Supplementary Material: Holistic Guidance for Occluded Person Re-Identification

Madhu Kiran[1]
madhu_sajc@hotmail.com
R Gnana Praveen[1]
praveengnan.24@gmail.com
Le Thanh Nguyen-Meidine[1]
le-thanh.nguyen-meidine.1@ens.etsmtl.ca
Soufiane Belharbi[1]
soufiane.belharbi.1@ens.etsmtl.ca
Louis-Antoine Blais-Morin[2]
lablaismorin@GENETEC.COM
Eric Granger[1]
Eric.Granger@etsmtl.ca

[1] LIVIA, Dept. of Systems Engineering
École de technologie supérieure
Montreal, Canada

[2] Genetec Inc.
Montreal, Canada

Abstract

This supplementary material complements our main paper with further results by providing additional details on our training protocol, description of datasets, ablation study on λ and α parameter(s),batch size vs result analysis, partial-ReID dataset, analysis of the effects of occlusion using different test datasets, our qualitative (visual) results, and impact of within/between class overlap in different data. **Our GitHub code is available** [1].

1 Additional Details on Experimental Setting:

The decoder network used in our experiments consists of a set of three transposed convolution layers, each one followed by Batchnorm and relu layers with kernel sizes 5, 3 and 3. The last convolution layer alone has a sigmoid activation function. We use a validation set during the training of Teacher Student model for early stopping. In addition, while using the holistic dataset with the teacher model, we use a pre-trained holistic network but early stopped during pre-training to allow some overlap between DCDs. This makes the separation between the DCDs overlap a bit at the start of teacher-student training, acting as a soft label for the student model. This simultaneous learning is expected to increase the model robustness to occlusion.

[1] https://anonymous.4open.science/r/HolisitcGuidanceOccReID2-7052/

2 Descriptions of Other Datasets:

Holistic ReID: We refer to Market-1501 [9] dataset and Duke-MTMC [7] dataset as holistic datasets or regular Person Re-ID datasets. Holistic datasets have a small percentage of occluded images which is negligible compared to the size of the database. Market1501 dataset has its bounding boxes labelled by object detector with a total of 1501 identities. It provides a realistic benchmark provided by a total of 6 cameras. Duke-MTMC dataset has been build from videos captured by 8 cameras at Duke University campus. It has a total of 1812 identities.

Partial-ReID: Partial-ILIDS [6, 10] and Partial-ReiD dataset [11]. Parial-ILIDS is simulated partial dataset. Partial-ReID dataset is specially used in partial person ReID, Each identity has five partial and five full body images. The images were collected at a university campus. For both Occluded ReID and Partial-ReID datasets, we show two sets of results. One by using only Holistic dataset in Teacher model and augmented Holistic dataset on Student model with random erasing as described above. In the second set of experiment we train the Teacher on holistic dataset (Market1501) and student on occluded samples from Occluded-Duke dataset.

3 Impact of λ and α:

The λ parameters, i.e., the weights of the loss function in Eqn. 7 and 8 play an important role in balancing the contribution of different components of the loss functions during optimization. λ(s) in Eqn. 8 balances the effects of classification loss and auto-encoder reconstruction loss. Since our main task is classification and discriminative capability of the deep CNN, we keep the weights of the auto-encoder loss small. Fig. 3 shows the variation of Rank-1 accuracy on Occluded-Duke dataset with respect to the variation of the parameter λ. We evaluate the effect on two different scenarios. One is the "lower-bound" scenario where we train our network on the Market1501 dataset without any tuning on the occlusion dataset and test on Occluded-Duke dataset. But this training also included the Joint loss function, i.e. reconstruction loss, along with classification loss. We can observe that a λ of 0.01 has improved the lower bound result compared to lower values of lambda i.e. denoising auto-encoder had added some ability to recover from occlusion. In the second setting –end-to-end setting, i.e. the whole system with teacher Student model, Teacher is trained on Market1501 dataset and Student is trained on Occluded-Duke dataset, and the test is performed on Occluded-Duke test dataset. Once again λ plays a role in contributing to the overall results by improving the model's capacity to be invariant to occlusions. A λ greater than 0.01 brings down the overall Rank-1 accuracy in both the cases. This could be due to the fact that the system starts to learn more of generative property compared to the discriminative property.

