Progressive Learning Algorithm for Efficient Person Re-Identification

Zhen Li[1,2], Liang Niu[3,4], Hanyang Shao[1], Nian Xue[3,4]

Shanghai Grandhonor Information Technology[1]
Nanjing University of Aeronautics and Astronautic[2]
New York University[3]
New York University Abu Dhabi[4]

ITALY 10 - 15 January 2021

Motivations

- We develop a novel learning strategy to find efficient feature embeddings while maintaining the balance of accuracy and model complexity.
- Existing triplet loss methods select only the hard identity examples which may not be optimal without considering easy examples in the triplet anchor.

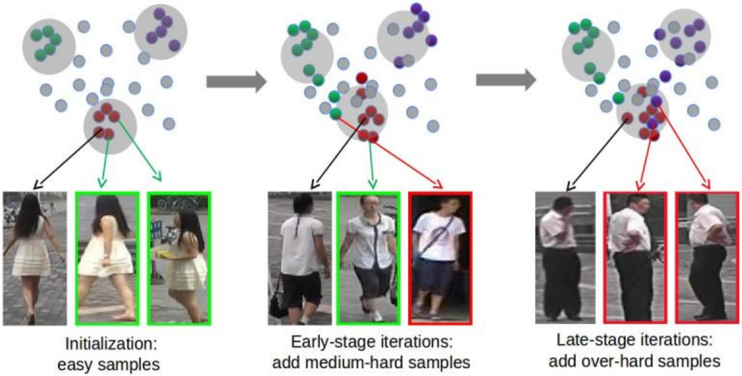

Fig 1. Training mini-batches consisting of easy, medium-hard and over-hardexamples.

Framework

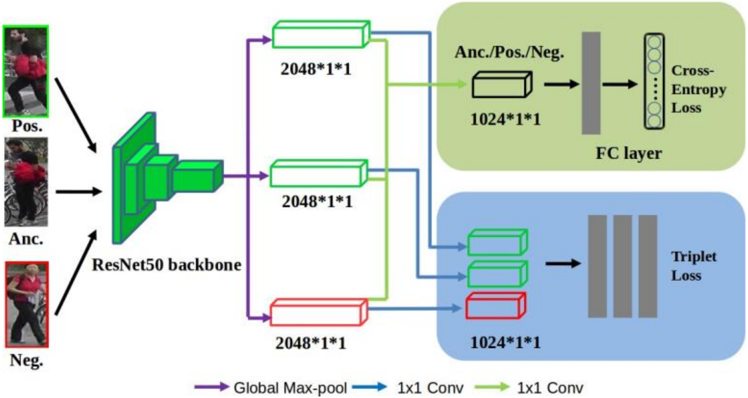

Fig. 2: Overview of the proposed ReID network architecture. “Anc.”, “Pos.” and “Neg.” represent anchor image, positive images that belong to the same identity and negative images that belong to different identities, respectively.

Problem Formulation

Main idea: Training the ReID model based on Progressive Learning Algorithm.

- **Input**: A fixed-size mini-batch consisting of $P = 16$ randomly selected identities and $K = 8$ randomly selected images per identity from the training set.
- **Output**: The optimal hyperparameter $\boldsymbol{w}^*$ along with the well trained CNN.
- Initialization: Randomly initialize N sets of hyper-parameters $\mathbb{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_N\}$ where $\boldsymbol{w}_i = (\lambda_i, m_i, k_i, p_i)$, $\lambda_i \in [0, 2]$, $m_i \in [-0.1, 0.3]$, $k_i \in [1, 8]$, $p_i \in [1, 16]$ for $i = 1, \cdots, N$.
- Loss function:

$$\mathcal{L}_{GBH}^{k,p}(\theta; X) = \sum_{l=1}^{P} \sum_{\substack{a,b \\ y_a = y_b = l}} \ln\left(1 + e^{m + T_{k,p}(a,b,n)}\right) \tag{1}$$

$$\mathcal{L}^{k,p}(\theta; X) = \mathcal{L}_{softmax}(\theta; X) + \lambda \mathcal{L}_{GBH}^{k,p}(\theta; X) \tag{2}$$

Progressive Learning Algorithm

NEW YORK UNIVERSITY

Repeat:

for each hyperparameter $i = 1$ to N **do**

Exploration: Backpropagate CNN in 20 epochs and evaluate the loss $\mathcal{L}$ according to Eq. 1 and Eq. 2, and evaluate the Bayesian objective $f(\boldsymbol{w}_i)$.

