Robust Re-Identification by Multiple Views Knowledge Distillation - Supplementary Material

Angelo Porrello[0000−0002−9022−8484], Luca Bergamini[0000−0003−1221−8640], Simone Calderara[0000−0001−9056−1538]

AImageLab, University of Modena and Reggio Emilia
{angelo.porrello, luca.bergamini24, simone.calderara}@unimore.it

Distilling viewpoints *vs* time: impact on camera bias. As discussed in the main paper (Introduction and Sec. 4.4), limiting the teacher-student transfer to the temporal axis does not explicitly encourage invariance and robustness to different viewpoints. To further prove such a claim, we again measure the camera bias lying in high-level features, in the same manner as described in Sec. 4.4 of the main paper. This time, though, we focus on a student accessing fewer frames from the same tracklet, thus being educated to capture time information solely. Table 1 compares this strategy (third row) with our proposal (fourth row), which instead forces the transfer at viewpoint level. As expected: *i)* time-based distillation performs similarly to the teacher, confirming its poor ability to confer robustness to shifts in background appearance; *ii)* as advocated by our work, a student shows a lower camera bias when trained on different viewpoints instead of using temporal information only.

Student explanation - other examples. In Sec. 4.4 of the main paper, we investigate which regions the student focuses on, showing that it pays higher attention to foreground details when compared to its teacher. We observe that this happens systematically, especially when dealing with person Re-ID. Figure 1 reports additional comparisons between the explanations provided by the teacher and its student on Duke-Video-ReID [1].

Errors Analysis We provide here some visual examples of the errors of our method and try to investigate their nature. With reference to the Video-To-Video setting on MARS [2], our model (ResVKD-50) misidentifies 223 out of

Table 1. Analysis on camera bias – in terms of viewpoint classification accuracy – for different methods. We indicate with "ResTKD-50" a student restricted to time information solely.

	MARS	Duke
Prior Class.	0.19	0.14
ResNet-50 (teacher)	**0.74**	**0.76**
ResTKD-50 (time-based distillation)	0.69	**0.76**
ResVKD-50 (viewpoints-based distillation)	0.49	0.69

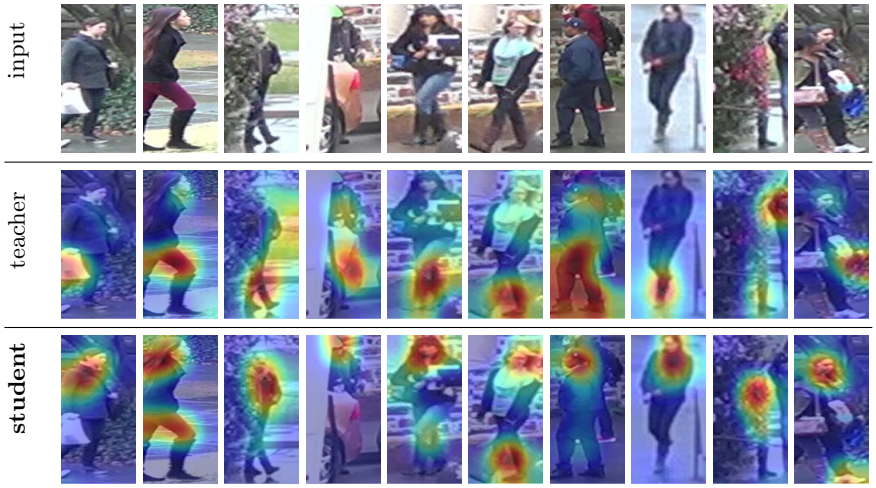
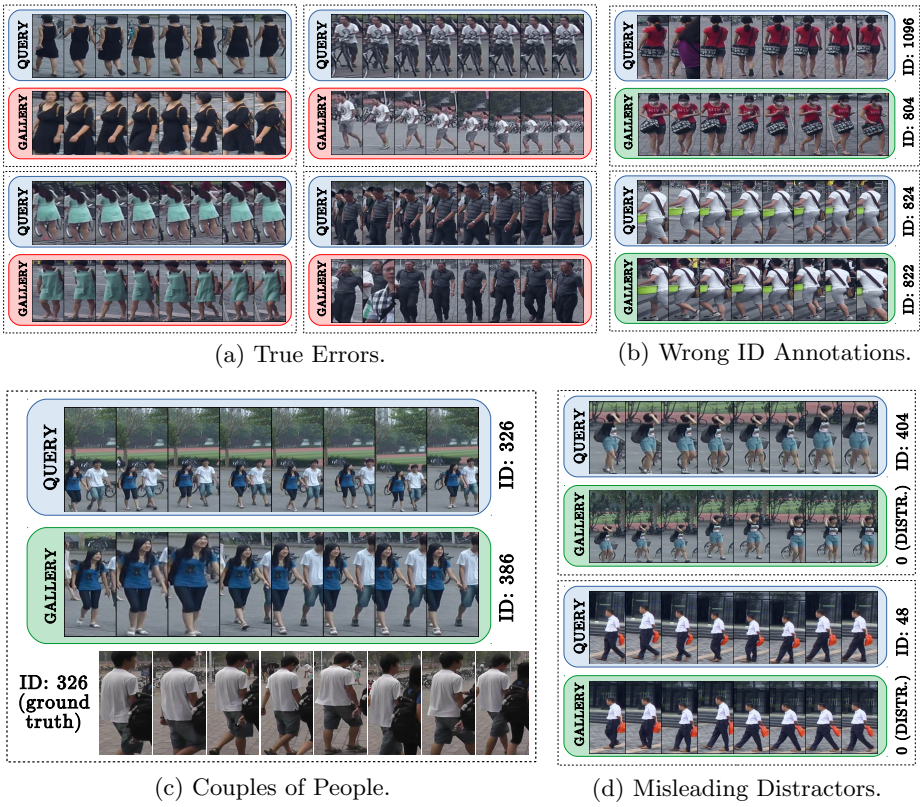


