Supplementary Material: Learning Discriminative and Generalizable Representations by Spatial-Channel Partition for Person Re-Identification

Hao Chen[1,2], Benoit Lagadec[2], and Francois Bremond[1]
[1]University of Côte d'Azur, Inria, Stars Project-Team, France
{hao.chen,francois.bremond}@inria.fr
[2]European Systems Integration, France
benoit.lagadec@esifrance.net

1. More comparison between PCB and SCR

More examples are illustrated in Figure 1 to demonstrate that salient secondary information neglected in the previous state-of-the-art PCB [3] can be useful for the SCR model to match people with similar appearance. For the first query, our SCR model notices the bike as the salient secondary information. For the second query, our SCR model succeeds to consider the short hair. For the third query, the t-shirt color and body shape are slightly different. For the fourth query, the important secondary information is the backpack. For the fifth query, the SCR model shows a better recovery of misalignment.

These examples confirms the strong capacity of the SCR model to keep salient secondary information and to deal with misalignment.

2. Analysis of remaining mismatches

Previous state-of-the-art, *i.e.*, MGN [5] and CPM [6], and our proposed SCR achieve the same Rank1 accuracy of 95.7% on Market-1501 dataset. To understand what causes this saturation problem, we have visualized all 146 remaining mismatched samples of the SCR model. The 146 Rank1 mismatched samples can be roughly categorized into

- 8 mismatches due to misalignment.
- 9 annotation errors.
- 14 distractors produced by DPM false detection.
- 8 mismatches due to occlusion.
- 73 mismatches due to very similar clothes.
- 34 mismatches which we are not sure if they corresponds to annotation errors or to similar appearance.

We illustrate one mismatched example of each category in Figure 2. As we can observe in the figure, most remaining mismatched samples are difficult to be distinguished, even for human.

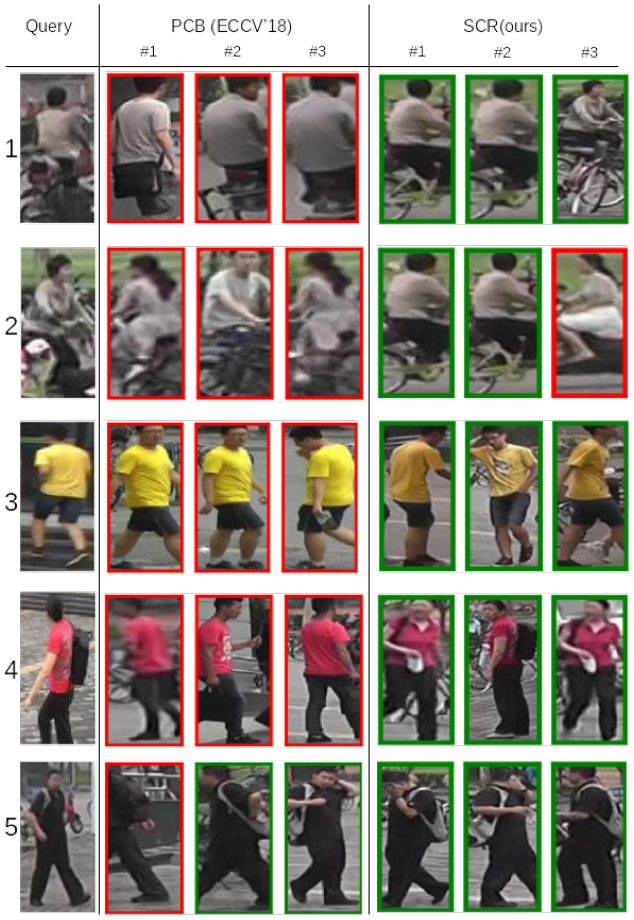

Figure 1. Examples of several mismatched samples in PCB on Market-1501 dataset, which are addressed by our proposed SCR. Red borders refers to mismatched samples. "#1", "#2" and "#3" correspond to top 3 retrieved gallery samples.

3. More saliency maps

More saliency maps are illustrated in Figure 3. Attention mechanism guides neural networks in extracting fea-

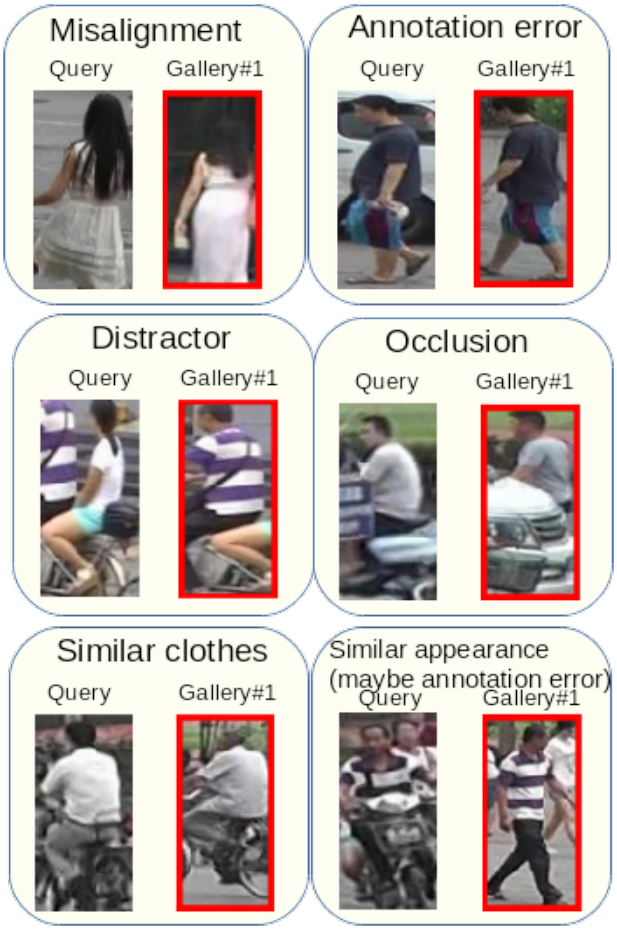

Figure 2. Examples of mismatched samples in our proposed SCR on Market-1501 dataset.

