Applying Hard Positive Mining and its Evaluation for Person Re-identification

Yuki Hiroi\textsuperscript{1a)} and Wataru Kameyama\textsuperscript{2}

\textsuperscript{1}Department of Computer Science and Communications Engineering, Graduate School of Fundamental Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

\textsuperscript{2}Faculty of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

\textsuperscript{a)} yuu-hiro@ruri.waseda.jp

Abstract: “Person Re-identification” technology has attracted more attention because of the proliferation of surveillance cameras and the increased awareness of crime prevention. However, there are various problems. One of them is that the same person is likely to be recognized erroneously as a different person due to the differences in persons’ appearance. Therefore, in this paper, we propose a method of hard positive mining to achieve better performance in identifying the same person images that are difficult to identify as the same person. The evaluation results using MARS dataset show that the CMS by applying the proposed hard positive mining is higher than the existing method of CNN with temporal pooling. It suggests that it is effective to apply hard positive mining in MARS dataset.

Keywords: Person re-identification, hard positive mining, hard sample mining, convolutional neural network

Classification: Multimedia systems for communications

References

\cite{1} K. Chen, Y. Chen, C. Han, N. Sang, C. Gao and R. Wan, “Improving Person Re-Identification by Adaptive Hard Sample Mining,” in 2018 IEEE International Conference on Image Processing (ICIP), pp.1638-1642, 7-10 October 2018, Athens, Greece.

\cite{2} H. Sheng, Y. Zheng, W. Ke, D. Yu, X. Cheng, W. Lyu, and Z. Xiong, X. Cheng, W. Lyu and Z. Xiong, “Mining Hard Samples Globally and Efficiently for Person Re-identification,” in IEEE Internet of Things Journal, 13 March 2020, DOI: 10.1109/JIOT.2020.2980549.

\cite{3} Q. Qian, L. Shang, B. Sun, J. Hu, H. Li and R. Jin, “SoftTriple Loss: Deep Metric Learning Without Triplet Sampling,” in 2017 IEEE International Conference on Computer Vision (ICCV), pp.6450–6458, 27 Oct.–2 Nov. 2019, Seoul, Korea.

\cite{5} J. Gao and R. Nevatia, “Revisiting Temporal Modeling for Video-based Person
ReID,” arXiv:1805.02104v2, 2018.
[4] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, “Mars: A Video Benchmark for Large-scale Person Re-identification,” in European Conference on Computer Vision, pp.868–884, Springer, 2016.
[6] Z. Zhong, L. Zheng, D. Cao and S. Li, “Re-ranking Person Re-identification with k-reciprocal Encoding,” in Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pp. 3652–3661, 21-26 July 2017, Honolulu, HI, USA.

1 Introduction

In recent years, with the proliferation of surveillance cameras and the increased awareness of crime prevention, “Person Re-identification (ReID)” becomes one of the hot research issues, which is the task to identify the same person between non-overlapped cameras. It is useful for many different applications including crime prevention and marketing.

A lot of studies have been done on ReID, but it’s still difficult to identify the same person because of persons having some item or not, different postures, and different view angles taken by different cameras. A sample image in which the same person is recognized erroneously as a different person is called a hard positive sample. On the other hand, a sample image that is likely to identify another person erroneously as the same person is called a hard negative sample. The hard positive samples and the hard negative samples are together called to hard samples in ReID. Some methods of hard sample mining which improve to recognize hard samples are proposed, such as [1] and [2]. [3] focuses on hard negative mining which improve to recognize hard negative samples by using the proposed triplet loss. However, the ever-proposed methods need to modify network models and loss functions, which make it difficult to include them into the existing network models. And the more the data is required for hard sample mining to be processed, the more the computational cost is required. Moreover, nowadays, since a huge size of persons’ images can be collected, it makes much larger diversity in images for a single person. Therefore, unless selecting good hard samples for training without considering the diversity, a non-essential model may be created to produce poor performance.

