

Supplementary Material for Shallow-Deep Collaborative 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/SDCL>

Shallow and Deep Feature Map. To illustrate the rationale for selecting the shallow embedding as a collaborative element, we visualize the part shallow and deep feature maps for one visible and one infrared sample, as shown in Fig. 1. Shallow features provide a more locally detailed representation of input data for capturing cross-modality invariance, as noted in red boxes in Fig. 1, offering a balance to the intricate and abstract representations generated by deep features in the transformer. By integrating these complementary aspects, our approach achieves a nuanced and comprehensive representation. These intra-modality and inter-modality collaborations leverage the strengths of both shallow and deep features, harnessing the immediate contextual understanding of the shallow layers and the complex, high-level abstractions captured by the deep layers. that enhances the overall performance in the context of VI-ReID This deliberate choice contributes to the unique character of our method, enhancing the cross-modality retrieval performance and distinguishing it from approaches that solely rely on deep features [1–4].

*Corresponding Author.

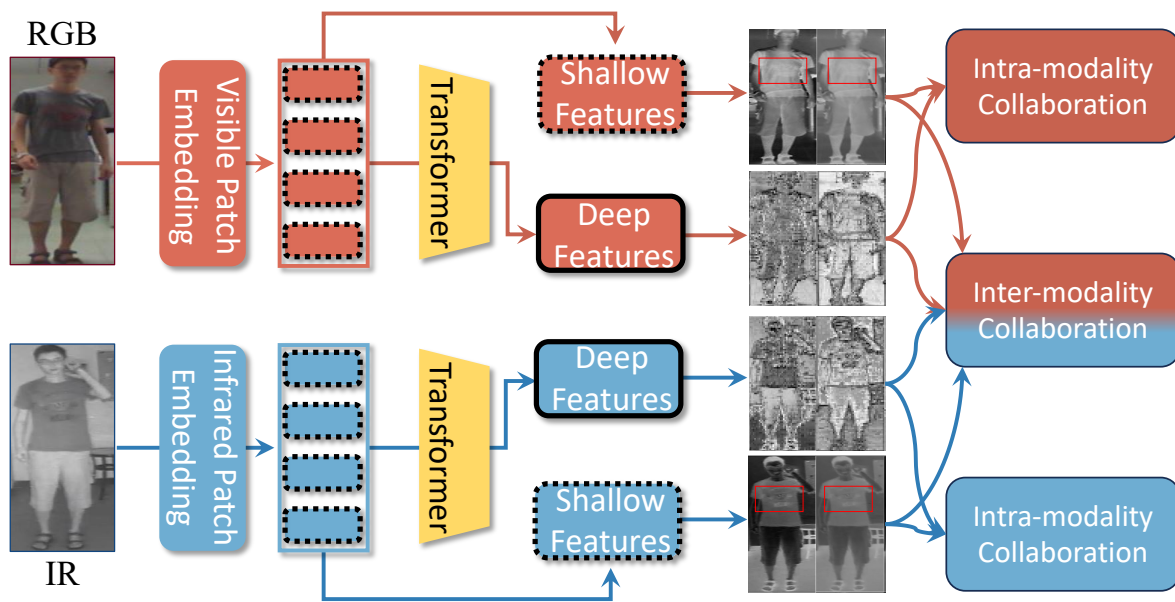


Figure 1. Illustration of the collaboration of shallow and deep features. Shallow features retain more local details on output feature maps for capturing cross-modality invariance in contrast to deep features, as noted in red boxes. In contrast, deep features generate intricate and abstract representations, emphasizing semantic information. The interplay between deep and shallow features seamlessly integrates complementary aspects, resulting in a nuanced and comprehensive representation. For better visualization, the input image is scaled to size 1440×720 .

References

- [1] Jiangming Wang, Zhizhong Zhang, Mingang Chen, Yi Zhang, Cong Wang, Bin Sheng, Yanyun Qu, and Yuan Xie. Optimal transport for label-efficient visible-infrared person re-identification. *ECCV*, 2022. [1](#)
- [2] Zesen Wu and Mang Ye. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning. In *CVPR*, 2023.
- [3] Bin Yang, Mang Ye, Jun Chen, and Zesen Wu. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification. In *ACM MM*, 2022.
- [4] Bin Yang, Jun Chen, and Mang Ye. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *ICCV*, 2023. [1](#)