tures from the most important region in an image. As a result, only primary information is kept and fed into the classifiers. This actually reduces the influence of secondary information. For some hard samples where people are wearing similar clothes, only using the primary information is not enough. Therefore, using partitions to keep also salient secondary information can provide complementary clues. As we can see in Figure 3 (b) to (e), channel partition activates different regions as compared to conventional spatial partition. Therefore, when spatial and channel partitions are combined together, more salient information are maintained in the person visual representation. Spatial-channel partitions keep both primary and salient secondary information, which makes visual representation more robust.

References

[1] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *CVPR*, 2018.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[3] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, 2018.

[4] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6458, 2017.

[5] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou. Learning discriminative features with multiple granularities for person re-identification. In *ACM Multimedia*, 2018.

[6] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji. Pyramidal person re-identification via multi-loss dynamic training. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

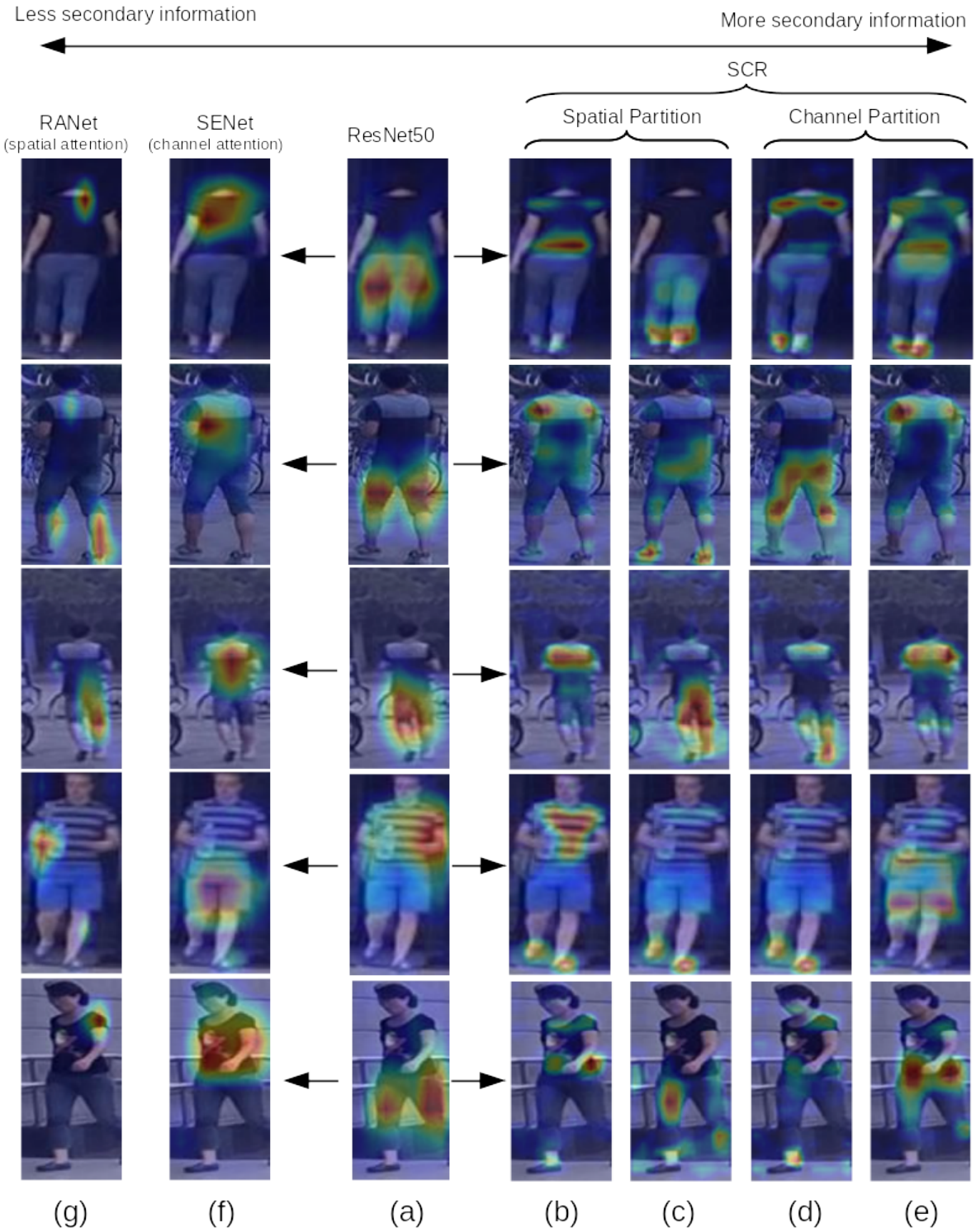

Figure 3. Comparisons of saliency maps generated by Grad-CAM [2] applied on 4 CNN models on Market-1501 test set. (a): A vanilla ResNet-50 w/o partition nor attention mechanism. (b) to (e): Our proposed SCR model, where (b) and (c) are saliency maps on two spatial parts in the second branch of the SCR model, (d) and (f) are saliency maps on two channel groups in the second branch of the SCR model. (f): Squeeze-and-Excitation Network [1]. (g): Residual Attention Network [4]. When (b) to (e) are combined together, more salient information is maintained in the SCR model as compared to attention models (g) and (f).