On the weights of components in distribution learning. From the paper, Eqn. 7 is: $\mathcal{L}_D = \lambda_{\text{wc}}\mathcal{L}_D^{\text{wc}} + \lambda_{\text{bc}}\mathcal{L}_D^{\text{bc}} + \lambda_{\text{g}}\mathcal{L}_{\text{global}}$. The class distance distribution is affected by both $\lambda_{\text{wc}}, \lambda_{\text{bc}}$, but λ_{g} controls the amount of data from holistic dataset to be retained in the student model. From Fig. 3 (results obtained by training the teacher on Market1501, and student on Occluded-Duke data), a lower value of λ_{bc} affects the performance severely compared to a lower value of λ_{wc} because λ_{bc} is responsible for emphasizing the between-class distance, which is the main challenge in occluded ReID. Hence λ_{bc} has a comparatively larger influence. It can also be seen that a sub-optimal λ_{g} can overfit the model on either occluded or holistic ReID datasets, and lead to poor performance.

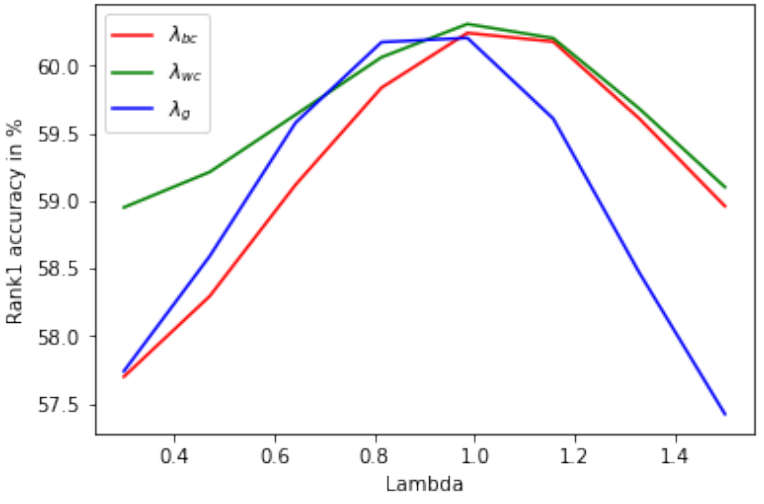


Figure 1: Impact on accuracy of $\lambda_{wc}, \lambda_{bc}$ and λ_g in loss function.

4 Batch Size vs Accuracy:

In order to analyse the effect of batch size in distilling the knowledge from teacher to student, we perform an experiment where we train different model with different batch sizes and observe its effect on the overall result. We can observe that a larger batch size shows better overall performance in terms of Rank-1 due to the fact that we depend on matching distribution of distances across the teacher and student and a larger batch size tends to represent the respective dataset better in-terms of overall distribution.

5 Results on Partial-ReID Dataset:

In Tab. 1 we compare with existing methods including MTRC [5], AMC+SWM [11]. It is worth mentioning that these Partial-REID methods use hand-cropped probe images. We achieve a better performance without the use of hand-cropped images. Other methods in the Table are more designed for occluded-ReID often using pose supervision additionally as described in the previous subsection. Similar to the evaluation on Occluded ReID datasets, we conduct two experiments one with the student trained on artificially occluded Market1501 data-HG(Unsup) and other experiment where the student was trained on occluded examples from Occluded Duke dataset-HG(sup). The results once again show that although the HG(Unsup) outperforms many SOA, the HG(Sup) though is trained by occluded examples from totally different dataset from the test set has still improved the overall performance by a good margin.

