last matching identity within the gallery. We display the worst last-match rankings that each method has yielded.

Recall-to-rank curve We examine the recall-to-rank curve to identify the portion of matching images a human operator can find for a given number of feedback rounds. We use this curve to evaluate our methods in terms of the first challenge mentioned above – achieving high recall when there is a limited time budget.

3.3 General framework

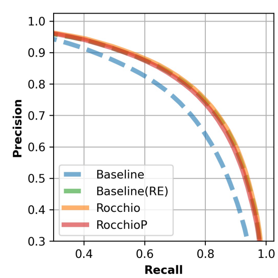
To sum up, we propose a general framework that reflects the aforementioned considerations.

1. Object detection: Human targets have to be identified in a bounding box manner to be used in state-of-the-art deep Re-ID methods.
2. Feature extraction: Informative features must be extracted from the images before re-identification. Recent deep models learn embedding spaces optimized for the Re-ID task directly, skipping the next step.
3. Feature vector adjustment: The embedding vectors are further adjusted so that the sample distances follow the semantic similarities more closely, exemplified by HVIL[20] or the Rocchio algorithm[10, 16].
4. Generating a distance matrix using the feature vectors: Modern methods often use Euclidean distance or cosine distance directly, as deep models already learn an embedding space optimized for re-identification. Approximated nearest neighbor algorithms [3, 14] can be used to avoid calculating the matrix as a whole, despite having potential limitations in the later procedures.
5. Distance matrix adjustment: The distance matrix can be further tweaked to enhance the rankings for identification. K-reciprocal encoding based re-ranking [27] is a popular method that is used to enhance the distance matrix for re-identification without human interaction.
6. Ranking: Final ranks are derived from the distance matrix by ordering the images in ascending order of distance. Ranked gallery images are shown to the human oracle for subsequent annotation.

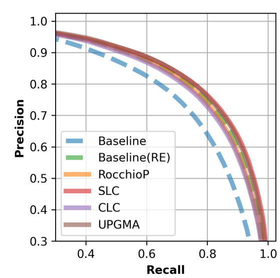
HITL methods can be considered in both the 3. Feature vector adjustment and 5. Distance matrix adjustment stages. Various methods can be applied in tandem within the loop. For example, it is possible to append HITL methods for distance matrix adjustment after k-reciprocal re-ranking, as the outputted result is also a distance matrix. Note that the time constraints must be considered when mixing such adjustment methods. E.g., it is not feasible to pair feature adjustment HITL methods with k-reciprocal re-ranking, as each feature adjustment would yield a different distance matrix, requiring another round of k-reciprocal re-ranking. This makes it infeasible for Re-ID systems in police investigations.

	Method	mAP	mINP		Method	mAP	mINP
Market	Baseline	0.8936	0.6072	Duke	Baseline	0.8389	0.4008
	Baseline(RE)	0.9549	**0.8822**		Baseline(RE)	0.9226	0.7204
	Rocchio	0.9529	0.7957		Rocchio	0.9212	0.5964
	RocchioP	0.9454	0.7634		RocchioP	0.9080	0.5964
	SLC	0.9620	0.8676		SLC	**0.9420**	**0.7222**
	CLC	0.9400	0.7437		CLC	0.8887	0.4938
	UPGMA	0.9611	0.8409		UPGMA	0.9351	0.6720
	SLC(RE)	0.9617	0.8645		SLC(RE)	0.9393	0.7160
	CLC(RE)	0.9549	**0.8822**		CLC(RE)	0.9226	0.7204
	UPGMA(RE)	**0.9635**	0.8799		UPGMA(RE)	0.9384	0.7188

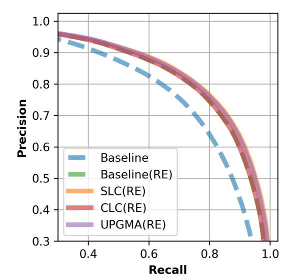
Table 1: mAP and mINP scores on Market and Duke dataset.