Restoration: CNN weights are restored to that before 20 epochs of exploration.

end for

Exploitation: Based on $f(\mathbb{W})$, obtain a new improved candidate $\boldsymbol{w}^{'}$ and update Gaussian process according to Eq. 3 and Eq. 4, and add $\boldsymbol{w}^{'}$ to $\mathbb{W}$, and update $\hat{\boldsymbol{w}}$ as well;

Backpropagate to update CNN weights for 300 epochs based on the new hyperparameter $\boldsymbol{w}$ and the feed-forward loss $\mathcal{L}$;

Save the model with lowest loss $\mathcal{L}$ for the current hyperparameter $\boldsymbol{w}$;

Until maximum epochs ($M = 3,000$) reached

Bayesian Algorithm

For N sets of such parameters we denote $\mathbb{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_N\}$, and the corresponding $f(\mathbb{W})$, the posterior belief of f at a new candidate $\boldsymbol{w}$ is given by

$$\begin{cases} \tilde{f}(\boldsymbol{w}) \sim \mathcal{GP}(\mu(\hat{\boldsymbol{w}}) + \triangle\mu, \mathcal{K}(\boldsymbol{w}) - \triangle\mathcal{K}) \\ \triangle\mu = \mathcal{K}(\boldsymbol{w}, \mathbb{W})\,\mathcal{K}(\mathbb{W})^{-1}(f(\mathbb{W}) - \mu(\mathbb{W})) \\ \triangle\mathcal{K} = \mathcal{K}(\boldsymbol{w}, \mathbb{W})\,\mathcal{K}(\mathbb{W})^{-1}\,\mathcal{K}(\mathbb{W}, \boldsymbol{w}) \end{cases} \tag{3}$$

The expected improvement of a candidate $\hat{\boldsymbol{w}}$ is defined as

$$\mathcal{EI}(\boldsymbol{w}) = (\mathcal{K}(\boldsymbol{w}) - \triangle\mathcal{K})^{\frac{1}{2}}(Z\Phi(Z) + \varphi(Z)) \tag{4}$$

Bayesian optimization minimizes the following objective function:

$$f(\boldsymbol{w}) = \left|\frac{\bar{\mathcal{L}}_t^{k,p}(\theta; \boldsymbol{w}; X) - \bar{\mathcal{L}}_{t'}^{k,p}(\theta; \boldsymbol{w}; X)}{\bar{\mathcal{L}}_t^{k,p}(\theta; \boldsymbol{w}; X)} - \mathcal{ED}\right| \tag{5}$$

Datasets

NEW YORK UNIVERSITY

We perform all the experiments on three commonly used benchmarks: Market-1501, DukeMTMC-ReID (briefed as DukeMTMC), and CUHK03(D) & CUHK03(L) datasets. These ReID datasets are summarized in Table 1.

Table 1: ReID Benchmark datasets used in our experiments.

Dataset	Market1501	DukeMTMC	CUHK03(D/L)
Identities	1,501	1,812	1,360
Bboxes	32,668	36,411	13,164
Camera	6	8	6
Train images	12,936	16,522	7,365/7,368
Train ids	751	702	767
Query images	3,368	2,228	1,400
Query ids	750	702	700
Gallery images	19,732	17,661	5,332

Precision Performance

Category	Methods	Market1501(SQ)		Market1501(MQ)		CUHK03(D)		CUHK03(L)		DukeMTMC	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
part	HA-CNN[7]	75.7	91.2	82.8	93.8	38.6	41.7	41.0	44.4	63.8	80.5
	Deep-Person[36]	79.6	92.3	85.1	94.5	-	-	-	-	64.8	80.9
	PCB[1]	77.4	92.3	-	-	54.2	61.3	-	-	66.1	81.7
	PCB+RPP[1]	81.6	93.8	-	-	57.5	63.7	-	-	69.2	83.3
	Aligned-ReID[35]	82.3	92.6	-	-	-	-	-	-	-	-
	MGN[2]	**86.9**	**95.7**	**90.7**	**96.9**	**66.0**	**66.8**	**67.4**	**68.0**	**78.4**	**88.7**
global	SVDNet[5]	62.1	82.3	-	-	37.2	41.5	37.8	40.9	56.8	76.7
	TriNet[6]	69.1	84.9	76.4	90.5	-	-	-	-	-	-
	GP-reid[24]	81.2	92.2	82.8	93.8	-	-	-	-	**72.8**	**85.2**
	DaRe[9]	74.2	88.5	-	-	58.1	61.6	60.2	64.5	63.0	79.1
	PLA	**83.6**	**93.7**	**88.4**	**95.2**	**63.2**	**67.2**	**67.5**	**71.5**	72.5	84.3
RK	Trinet [6]	81.1	86.7	87.2	91.8	-	-	-	-	-	-
	DaRe [9]	85.9	90.8	-	-	71.2	69.8	73.7	72.9	79.6	84.4
	MGN [2]	**94.2**	**96.6**	**95.9**	**97.1**	-	-	-	-	-	-
	PLA	89.4	94.7	92.9	95.7	**77.2**	**75.5**	**81.0**	**79.6**	**80.1**	**87.0**