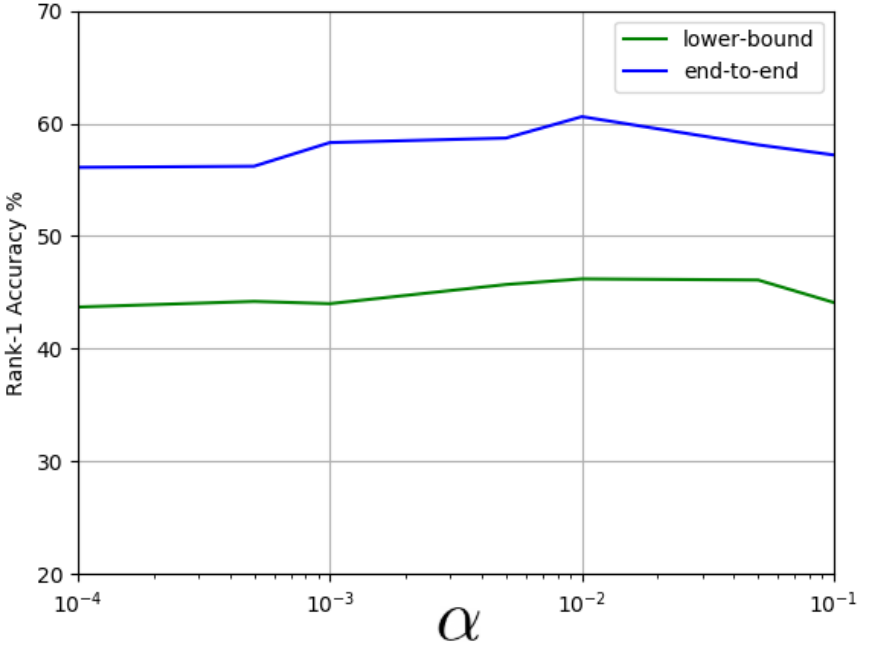


Figure 2: An ablation study on the α parameter used in the joint loss function. It can be observed that lambda must be small (much less than 1) as it reflects on the discriminative CNN learning. In the lower-bound setting, the CNN backbone alone was trained on holistic data with reconstruction and classification losses, and tested directly on occluded data. In the end-to-end setting, the full ReID system with student-teacher model was employed.

6 Visual Results:

Fig. 4 shows the comparison of activation maps from the student model tested on examples with Partial-ReID dataset and activation maps of [6]. [6] use pose maps for attention and hence dependent on appearance to obtain a good quality pose map. Therefore in the last example of Fig. 4 with both head and legs occluded, [6] show high activation for occluded region. But from the activation maps of our proposed HG method, it can be seen that our method is good at localising non-occluded regions. This is because our method does not depend on pose maps and was trained by using guidance from holistic data.

7 Impact of Occlusion on Within/Between-Class Distance Distributions:

In general, deep CNN features of the same person tend to have a more similar appearance – for example in the Euclidean space – than features of different people. In other words, within-class distances tend to be smaller than those of between-class distances. Our proposal

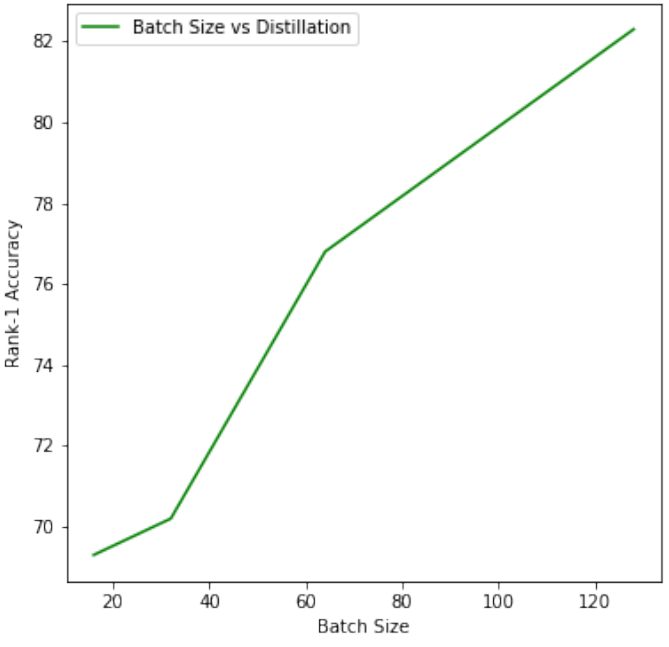


Figure 3: A study to analyse the effect of batch size in distilling knowledge from the Teacher to the Student and hence the overall performance in Rank-1. This experiment was performed on Occluded Re-ID dataset.

stands on the assumption that identities of occluded persons often tend to have some overlap with those of other classes due to feature corruption.

In order to asses the impact of occlusion within our test dataset, i.e., Occluded ReID [13], Occluded Duke-MTMC [6], Partial Re-ID [2, 4], and Partial-ILIDS [2, 4], we propose to visualize the within/between-class distances of these datasets. To provide a reference point for comparing the difficulty levels, we visualize the class distance distribution of our pre-trained baseline network [12] on the Market1501 dataset, and compute the Euclidean distances between features extracted from the Market and the test datasets mentioned above. Fig. 5 shows that within-class distributions overlap considerably with between-class distribution. Based on this empirical evidence, we therefore conclude that occlusions contribute to corrupt features, and lead to a greater overlap between feature representations from different classes.