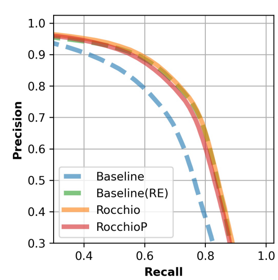
(a) Market - Rocchio



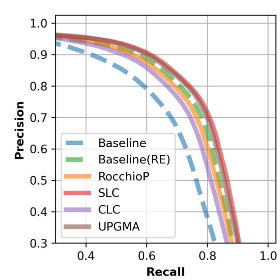
(b) Market - Cluster



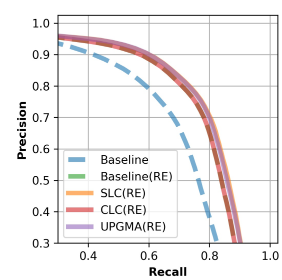
(c) Market - Cluster (RE)



(d) Duke - Rocchio

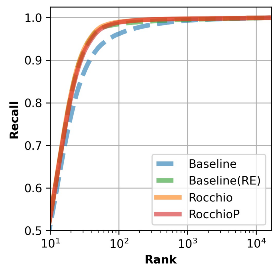


(e) Duke - Cluster

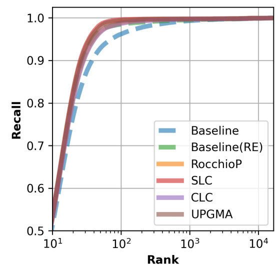


(f) Duke - Cluster (RE)

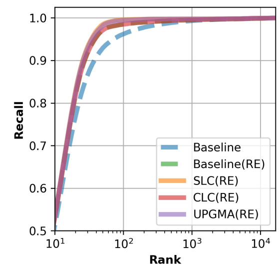
Figure 1: Precision-recall curve on Market and Duke dataset. We show only the partial curve to show the distinction between methods. Curves become hard to differentiate near Precision= 1.0 and Recall= 1.0.



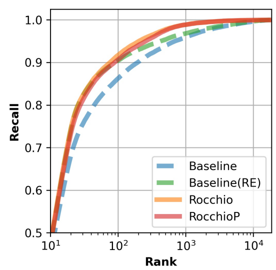
(a) Market - Rocchio



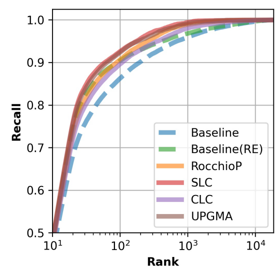
(b) Market - Cluster



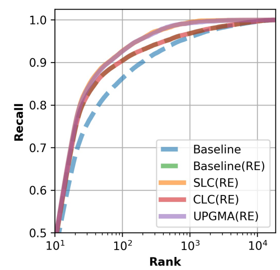
(c) Market - Cluster (RE)



(d) Duke - Rocchio

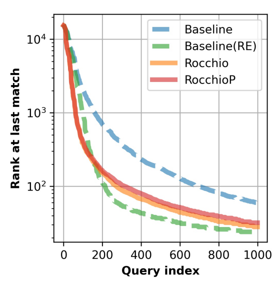


(e) Duke - Cluster

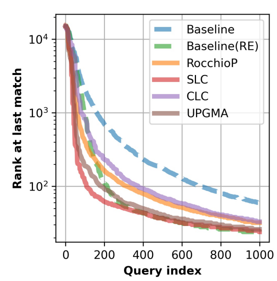


(f) Duke - Cluster (RE)

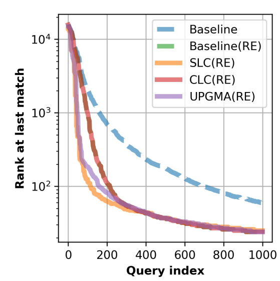
Figure 2: Recall-rank curves on Market and Duke dataset. The x-axis is plotted in logarithmic scale for visual clarity. We show only the partial curve to show the distinction between methods. Curves become hard to differentiate for Recall < 0.5.