Table 2 : Comparing PLA with different global models on all datasets. “RK” stands for reranking.

Computation and Memory Cost

NEW YORK UNIVERSITY

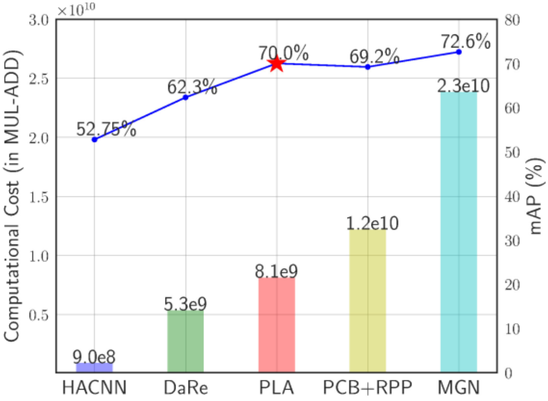

Fig. 3 a)

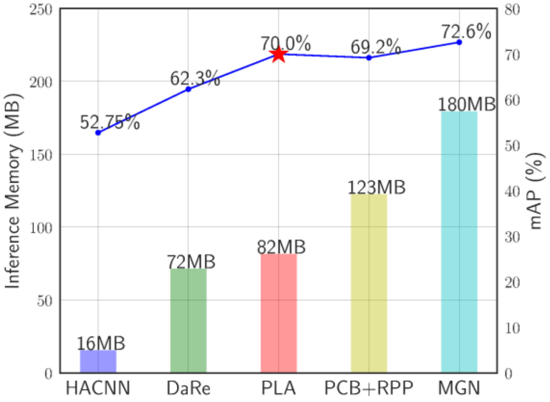

Fig. 3 b)

a) Accuracy vs. computation cost (number of Mul-Add); b) Accuracy vs. inference memory (MB). Accuracy is reported as the average mAP on all datasets.

Conclusions

NEW YORK UNIVERSITY

- A novel learning approach is proposed to find efficient feature embeddings while maintaining the balance of accuracy and model complexity.
- A novel method is developed to explore the hard examples and build a discriminant feature embedding yet compact enough for large-scale applications.
- A novel Bayesian approach is employed to progressively learn the triplet loss from simple to hard samples.
- The developed reid system is efficient in both computation and memory, rendering it a commercial system.

References I

- Y. Sun, L. Zheng, W. Deng, and S. Wang, “Svdnet for pedestrian retrieval,” in *IEEE International Conference on Computer Vision*, 2017, pp. 3820–3828.
- A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*, 2017.
- J. Almazan, B. Gajic, N. Murray, and D. Larlus, “Re-id done right: towards good practices for person re-identification,” *arXiv preprint arXiv:1801.05339*, 2018.
- Y. Wang, L. Wang, Y. You, X. Zou, V. Chen, S. Li, G. Huang, B. Hariharan, and K. Q. Weinberger, “Resource aware person re-identification across multiple resolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8042–8051.

References II

- W. Li, X. Zhu, and S. Gong, “Harmonious attention network for person re-identification,” in *Computer Vision and Pattern Recognition*, vol. 1, 2018, p. 2.
- X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, “Deep-person: Learning discriminative deep features for person re-identification,” *arXiv preprint arXiv:1711.10658*, 2017.
- Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond part models: Person retrieval with refined part pooling,” *arXiv preprint arXiv:1711.09349*, 2017.
- X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun, “Alignedreid: Surpassing human-level performance in person re-identification,” *arXiv preprint arXiv:1711.08184*, 2017.

References III

G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, “Learning discriminative features with multiple granularities for person re-identification,” in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 274–282.