Table 1: Accuracy of our proposed compared with state-of-the-art methods on the Partial-ReID datasets.

Method	Partial-ILIDS		Partial-ReID		Supervision
	Rank-1	Rank-3	Rank-1	Rank-3	
MTRC [5]	17.7	26.1	23.7	27.3	None
AMC+SWM [11]	21	32.8	37.7	46	None
DSR [2]	58.8	67.2	50.7	70	None
SFR [3]	63.9	74.8	56.9	78.5	None
PGFA [6]	69.1	80.9	68	80	Pose Maps
PVPM [1]	-	-	78.3	-	Pose Maps
FPR [4]	68.1	-	81	-	None
Teacher-S [14]	-	-	82.7	-	Occluded Data
HOREID [8]	72.6	86.4	85.3	91	Key Points
Ours(HG-Unsup)	71.1	85.8	82.3	87.2	None
Ours(HG-Sup)	**74.8**	**87.3**	**86.1**	**91.8**	Occluded Data

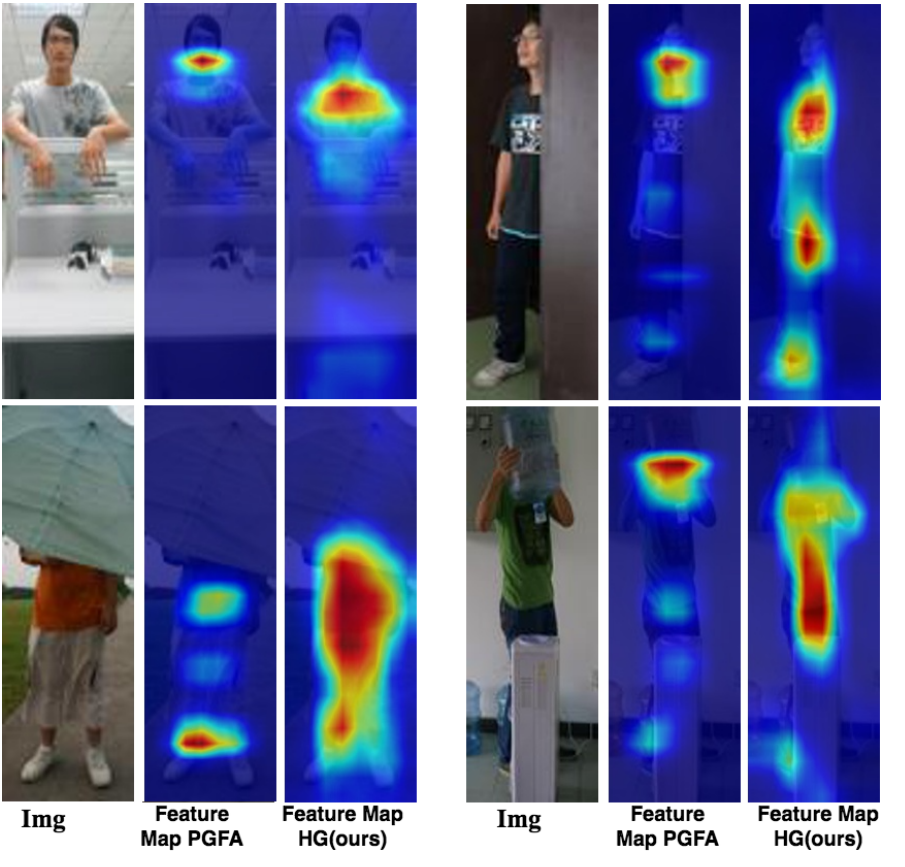
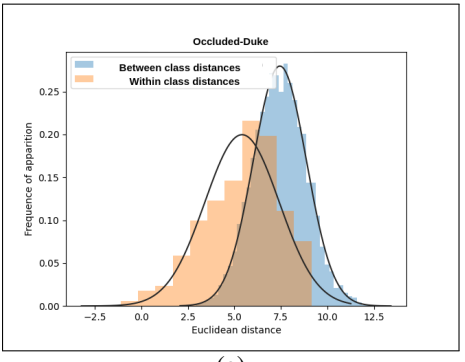
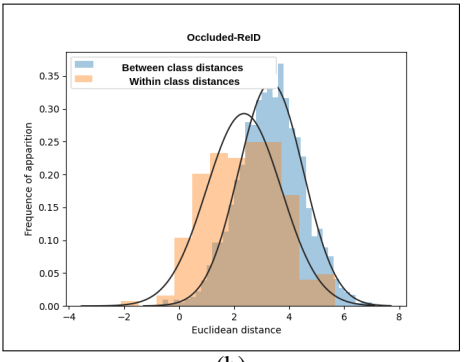