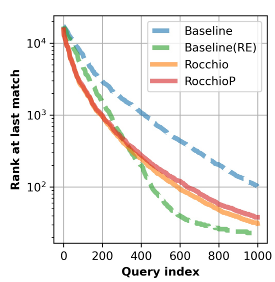
(a) Market - Rocchio



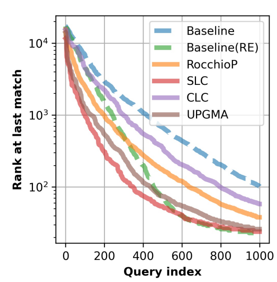
(b) Market - Cluster



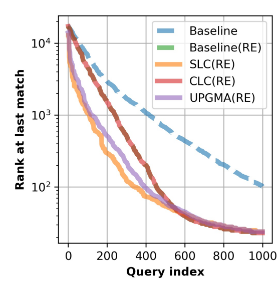
(c) Market - Cluster (RE)



(d) Duke - Rocchio



(e) Duke - Cluster



(f) Duke - Cluster (RE)

Figure 3: Last-match rankings on Market and Duke dataset. 1000 most difficult queries are plotted for each method, sorted in descending order. The y-axis is plotted in logarithmic scale for visual clarity. We show rankings for only 1000 queries to focus on difficult queries that require more than 100 feedback rounds for full match retrieval.

4 Experimental Results

We concentrate on the HITL adjustment methods of the framework above, comparing the effectiveness of Rocchio and hierarchical clustering HITL methods on deep Re-ID systems. We use the Market-1501[25] (Market) and DukeMTMC-ReID[15] (Duke) datasets and apply the aforementioned framework as follows:

1. As with most person re-identification datasets, object detection is performed during dataset development, providing us with bounded box images of each person in a CCTV frame.
2. Deep image feature vectors are extracted using a baseline method for deep re-ID [13].
3. For HITL feature vector adjustment, we consider the Rocchio algorithm, using parameters suggested in [10].
4. Euclidean distance is used to build the distance matrix.
5. For distance matrix adjustment, we consider k-reciprocal re-ranking as an offline method, as well as three HITL methods based on the hierarchical clustering methods: SLC, CLC, and UPGMA.
6. Finally, we rank the gallery images in ascending order of distance and simulate human oracle feedback by providing ground-truth annotation for the top-ranked gallery image, for each query.

As hierarchical clustering methods only make use of positive feedback, we additionally test a variant of Rocchio that does not utilize negative feedback for comparison, by setting the γ parameter to 0. We denote this method as RocchioP. Also, we consider a mixture of k-reciprocal re-ranking and hierarchical clustering methods, by applying the HITL methods on top of re-ranked distance matrices. We append (RE) to denote methods that use re-ranking.

Advantages of HITL methods in police investigation All post-ranking methods show better overall performance compared to the baseline. In terms of precision-recall, we find that SLC and UPGMA methods perform best, while Rocchio and CLC perform sub-optimally to Baseline(RE) that does not utilize human feedback (Table 1, Fig. 1). mINP results (Table 1) show that Baseline(RE) and SLC perform similarly. However, last-match rankings (Fig. 3) reveal that SLC is more effective at reducing the number of annotations required to identify all matches for difficult queries without hurting the general performance.

Existence of local mini-clusters Among the hierarchical clustering methods, we find that SLC outperforms all methods, with the exception of UPGMA(RE) on Market by a small margin. We also find that CLC shows the lowest performance among clustering methods. This suggests the existence of multiple mini-clusters for a single identity within the embedding space, rather than a single cluster.

Effects of k-reciprocal re-ranking on HITL methods We find that prepending k-reciprocal re-ranking to CLC and UPGMA consistently improves results, while the performance of SLC is marginally degraded. The reason for this is unclear, but we suspect that the use of k-reciprocal neighbors encourages the discovery of local clusters, whilst interfering with SLC in this regard, due to the inclusion of irrelevant images in k-reciprocal neighbors.

Utilizing negative samples RocchioP shows consistent performance degradation compared to Rocchio, which suggests that modifying hierarchical clustering methods to additionally use negative samples may yield better results.