Therefore, in this paper, we focus on hard positive samples to consider the diversity in images for one person, and propose a hard positive mining (hereafter referred to as HPM) method to achieve better performance in recognition of hard positive samples. The proposed method trains the networks with optimized learning-image orders and sets considering the hard positive samples, and it is independent from any network architectures and loss functions. It utilizes CMS (Cumulative Match Score), which is mainly used as the evaluation metrics in ReID, as the index for detecting hard positive samples. Using the CMS, time-series of images (hereafter referred to a clip) of the same person are ranked, and the clips of lower ranks that are taken as the detected hard positive samples are trained at the end of the epochs where the loss function is converging in order to reduce the
computational cost and to prevent the overfitting. We compare the results using the MARS dataset [4] with [5] which achieves the highest performance in the dataset where CNN with temporal pooling are used without HPM.

2 Proposed Method

In [5], clips of the same class and different classes are trained per batch, where the batch size is $B$ and the number of clips per class in each batch is $N$. Each batch contains $B/N$ different classes that are randomly selected, and each class contains $N$ clips. In addition, the number of clips per a class in dataset is $M$ whose value varies depending on the class in the dataset, but the value of $N$ is determined as a fixed one depending on, for example, memory and computational costs. And $N$ clips of the same class to be included in a batch are randomly selected from the $M$ clips. Due to this random selection, there is an issue to increase the possibility of recognizing the same class as a different class. In order to improve this, we propose an HPM method that positively trains the same person clips that are difficult to identify as the same class. In order to detect the hard positive samples, the CMS values are used to rank $M$ clips of each class and to select $N$ lower rank clips that are difficult to identify among $M$ clips.

In the proposed method, first, $N$ and $M$ are compared. If $N$ is greater than or equal to $M$, classes to be in a batch are randomly selected, then clips in a selected class are also randomly selected, allowing duplications, as same in [5]. However, if $N$ is less than $M$, first, classes are randomly selected as same in [5], but after that, $M$ clips of each class are ranked by CMS. In general, CMS is used for the evaluation index of ReID, which is calculated from samples (images or clips) in various different classes, and those of all classes are ranked according to the distance of features from small to large for each sample in a class by CMS. However, in this proposal, CMS is used to detect clips including hard positive samples by calculating it from clips in the same class, where the clips in the same class are ranked according to the distance of features from small to large for each clip in a class by the calculated CMS. Then, the HPM index is calculated by Eq. (1).

$$ HP_K = \sum_{i=0}^{N-1} x_K^{M-i} (N - i) \alpha $$

Where $HP_K$ is the HPM index of clip $K$ ($K \in \mathbb{Z} | K \in [0, M - 1]$), $x_K^R$ is the number of clip $K$ in all the Rank $R$ of the clips calculated by CMS, and $\alpha$ is a weight. After calculating $HP_K$, $N$ clips are selected in the descending order of $HP_K$ for each class. As mentioned above, each batch contains $B/N$ different classes. Therefore, CMS and Eq. (1) are calculated $B/N$ times per batch in the proposed method. Finally, all the selected $B$ clips are put in a batch for training.

Table I shows an example of the proposed HPM method, where a class is randomly selected, $M = 5$ (Clip 0 to Clip 4), $N = 3$ and $\alpha = 10$. Because $N < M$, each $HP_K$ has to be calculated. As shown in Table I, Clip 0 appears once in Rank 5, none in Rank 4 and twice in Rank 3. Therefore, $x_0^5 = 1$, $x_0^4 = 0$ and $x_0^3 = 2$. Then, $HP_0$ is calculated to 50 by Eq. (1). Similarly, $HP_1$, $HP_2$, $HP_3$ and $HP_4$ are
40, 60, 90 and 80, respectively. Because \( N = 3 \), i.e. three clips are to be selected, Clip 3, Clip 4 and Clip 2 are selected in the descending order of \( H_{P_k} \) as they are to contain hard positive samples for training.

