Learning to Learn in a Semi-Supervised Fashion Supplementary Material

Anonymous ECCV submission

Paper ID 3007

Overview

In this supplementary document, we present additional details and results to complement the main paper.

- We describe the details of the proposed meta-learning algorithm.
- We include more experimental results and analysis.

1 Details of the Proposed Method

We present the pseudo code of our proposed method in this section.

1.1 Meta-Learning on X_L

The pseudo code of meta-learning on X_L (i.e., the first stage of the proposed meta-learning algorithm) is summarized in Algorithm 1.

Following an existing episodic learning paradigm [2, 5], in each episode we partition the labeled set X_L into a meta-training set M_T with $\lfloor \frac{C_L}{2} \rfloor$ classes (i.e., $C_{M_T} = \lfloor \frac{C_L}{2} \rfloor$) and a meta-validation set M_V with the remaining $C_L - \lfloor \frac{C_L}{2} \rfloor$ classes (i.e., $C_{M_V} = C_L - \lfloor \frac{C_L}{2} \rfloor$) based on the ground truth labels so that the meta-training set and meta-validation set share disjoint ground truth label sets.

1.2 Meta Semi-Supervised Learning on X_L and X_U

The pseudo code of meta semi-supervised learning on X_L and X_U (i.e., the second stage of the proposed meta-learning algorithm) is summarized in Algorithm 2.

Algorithm 1: Meta-Learning on X_L

Input: Labeled images X_L, label set Y_L, and model F

for *each episode* **do**
  M_T, $M_V \leftarrow$ DataSplit(X_L)
  M_T and M_V share non-overlapping category labels.
  for *each batch in* M_V **do**
    Sample one reference image from each class in M_T.
    Extract feature for each sampled reference image and each image in the current batch.
    Compute semantics-oriented similarity representation s.
    Update the model by minimizing Equations (1) and (3).
  end
end

Algorithm 2: Meta Semi-Supervised Learning on X_L and X_U

Input: Labeled images X_L, label set Y_L, unlabeled images X_U, and model F

for *each iteration* **do**
  for *each batch in* X_U **do**
    Sample C_{M_T} classes in X_L.
    For each sampled class, sample one reference image.
    Compute semantics-oriented similarity representation s for this batch of images in X_U.
    Determine positive and negative pairs based on Equation (4).
    Update the model by minimizing $\mathcal{L}_{\text{feat}-\text{contrast}}(X_U; F)$ based on the assigned positive/negative pair labels.
  end
  for *each batch in* X_L **do**
    Update the model by minimizing $\mathcal{L}_{\text{feat}-\text{contrast}}(X_L; F)$ based on the label set Y_L.
  end
end

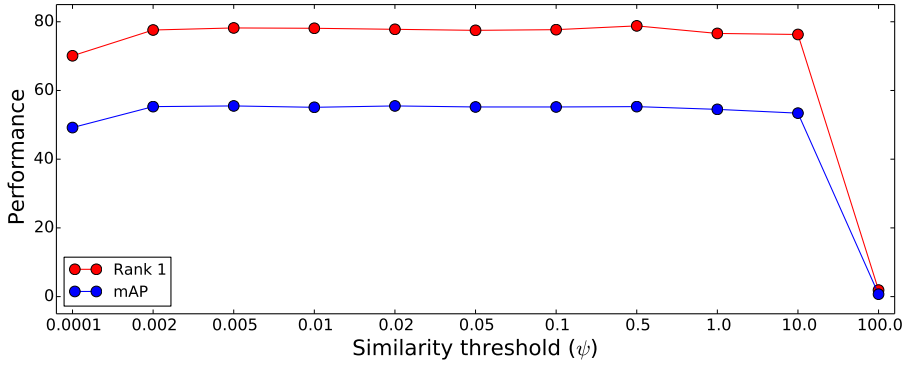

Fig. 1: **Sensitivity analysis of the similarity threshold ψ.** The performance of our approach remains stable when the similarity threshold ψ is within a reasonable range.

2 Experimental Results and Analysis

2.1 Sensitivity analysis on the similarity threshold ψ

To analyze the sensitivity of our algorithm with respect to the similarity threshold ψ which is used to determine whether or not a given image pair is of the same class, we conduct a sensitivity analysis by varying the value of ψ.

Figure 1 presents the experimental results on the Market-1501 dataset [20]. When the similarity threshold ψ is set to a very small value, e.g., 1×10^{-4}, almost all of the image pairs in the unlabeled set are recognized as negative pairs, since the distance between the semantics-oriented similarity representations of an image pair from the unlabeled set should be less than the threshold ψ in order to be considered as a positive pair. Similarly, when the similarity threshold ψ is set to a large value, e.g., 100, almost all the image pairs in the unlabeled set are recognized as positive pairs. Our results suffer from significant performance drops on both rank 1 and mAP. When the similarity threshold ψ is set within a reasonable range, the performance is improved and not sensitive to its value.

