Supplementary Material for Unsupervised Visible-Infrared Person Re-Identification via Progressive Graph Matching and Alternate Learning

Zesen Wu[1], Mang Ye[1,2*]
[1]National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence,
Hubei Key Laboratory of Multimedia and Network Communication Engineering,
School of Computer Science, Wuhan University, Wuhan, China
[2] Hubei Luojia Laboratory, Wuhan, China

1. Details of CA Assisted Learning in DCL

Channel augmentation [1] is a common and powerful data augmentation to bridge the gap between visible and infrared images, and thus channel augmented (CA) images are used to assist in the learning process of visible streams. CA dataset is denoted as $\mathcal{T}^a = \{x_i^a | i = 1, 2, \cdots, N\}$ with the same number of images as the visible dataset. In the DCL framework, visible images have their augmented CA images to assist the training progress, which uses the following equations:

$$\mathcal{L}_{vis} = -\sum_{i=1}^{N_B/2} \log\left(\frac{\exp(\mathcal{K}^v[\widetilde{y_i^v}]^T \cdot f(x_i^v)/\tau)}{\sum_{k=1}^{Y^v} \exp(\mathcal{K}^v[k]^T \cdot f(x_i^v)/\tau)}\right), \quad (1)$$

$$\mathcal{L}_{ca} = -\sum_{i=1}^{N_B/2} \log\left(\frac{\exp(\mathcal{K}^v[\widetilde{y_i^v}]^T \cdot f(x_i^a)/\tau)}{\sum_{k=1}^{Y^v} \exp(\mathcal{K}^v[k]^T \cdot f(x_i^a)/\tau)}\right), \quad (2)$$

where $\widetilde{y_i^v}$ is the class (pseudo label) for image x_i^v and its channel augmented image x_i^a. Besides, τ is a temperature factor.

2. Details of CA Assisted Learning in ACCL

Visible to infrared learning exhibits a similar form like *infrared to visible* learning described in main text of the paper. The difference is that half of the N_B images are visible images and the other half are their corresponding CA images. It consists of visible and CA learning, denoted as $\mathcal{L}_{v2r}$ and $\mathcal{L}_{a2r}$, respectively. The equations are:

$$\mathcal{L}_{V2R} = \mathcal{L}_{v2r} + \mathcal{L}_{a2r},$$

$$\mathcal{L}_{v2r} = -\sum_{i=1}^{N_B/2} \log\left(\frac{\exp(\mathcal{K}^r[\hat{y_i^v}]^T \cdot f(x_i^v)/\tau)}{\sum_{k=1}^{Y^r} \exp(\mathcal{K}^r[k]^T \cdot f(x_i^v)/\tau)}\right), \quad (3)$$

$$\mathcal{L}_{a2r} = -\sum_{i=1}^{N_B/2} \log\left(\frac{\exp(\mathcal{K}^r[\hat{y_i^v}]^T \cdot f(x_i^a)/\tau)}{\sum_{k=1}^{Y^r} \exp(\mathcal{K}^r[k]^T \cdot f(x_i^a)/\tau)}\right),$$

where $\hat{y_i^v} = \mathbf{V2R}[\widetilde{y_i^v}]$. $\widetilde{y_i^v}$ is the pseudo label for the infrared image x_i^v, and $\hat{y_i^v}$ is the cross-modality correspondence for $\widetilde{y_i^v}$, also the cross-modality label for image x_i^v and x_i^a. τ is a temperature factor. The *visible to infrared* aims to bridge modality gap by gathering the given visible (CA) sample to its corresponding cross-modality proxy while scattering other proxies.

*Corresponding Author: Mang Ye

References

[1] Mang Ye, Weijian Ruan, Bo Du, and Mike Zheng Shou. Channel augmented joint learning for visible-infrared recognition. In *ICCV*, pages 13567–13576, 2021. 1