

Supplementary Material for Towards Grand Unified Representation Learning for Unsupervised Visible-Infrared Person Re-Identification

Bin Yang Jun Chen[†] Mang Ye[†]

National Engineering Research Center for Multimedia Software,
School of Computer Science, Hubei LuoJia Laboratory, Wuhan University, Wuhan, China

<https://github.com/yangbincv/GUR>

1. Bottom-up Domain Learning Strategy

Bottom-up domain learning strategy has three training stages, *i.e.*, *intra-camera training*, *inter-camera* and *inter-modality training*, as shown in Fig. 1. In each stage, the augmented dual-contrastive (ADC) learning [2] is conducted for pseudo-label-based unsupervised learning. In the **intra-camera training stage**, the ADC is executed alternately in each camera domain separately via clustering the intra-camera similarity by DBSCAN clustering algorithm [1] to assign pseudo-labels. In the **inter-camera training**, the cross-memory association embedding (CAE) module at the camera level calculates the association embedding of persons with each camera memory for inter-camera (intra-modality) DBSCAN clustering (*i.e.*, clustering the data of each modality separately). Then, ADC with two modality-specific memories is performed to learn camera-invariant features within each modality. Similarly, during the **inter-modality training**, the CAE module at the modality level computes the association embedding of pedestrians with each modality memory for inter-modality clustering (*i.e.*, simultaneously input all data into DBSCAN for clustering without considering domains), and the ADC with a modality-shared memory is conducted to learn the final unified features. From intra-camera to inter-modality training, the model progressively captures camera-invariant and modality-invariant features. Three stages are executed in an alternate manner during one training epoch.

[†]Corresponding Author.

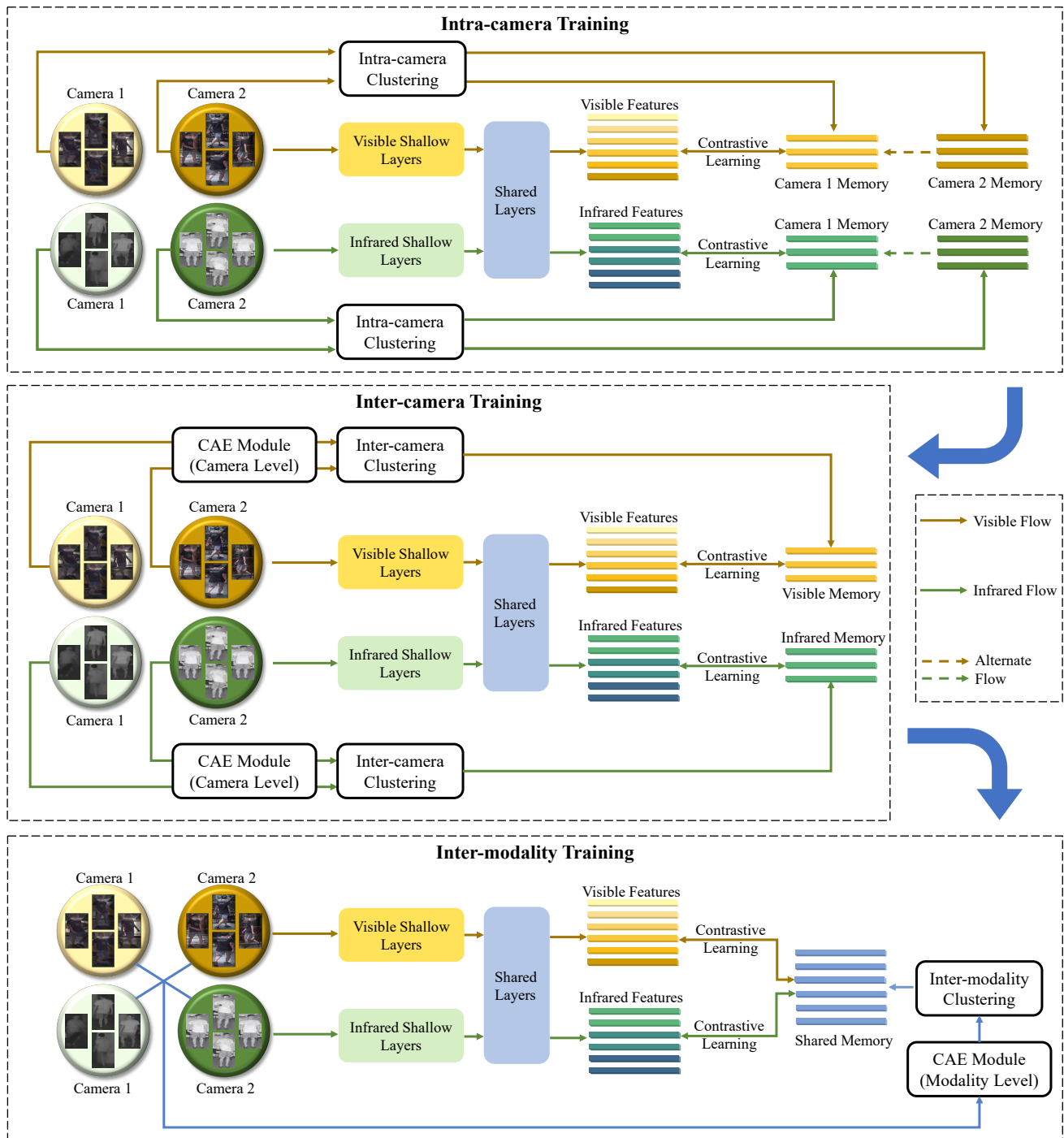


Figure 1. Illustration of bottom-up domain learning strategy with two cameras within each modality as an example.

References

- [1] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996. [1](#)
- [2] Bin Yang, Mang Ye, Jun Chen, and Zesen Wu. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification. In *ACM MM*, page 2843–2851, 2022. [1](#)