Fig. 1. Model explanation (Duke-Video-ReID) on ResNet-50 (teacher) and ResVKD-50 (student).

1980 top-1 matchings. From an analysis computed on top of these 223 cases, we identify four different categories of errors. We also asked two external researchers to annotate the errors according to these four classes as follows:

a) **True errors**: the network associates the query to a wrong identity from the gallery set (Figure 2a). This often happens when similar clothes and appearances between the two identities fool the network. Out of 223, 103 (**46.2%**) were identified as true errors;
b) **Wrong ID Annotations**: the ground truth indicates that the network associates the query to a wrong identity from the gallery set. However – for a limited set of queries – this does not hold true when visually inspecting the gallery identity. This is due to annotation errors, probably caused by a drift in the tracker (Figure 2b). Out of 223, 29 (**13.0%**) were identified as true errors;
c) **Couples of People**: some crops depict more than one subject (*e.g.* two) but only one can be associated with the tracklet id (Figure 2c). Out of 223, 37 (**16.6%**) were identified as errors involving frames with more than one person;
d) **Misleading Distractors**: cases in which the subject has been correctly identified, but the gallery tracklet was erroneously indicated as a distractor. Again, because this set has not been manually checked, some distractors are valid as they depict people (Figure 2d). Out of 223, 54 (**24.2%**) were identified as misleading distractors;

It is worth noting that the presence of the last three types of errors places a limit on the maximum score a method can obtain.



(a) True Errors. (b) Wrong ID Annotations.

(c) Couples of People. (d) Misleading Distractors.

Fig. 2. Different categories of errors on MARS. While almost half of them can be attributed to our method misidentifying between similar appearances (a), the other half are due to the automatic annotation process. In particular, wrong annotation caused by tracking drift (b), more than one identity in the same tracklet (c) and misleading distractors (d).

References

1. Wu, Y., Lin, Y., Dong, X., Yan, Y., Ouyang, W., Yang, Y.: Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In: IEEE International Conference on Computer Vision and Pattern Recognition (2018)
2. Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., Tian, Q.: Mars: A video benchmark for large-scale person re-identification. In: European Conference on Computer Vision (2016)