### Table I. An Example of the Proposed Hard Positive Mining

| Clip | Rank 1 | Rank 2 | Rank 3 | Rank 4 | Rank 5 | \( H_{P_k} \) \( (\alpha = 10, N = 3) \) |
|------|--------|--------|--------|--------|--------|----------------------------------|
| 0    | 0      | 1      | 3      | 2      | 4      | \( 1 \times 3 \alpha + 0 \times 2 \alpha + 2 \times \alpha = 50 \) |
| 1    | 1      | 0      | 2      | 4      | 3      | \( 0 \times 3 \alpha + 2 \times 2 \alpha + 0 \times \alpha = 40 \) |
| 2    | 2      | 1      | 0      | 3      | 4      | \( 1 \times 3 \alpha + 1 \times 2 \alpha + 1 \times \alpha = 60 \) |
| 3    | 3      | 4      | 0      | 1      | 2      | \( 1 \times 3 \alpha + 1 \times 2 \alpha + 2 \times \alpha = 90 \) |
| 4    | 4      | 2      | 3      | 1      | 0      | \( 2 \times 3 \alpha + 1 \times 2 \alpha + 0 \times \alpha = 80 \) |

### 3 Experiment

In the experiment, we use MARS which is a large person image dataset taken by different surveillance cameras. In MARS, there are about 1,260 persons and about 1 million images in total. The ratio of the training data and the test data used in this experiment is 5:5 of random selection. This experiment is to compare the baseline [5] applying the proposed HPM with the baseline of no HPM. CMS is used for the evaluation metrics. In addition, we apply Re-rank [6] to both which improves CMS [5]. Table II shows the parameters of this experiment. As shown in Table II, applying HPM is started from the 700th epoch in order to positively train the hard positive samples only at the end of epochs where the loss function is converging. In addition, the value of weight \( \alpha \) is set to 10 from the results of our preliminary experiment.

### Table II. Parameters of Experiment

| Parameter                          | Baseline[5] | Baseline[5] + HPM |
|------------------------------------|-------------|------------------|
| Epoch                              | 800         | 800              |
| \( N \)                            | 4           | 4                |
| \( M \) in Average                  | 13.2        | 13.2             |
| \( B \)                            | 32          | 32               |
| The Number of Frame per Clip       | 3           | 3                |
| Input Image Size                   | (112, 224)  | (112, 224)       |
| Optimization                       | Adam        | Adam             |
| Activation                         | ReLU        | ReLU             |
| \( \alpha \)                        | -           | 10               |
| Starting Epoch of HPM              | -           | 700              |

### 4 Results

Table III shows the results in mAP (mean Average Precision) and CMS of ranks.
As shown in Table III, the accuracies with the proposed HPM are higher than those of the baseline except for Rank 10. In particular, in Rank 1, the accuracy of the proposed HPM is about 2.3 points higher than that of the baseline. Therefore, it is considered that the proposed HPM is effective in MARS.

Table III. Result of Experiment

| mAP / Rank | Baseline[5] | Baseline[5] + HPM |
|------------|-------------|-------------------|
| mAP        | 84.5812     | 85.3410           |
| Rank 1     | 84.2935     | 86.5761           |
| Rank 2     | 90.4348     | 92.0109           |
| Rank 3     | 92.1196     | 92.9891           |
| Rank 4     | 92.9348     | 93.5326           |
| Rank 5     | 93.7500     | 94.1304           |
| Rank 10    | 95.8696     | 95.8152           |
| Rank 20    | 97.3913     | 97.4457           |

5 Conclusion

In this paper, we propose an HPM method by using the CMS values in order to achieve better performance in recognizing the hard positive samples. The results suggest that it is effective to use the proposed HPM method in MARS.

Since our experiment up to now has been done only with MARS, we will carry out a further experiment on another dataset to investigate whether the proposed HPM method is stably effective even in a different dataset. In addition, we will study to reduce the computational cost of the HPM indices much more.