5 Conclusion

In this paper, we propose a HITL Re-ID framework specialized for the police investigation and its evaluation criteria based on the analysis of the unique characteristics of the setting. Our findings suggest the existence of multiple mini-clusters in the embedding space of deep Re-ID models, where SLC handles most effectively. Furthermore, our results show that various re-ranking methods in the distance matrix optimization stage can be mixed to boost performance. Future work will focus on understanding the characteristics of the embedding space and designing novel methods that can exploit them.

Broader Impact

Our methods can help investigators swiftly identify CCTV appearances of a suspect, allowing the timely retrieval of evidence which may be critical to the outcome of criminal investigations. Our framework and findings can help guide future development of human-in-the-loop Re-ID systems for police surveillance, and possibly other applications.

On the other hand, a malicious actor can take advantage of Re-ID systems to monitor unknowing pedestrians or vehicles without consent. Therefore, caution must be taken to establish regulations and implement safeguards to control the use of Re-ID systems.

References

[1] Matthew PJ Ashby. The value of cctv surveillance cameras as an investigative tool: An empirical analysis. *European Journal on Criminal Policy and Research*, 23(3):441–459, 2017.

[2] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European conference on computer vision*, pages 262–275. Springer, 2008.

[3] Ruiqi Guo, Philip Sun, Erik Lindgren, Quan Geng, David Simcha, Felix Chern, and Sanjiv Kumar. Accelerating large-scale inference with anisotropic vector quantization. In *Proc. of the 37th International Conference on Machine Learning*, 2020.

[4] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification, 2017.

[5] Garda Inspectorate. Crime investigation. *Dublin: Garda Síochána Inspectorate*, 2014.

[6] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2288–2295. IEEE, 2012.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[8] Tom Levesley, Amanda Martin, and G Britain. *Police Attitudes to and Use of CCTV*. Home Office London, 2005.

[9] Wei-Chao Lin, Zong-Yao Chen, Shih-Wen Ke, Chih-Fong Tsai, and Wei-Yang Lin. The effect of low-level image features on pseudo relevance feedback. *Neurocomputing*, 166:26 – 37, 2015. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2015.04.037. URL `http://www.sciencedirect.com/science/article/pii/S0925231215005007`.

[10] Wei-Chao Lin, Zong-Yao Chen, Shih-Wen Ke, Chih-Fong Tsai, and Wei-Yang Lin. The effect of low-level image features on pseudo relevance feedback. *Neurocomputing*, 166:26–37, 2015.

[11] Chunxiao Liu, Chen Change Loy, Shaogang Gong, and Guijin Wang. Pop: Person re-identification post-rank optimisation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[12] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[13] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification, 2019.

[14] Bilegsaikhan Naidan, Leonid Boytsov, Yury Malkov, and David Novak. Non-metric space library manual. *arXiv preprint arXiv:1508.05470*, 2015.

[15] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*, pages 17–35. Springer, 2016.

[16] Joseph Rocchio. Relevance feedback in information retrieval. *The Smart retrieval system-experiments in automatic document processing*, pages 313–323, 1971.

[17] Peter M Roth, Martin Hirzer, Martin Köstinger, Csaba Beleznai, and Horst Bischof. Mahalanobis distance learning for person re-identification. In *Person re-identification*, pages 247–267. Springer, 2014.

[18] Peter HA Sneath, Robert R Sokal, et al. *Numerical taxonomy. The principles and practice of numerical classification.* 1973.

[19] Guangcong Wang, Jianhuang Lai, Peigen Huang, and Xiaohua Xie. Spatial-temporal person re-identification, 2018.

[20] Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang. Human-in-the-loop person re-identification, 2018.

[21] Bin Xu, Jiajun Bu, Chun Chen, Deng Cai, Xiaofei He, Wei Liu, and Jiebo Luo. Efficient manifold ranking for image retrieval. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 525–534, 2011.

[22] Mang Ye, Chao Liang, Zheng Wang, Qingming Leng, and Jun Chen. Ranking optimization for person re-identification via similarity and dissimilarity. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1239–1242, 2015.

[23] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook, 2020.

[24] Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, and Jian Sun. Alignedreid: Surpassing human-level performance in person re-identification, 2018.

[25] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

[26] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification, 2019.

[27] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding, 2017.