Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification

Fengxiang Yang[1]*, Zhun Zhong[2]*, Zhiming Luo[3]†, Yuanzheng Cai[4], Yaojin Lin[5], Shaozi Li[1]†, Nicu Sebe[2]

1 Artificial Intelligence Department, Xiamen University

2 Department of Information Engineering and Computer Science, University of Trento

3 Postdoc Center of Information and Communication Engineering, Xiamen University

4 Minjiang University 5 Minnan Normal University

Project: https://github.com/FlyingRoastDuck/MetaCam_DSCE

Appendix A. Detailed Explanation of DSCE

We use $\mathbf{C}_j$ to denote the j-th centroid ($1 \leq j \leq N_c$). $p_j = \frac{\exp(\mathbf{C}_j^{\mathrm{T}}\mathbf{f}/\tau)}{\sum_{m=1}^{N_c}\exp(\mathbf{C}_m^{\mathrm{T}}\mathbf{f}/\tau)}$ is the probability of assigning $\mathbf{f}$ to the j-th class and $\sum_{j=1}^{N_c} p_j = 1$. $\hat{\mathbf{y}}$ is the one-hot vector of $\mathbf{f}$ obtained by clustering. $\widetilde{y}_j = \frac{\exp(\hat{y}_j)}{\sum_{m=1}^{N_c}\exp(\hat{y}_m)}$ is the j-th element of softmax-normalized $\hat{\mathbf{y}}$. Then, we have:

$$\widetilde{y}_j = \begin{cases} \frac{1}{N_c-1+e}, & \hat{y}_j = 0 \\ \frac{e}{N_c-1+e}, & \hat{y}_j = 1 \end{cases}. \tag{1}$$

The DSCE loss in our paper can be reformulated as:

$$L_{dsce} = -\sum_{j=1}^{N_c} p_j \log \widetilde{y}_j. \tag{2}$$

[1] proves that a loss function L is robust to noisy labels if it satisfies the following constraint:

$$\sum_{k=1}^{N_c} L(\mathbf{f}, k) = Z, \tag{3}$$

where $L(\mathbf{f}, k)$ indicates the loss when the class label is k. Next, we will prove that DSCE loss satisfies Eq. 3.

Theorem 1 *In a multi-class classification problem, the proposed DSCE loss (Eq. 2) satisfies the constraint in Eq. 3.*

Proof 1 When the class label is k-th class, *i.e.*, $\hat{y}_k = 1$, Eq. 2 can be reformulated as:

$$L_{dsce}(\mathbf{f}, k) = -p_k \log(\widetilde{y}_k) - \sum_{j \neq k}^{N_c} p_j \log(\widetilde{y}_j). \tag{4}$$

For convenience, we define $Q = \frac{1}{N_c-1+e}$. According to Eq. 1, Eq. 4 can be simplified as:

$$\begin{aligned} L_{dsce}(\mathbf{f}, k) &= -p_k \log(eQ) - (\log Q) \sum_{j \neq k}^{N_c} p_j \\ &= -(1 + \log Q) p_k - (1 - p_k) \log Q \\ &= -p_k - \log Q. \end{aligned}$$

Then, we have:

$$\begin{aligned} \sum_{k=1}^{N_c} L_{dsce}(\mathbf{f}, k) &= -\sum_{k=1}^{N_c} p_k - \sum_{k=1}^{N_c} \log Q \\ &= -1 - N_c \log Q. \end{aligned}$$

Therefore, the proposed DSCE loss satisfies Eq. 3 and is robust to noisy labels. ■

References

[1] Aritra Ghosh, Himanshu Kumar, and PS Sastry. Robust loss functions under label noise for deep neural networks. In *AAAI*, 2017. 1

*Equal contribution: yangfx@stu.xmu.edu.cn

†Corresponding author: {zhiming.luo, szlig}@xmu.edu.cn