# Supplementary Material: Similarity Metric Learning For RGB-Infrared Group Re-Identification

Jianghao Xiong[1], Jianhuang Lai[1,2,3,4*]

[1]School of Computer Science and Engineering, Sun Yat-Sen University, China
[2]Guangdong Province Key Laboratory of Information Security Technology, China
[3]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
[4]Key Laboratory of Video and Image Intelligent Analysis and Applicaiton Technology,
Ministry of Public Security, China

xiongjh7@mail2.sysu.edu.cn, stsljh@mail.sysu.edu.cn

## 1. Overview

In this supplementary material, we provide more experiment details and ablation studies of CPM and CPD. Note that we do not optimize the hyper-parameters and all the networks can be tuned for better performance.

- We visualize several retrieval results by CPM on CM-Group.

- We conduct further ablation studies for different backbones with CPD on CSG.

- We present more comparisons with state-of-the-arts in Group ReID.

## 2. Retrieval Results by CPM on CM-Group

We present six retrieval results by CPM on CM-Group in Fig. 1. It is clear that all true matches have large variations in viewpoints and group layout. In addition, challenging situations such as clothes change (row 2), scale change (row 3 and row 6) and occlusions (row 5) can be effectively handled by our CPM.

| Method | Task | Rank1 | Rank10 | Rank20 | mAP |
|---|---|---|---|---|---|
| Swin Transformer [7] | R | 32.2 | 57.0 | 63.8 | 19.1 |
| Swin Transformer [7]+CPD | G | 59.6 | 78.7 | 83.2 | 34.5 |
| DenseNet-121 [2] | R | 59.0 | 81.8 | 86.2 | 34.9 |
| DenseNet-121 [2] +CPD | G | 87.9 | 96.1 | 97.6 | 51.1 |
| HRNet-18 [8] | R | 60.2 | 82.1 | 86.1 | 35.5 |
| HRNet-18 [8]+CPD | G | 88.2 | 95.9 | 96.8 | 51.6 |

Table 1. Results of three different backbones in terms of CMC (%) and mAP (%) performances on CSG. "R/G" stands for ReID and G-ReID respectively.

## 3. Ablation Studies for different backbones

We test three different backbones Swin Transformer [7], DenseNet-121 [2] and HRNet-18 [8] with CPD on CSG. As shown in Tab. 1, the performance of HRNet-18 is best on CSG (ReID) and the G-ReID performance with CPD is best among the three backbones. The results also indicate that with a better feature extraction network, our CPM is supposed to gain better performance in G-ReID task.

---

*Corresponding Author

# 4. More Comparisons with State-of-the-arts in Group ReID

We compare our modified CPM with more state-of-the-arts methods on CSG and RoadGroup for G-ReID. As shown in Tab. 2, our modified CPM achieves outstanding performances on both G-ReID datasets. It also demonstrates that our CPM framework is flexible to be applied to G-ReID task.

| Method | Venue | CSG | | | | RoadGroup | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Rank1 | Rank5 | rank10 | mAP | Rank1 | Rank5 | rank10 | mAP |
| CRRRO-BRO [11] | BMVC 2009 | 10.4 | 25.8 | 37.5 | - | 17.8 | 34.6 | 48.1 | - |
| Covariance [1] | ICPR 2010 | 16.5 | 34.1 | 47.9 | - | 38.0 | 61.0 | 73.1 | - |
| PREF [6] | ICCV 2017 | 19.2 | 36.4 | 51.8 | - | 43.0 | 68.7 | 77.9 | - |
| BSC+CM [12] | ICIP 2016 | 24.6 | 38.5 | 55.1 | - | 58.6 | 80.6 | 87.4 | - |
| DotGNN [3] | MM 2019 | - | - | - | - | 74.1 | 90.1 | 92.6 | - |
| GCGNN [13] | TMM 2020 | - | - | - | - | 81.7 | 94.3 | 96.5 | - |
| MGR [5] | TCYB 2021 | 57.8 | 71.6 | 76.5 | - | 80.2 | 93.8 | 96.3 | - |
| DotSCN [4] | TCSVT 2021 | - | - | - | - | 84.0 | 95.1 | 96.3 | - |
| MACG [9] | TPAMI 2020 | 63.2 | 75.4 | 79.7 | - | 84.5 | 95.0 | 96.9 | - |
| 3DT [10] | CVPR 2022 | **92.9** | **97.3** | **98.1** | **92.1** | 91.4 | 97.5 | 98.8 | 94.3 |
| **CPM*(Ours)** | - | 88.6 | 94.1 | 95.3 | 51.5 | **92.6** | **98.8** | **98.8** | **95.4** |

Table 2. Comparisons with state-of-the-arts methods on CSG and RoadGroup.



Figure 1. Examples of retrieval results by CPM on CM-Group. In each row, the leftmost image is the query followed by top-10 nearest neighboring images.The green digits indicate true matches while the red digits indicate false matches. Images are resized for better visualization.

# References

[1] Yinghao Cai, Valtteri Takala, and Matti Pietikäinen. Matching groups of people by covariance descriptor. In *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010*, pages 2744–2747. IEEE Computer Society, 2010. 2

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 1

[3] Ziling Huang, Zheng Wang, Wei Hu, Chia-Wen Lin, and Shin'ichi Satoh. Dot-gnn: Domain-transferred graph neural network for group re-identification. In *Proceedings of the 27th ACM International Conference on Multimedia*, page 1888–1896, 2019. 2

[4] Ziling Huang, Zheng Wang, Chung-Chi Tsai, Shin'ichi Satoh, and Chia-Wen Lin. Dotscn: Group re-identification via domain-transferred single and couple representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(7):2739–2750, 2020. 2

[5] Weiyao Lin, Yuxi Li, Hao Xiao, John See, Junni Zou, Hongkai Xiong, Jingdong Wang, and Tao Mei. Group reidentification with multigrained matching and integration. *IEEE Transactions on Cybernetics*, 51(3):1478–1492, 2021. 2

[6] Giuseppe Lisanti, Niki Martinel, Alberto Del Bimbo, and Gian Luca Foresti. Group re-identification via unsupervised transfer of sparse features encoding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2449–2458, 2017. 2

[7] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021. 1

[8] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5686–5696, 2019. 1

[9] Yichao Yan, Jie Qin, Bingbing Ni, Jiaxin Chen, Li Liu, Fan Zhu, Wei-Shi Zheng, Xiaokang Yang, and Ling Shao. Learning multi-attention context graph for group-based re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2

[10] Quan Zhang, Kaiheng Dang, Jian-Huang Lai, Zhanxiang Feng, and Xiaohua Xie. Modeling 3d layout for group re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7512–7520, June 2022. 2

[11] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *British Machine Vision Conference, BMVC 2009, London, UK, September 7-10, 2009. Proceedings*, pages 1–11, 2009. 2

[12] Feng Zhu, Qi Chu, and Nenghai Yu. Consistent matching based on boosted salience channels for group re-identification. In *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, pages 4279–4283. IEEE, 2016. 2

[13] Ji Zhu, Hua Yang, Weiyao Lin, Nian Liu, Jia Wang, and Wenjun Zhang. Group re-identification with group context graph neural networks. *IEEE Transactions on Multimedia*, 23:2614–2626, 2020. 2