We note that as shown in Figure 3 of the main manuscript, our learned semantics-oriented similarity representations are well-separated with respect to the ID labels. This explains why a wide range of ψ would be applicable in determining positive and negative image pairs.

2.2 Results of using different pre-trained models

Since our method can be applied to fully supervised settings, we conduct an experiment by adopting different pre-trained models and report the results with comparisons to the state-of-the-art approaches. We initialize our model from PyrNet [9], SPReID [15], and DMML [3], respectively, and apply our proposed algorithm on the entire training set. We denote the three variant methods as: (1) PyrNet [9] + Ours, (2) SPReID [15] + Ours, and (3) DMML [3] + Ours,

Table 1: **Results of fully-supervised person re-ID.** The bold and underlined numbers indicate the top two results, respectively.

Method	Market-1501 [20]			DukeMTMC-reID [10]		
	Rank 1	Rank 5	mAP	Rank 1	Rank 5	mAP
Part-Aligned [19]	81.0	92.0	63.4	-	-	-
SVDNet [16]	82.3	92.3	62.1	76.7	86.4	56.8
PAN [21]	82.8	93.5	63.4	71.6	83.9	51.5
MGCAM [14]	83.8	-	74.3	-	-	-
TriNet [6]	84.9	94.2	69.1	-	-	-
JLML [7]	85.1	-	65.5	-	-	-
PoseTransfer [22]	87.7	-	68.9	78.5	-	56.9
PSE [11]	87.7	94.5	69.0	79.8	89.7	62.0
CamStyle [22]	88.1	-	68.7	75.3	-	53.5
DPFL [4]	88.9	92.3	73.1	79.2	-	60.6
AlignedReID [17]	89.2	95.9	72.8	79.3	89.7	65.6
DML [18]	89.3	-	70.5	-	-	-
DKP [12]	90.1	96.7	75.3	80.3	89.5	63.2
DuATM [13]	91.4	97.1	76.6	81.8	90.2	68.6
RDR [1]	92.2	97.9	81.2	85.2	93.9	72.8
PyrNet [9]	93.6	98.2	81.7	<u>87.1</u>	<u>94.1</u>	74.0
SPReID [15]	93.7	97.6	83.4	85.9	92.9	73.3
DMML [3]	93.5	97.6	81.6	85.9	-	73.7
BoT [8]	<u>94.5</u>	98.2	<u>85.9</u>	86.4	93.6	<u>76.4</u>
AlignedReID [17] + Ours	91.1	96.3	78.1	81.7	91.0	67.7
DMML [3] + Ours	94.1	97.9	82.2	86.6	93.9	74.6
SPReID [15] + Ours	94.2	<u>98.4</u>	84.2	<u>87.1</u>	93.8	74.5
PyrNet [9] + Ours	94.2	**98.5**	82.1	**88.0**	**94.8**	74.7
BoT [8] + Ours	**94.8**	98.3	**86.1**	86.6	93.9	**76.8**

respectively. Table 1 presents the experimental results on the Market-1501 [20] and DukeMTMC-reID [10] datasets.

From the results, we observe that applying our method consistently improves the performance of PyrNet [9], SPReID [15], and DMML [3], achieving the state-of-the-art performance on both datasets.

References

1. Almazan, J., Gajic, B., Murray, N., Larlus, D.: Re-id done right: towards good practices for person re-identification. arXiv (2018)
2. Balaji, Y., Sankaranarayanan, S., Chellappa, R.: Metareg: Towards domain generalization using meta-regularization. In: NeurIPS (2018)
3. Chen, G., Zhang, T., Lu, J., Zhou, J.: Deep meta metric learning. In: ICCV (2019)
4. 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: CVPR (2016)
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML (2017)
6. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv (2017)
7. Li, W., Zhu, X., Gong, S.: Person re-identification by deep joint learning of multi-loss classification. IJCAI (2017)
8. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPRW (2019)
9. Martinel, N., Luca Foresti, G., Micheloni, C.: Aggregating deep pyramidal representations for person re-identification. In: CVPRW (2019)
10. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: ECCVW (2016)
11. Saquib Sarfraz, M., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: CVPR (2018)
12. Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X.: End-to-end deep kronecker-product matching for person re-identification. In: CVPR (2018)
13. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: CVPR (2018)
14. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: CVPR (2018)
15. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: CVPR (2018)
16. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval. In: CVPR (2017)
17. Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q., Jiang, W., Zhang, C., Sun, J.: Alignedreid: Surpassing human-level performance in person re-identification. arXiv (2017)
18. Zhang, Y., Xiang, T., Hospedales, T.M., Lu, H.: Deep mutual learning. In: CVPR (2018)
19. Zhao, L., Li, X., Zhuang, Y., Wang, J.: Deeply-learned part-aligned representations for person re-identification. In: ICCV (2017)
20. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV (2015)
21. Zheng, Z., Zheng, L., Yang, Y.: Pedestrian alignment network for large-scale person re-identification. TCSVT (2018)
22. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: CVPR (2018)