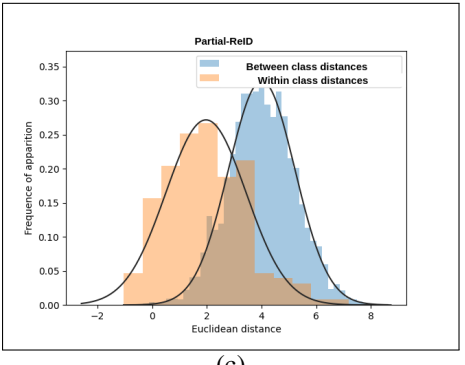
Figure 4: We compare the feature maps generated by one of the SOA(middle) [6] with that of our HG method. [6] uses pose estimation additionally.



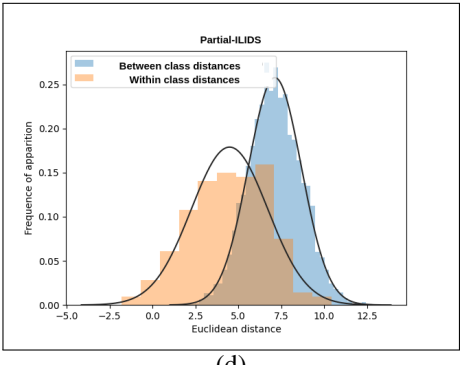
(a)



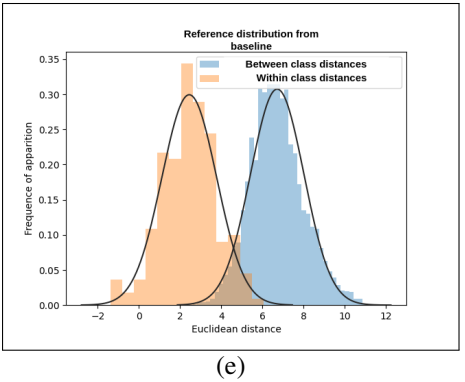
(b)



(c)



(d)



(e)

Figure 5: A visual comparison between different datasets to asses their level of difficulty of occlusion. **(e)** shows the reference distribution for the holistic dataset (Market1501 dataset). It can be observed that the reference dataset has a better separation between within and between class distributions compared to occluded and partial dataset. We hypothesise that the overlap of distributions of the later could be attributed to the corrupted features failing to represent a given class accurately.

References

[1] S. Gao, J. Wang, H. Lu, and Z. Liu. Pose-guided visible part matching for occluded person reid. In CVPR, 2020.

[2] L. He, J. Liang, H. Li, and Z. Sun. Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In CVPR, 2018.

[3] L. He, Z. Sun, Y. Zhu, and Y. Wang. Recognizing partial biometric patterns. CoRR, abs/1810.07399, 2018.

[4] L. He, Y. Wang, W. Liu, H. Zhao, Z. Sun, and J. Feng. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In ICCV, 2019.

[5] S. Liao, A.K. Jain, and S.Z. Li. Partial face recognition: Alignment-free approach. PAMI, 35(5):1193–1205, 2012.

[6] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang. Pose-guided feature alignment for occluded person re-identification. In ICCV, 2019.

[7] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In ECCV workshop, 2016.

[8] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, and J. Sun. High-order information matters: Learning relation and topology for occluded person re-identification. In CVPR, 2020.

[9] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In ICCV, 2015.

[10] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In CVPR, 2011.

[11] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong. Partial person re-identification. In ICCV, 2015.

[12] W. Zhong, L. Jiang, T. Zhang, J. Ji, and H. Xiong. A part-based attention network for person re-identification. Multimedia Tools and Applications, 79, 2020.

[13] J. Zhuo, Z. Chen, J. Lai, and G. Wang. Occluded person re-identification. In ICME, 2018.

[14] J. Zhuo, J. Lai, and P. Chen. A novel teacher-student learning framework for occluded person re-identification. CoRR, abs/1907.03253, 2019.