Occlude Them All: Occlusion-Aware Mask Network for Person Re-identification

	Occluded-Duke		Partial-REID		Partial-iLIDS	
	Rank-1	*m*AP	Rank-1	*m*AP	Rank-1	*m*AP
Classifier (OC)	55.6	41.2	82.3	76.1	73.1	76.0
Grader (OG)	**62.6**	**46.1**	**86.0**	**77.4**	**77.3**	**79.5**

Table 1: Comparing performance of different methods for identifying occlusion types.

images						
Classifier (OC)	1/4 right	holistic	holistic	holistic	holistic	holistic
Grader (OG)	holistic	1/4 down	1/4 down	1/2 top	1/4 top	1/2 down

Table 2: Case studies of different methods for identifying occlusion types.

	Occluded-Duke		Partial-REID		Partial-iLIDS	
	Rank-1	*m*AP	Rank-1	*m*AP	Rank-1	*m*AP
layer-3	**62.6**	**46.1**	**86.0**	**77.4**	**77.3**	**79.5**
layer-4	60.2	45.6	83.7	77.7	70.6	74.1

Table 3: Comparing performance when feeding different features to the occlusion grader.

Model	Params	Flops	Time-1 (s)	Time-119 (s)
HOReID [2]	161.05×10^6	15.08×10^9	9.83	23.20
PGFA [1]	85.72×10^6	22.13×10^9	6.34	12.71
Ours	$\mathbf{25.46 \times 10^6}$	$\mathbf{2.74 \times 10^9}$	**4.56**	**6.33**

Table 4: Comparing Params, Flops and Time of different occluded person ReID methods. Params is the model's number of parameters, Flops is the model's number of floating point operations, Time indicates the model's time efficiency of retrieving images.

1. Methods for Identifying Occlusion Types

In the main paper, we propose an occlusion grader (OG) to identify the occlusion's type. In this section, we quantitatively and qualitative show why the more straightforward occlusion classifier (OC) is not used. In specific, the classifier incorporates a nine dimensional FC layer for classification and the cross-entropy loss for constraint, defined as:

$$\mathcal{L}_{\mathrm{OC}} = \ell_{\mathrm{CE}}(T(f_i), t_i), \tag{1}$$

where f_i and t_i denote the input feature and its known occlusion type, respectively. In contrast, the proposed OG adopts a threshold method to avoid overfitting the limited occlusion types in OC (detailed in section 3.4).

Quantitative results are reported in Table 1. As we can observe, the proposed threshold-based grader shows a clear improvement to the classifier. Such results support the design of our approach. To explain, real-world occlusions are typically more complicated than the eight types we have used for augmentation. Therefore, directly learning to classify such types can overfit to these types, hence cannot precisely identify real-world occlusion types.

Qualitative results are reported in Table 2. In specific, we compare identified occlusion types from the above two methods. We elaborate on the six case studies sequentially as following. (1) OC incorrectly identifies the pedestrian in dark clothes as right-occluded, but OG is accuracy. (2) OC incorrectly reports a holistic image, failing to disambiguate the dark clothes and dark baggage. (3) While this picture does not include bottom-parts, the proposed OG can still recognize such an occlusion. Lastly, (4) to (6) demonstrate other cases where OC misclassifies occluded pedestrians to holistic ones.

2. Input Features of the Occlusion Grader

In this section, we compare the performance when feeding different intermediate features to the occlusion grader. We report the results in Table 3. As we can observe, feeding the masked feature g (layer-3) obtains better results than feeding the final output feature h (layer-4). We suspect the reason as final features contain (1) more high-dimensional and abstract semantic information, and (2) less contour and positional information. Since the latter information is more effective in determining the occlusion's location and area, using intermediate layers can achieve better performance.

3. Computational Efficiency

In this section, we demonstrate the efficiency of our approach. In particular, we examine three metrics, including (1) the model's number of parameters (Params), (2) the model's number of floating point operations (Flops), and (3) the time cost of identifying 1 and 119 probe images from 119 gallery images. The results are reported in Table 4. As we can observe, the proposed OAMN only needs 25.46×10^6 parameters and 2.74×10^9 Flops. Comparing with PGFA, we reduce the cost by 60.26×10^6 parameters and 19.39×10^9 Flops. We obtain more reduction of 135.59×10^5 parameters and 12.34×10^9 Flops when comparing to HOReID. Given 119 gallery images, our method only needs 4.56s and 6.33s to process 1 and 119 queries, whereas HOReID needs 9.83s and 23.20s, and PGFA needs 6.34s and 12.71s. Regarding the time cost of retrieval, we average the time over 5 experiments. Results show that our OAMN is 55% faster than HOReID and 1.7820s faster than PGFA. In summary, the results demonstrate that the proposed OAMN is more effective and efficient than state-of-the-art approaches.

References

[1] Jiaxu Miao, Yu Wu, Ping Liu, Yuhang Ding, and Yi Yang. Pose-guided feature alignment for occluded person re-identification. In *ICCV*, pages 542–551, 2019. 1

[2] Guan'an Wang, Shuo Yang, Huanyu Liu, Zhicheng Wang, Yang Yang, Shuliang Wang, Gang Yu, Erjin Zhou, and Jian Sun. High-order information matters: Learning relation and topology for occluded person re-identification. In *CVPR*, pages 6448–6457, 2020. 1