Salient-to-Broad Transition for Video Person Re-identification
Supplementary Material

Shutao Bai[1,2], Bingpeng Ma[2], Hong Chang[1,2], Rui Huang[3], Xilin Chen[1,2]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, 100190, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
[3]The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, China
shutao.bai@vipl.ict.ac.cn,bpma@ucas.ac.cn,ruihuang@cuhk.edu.cn, {changhong,xlchen}@ict.ac.cn

In this supplementary material, we provide more details and ablation studies about the mutual information loss $\mathcal{L}_{mi}$ (Equation 8 in the main paper).

A. Mutual Information Loss

In SINet, we leverage mutual information loss to maintain the diversities between frame-level embeddings. The formulation of $\mathcal{L}_{mi}$ is similar to Equation 3,4 in [1], except that our $\mathcal{L}_{mi}$ is conducted at frame-level. Formally, in a mini-batch, one identity has K clips and one clip has T frames. Let $\mathbf{z}_{k,t}$ be the normalized embedding of the t-th frame in k-th clip for this identity. Mutual information loss is conducted for each identity separately, and its formulation is defined as:

$$\mathcal{L}_{mi}(\mathbf{z}) = \frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}\mathcal{L}_{mi}(\mathbf{z}_{k,t})$$

$$\mathcal{L}_{mi}(\mathbf{z}_{k,t}) = \sum_{\substack{i=1\\ i\neq k}}^{K}\frac{\exp\left(\mathbf{z}_{k,t}^{T}\mathbf{z}_{i,t}/\tau\right)}{\exp\left(\mathbf{z}_{k,t}^{T}\mathbf{z}_{i,t}/\tau\right)+\sum_{\substack{j=1\\ j\neq t}}^{T}\left(\mathbf{z}_{k,t}^{T}\mathbf{z}_{k,j}/\tau\right)}$$

Here, τ is the temperature and is set to 0.07 in default. This loss works by requiring the embeddings with same temporal index in different clips are mapped close, while embeddings with difference temporal index in same clip are mapped sufficiently far apart. In this way, this loss assists the salient-to-broad transition for consecutive frames. For an input mini-batch, the $\mathcal{L}_{mi}$ in Equation 8 is the average of $\mathcal{L}_{mi}(\mathbf{z})$ for all identities.

In Equation 8, λ_2 controls the influence of mutual information loss. While a small λ_2 means no effective constrains, a large λ_2 encourages the later frame embeddings to be completely different with the former ones, which is achieved by focusing on the non-discriminative backgrounds. Empirically, we set $\lambda_2 = 0.01$ in our experiments (see Table 1).

Table 1. Influence of λ_2 in SBM on MARS.

λ_2	mAP	rank-1
0.003	85.3	89.5
0.01 (default)	85.7	90.2
0.03	85.5	89.8

References

[1] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*, 2020. 1