

附录 – 面向跨模态行人重识别的单模态自监督信息挖掘

在正文的实验中，ResNet-50 (He 等, 2016) 的骨干模型架构设计参照了 MGN (Wang 等, 2018) 的方法，在输出的特征图上分割水平条带提取特征。在为了更充分地验证提出方法的有效性，使用 ResNet-50 的默认架构（对最后一层特征图进行全局平均池化）进行与正文实验设置完全一致的消融实验和对比实验。

(1) 消融实验

参照 2.2 节的 1) 的实验设置进行基于 ResNet-50 默认架构的消融实验。实验结果如表 A1 所示。与正文表 1 的结果对比，使用默认架构的性能略低于使用 MGN 架构的性能，消融实验之间的对比结论不变，验证了方法每个部分的有效性。

(2) 对比实验

参照 2.2 节的 2) 的实验设置，基于 ResNet-50 默认架构实现本文方法，与当前先进方法进行对比实验。在 SYSU-MM01、RGBNT201 和 RegDB 数据集上的实验结果分别如表 A2、A3 和 A4 所示。为节约篇幅，对比方法的结果只展示性能最高和具有代表性的 LBA (unreal)、CM-NAS (unreal)、CAJL (unreal) 和 MPANet (unreal)。

基于 ResNet-50 的默认架构实现本文方法，三通道模型在 SYSU-MM01 数据集上取得与对比方法相当的性能，而在样本数更加受限的 RGBNT201 数据集和 RegDB 数据集上则能取得显著超过对比方法的性能。双模型融合的性能优于单模型的性能。实验结果表明了本文方法的有效性。

表A1 使用ResNet-50默认架构在SYSU-MM01数据集上的消融实验性能
Table A1 Ablation study performance for ResNet-50 (default architecture) on SYSU-MM01

| 实验 | 互学习类型 | 预训练互学习 | 微调互学习 | /% | | | |
|--------------|-------------------|--------|-------|------|--------|------|--------|
| | | | | 全搜索 | | 室内搜索 | |
| | | | | mAP | Rank-1 | mAP | Rank-1 |
| 0 (ImageNet) | 无 | 无 | 无 | 42.4 | 41.8 | 52.7 | 42.8 |
| 1 (基础模型) | 无 | 无 | 无 | 54.1 | 54.8 | 63.4 | 54.8 |
| 2 | 三通道-三通道 | 有 | 无 | 55.0 | 56.0 | 66.8 | 58.8 |
| 3 | 三通道-乱序通道 | 有 | 无 | 57.6 | 58.9 | 67.4 | 60.0 |
| 4 | 三通道-灰度图 | 有 | 无 | 57.4 | 58.2 | 69.1 | 62.0 |
| 5 | 三通道-跨光谱图像 (Fan 等) | 有 | 无 | 57.8 | 58.8 | 69.0 | 62.0 |
| 6 | 三通道-单通道掩膜 (本文) | 有 | 无 | 58.1 | 59.4 | 70.5 | 63.5 |
| 7 | 三通道-三通道 | 有 | 有 | 61.4 | 62.8 | 73.1 | 66.9 |
| 8 | 三通道-乱序通道 | 有 | 有 | 63.3 | 64.8 | 74.1 | 66.8 |
| 9 | 三通道-灰度图 | 有 | 有 | 62.1 | 64.9 | 73.8 | 67.2 |
| 10 | 三通道-跨光谱图像 (Fan 等) | 有 | 有 | 65.6 | 68.6 | 76.7 | 71.6 |
| 11 (三通道模型) | 三通道-单通道掩膜 (本文) | 有 | 有 | 66.5 | 69.6 | 78.2 | 73.2 |

表A2 使用ResNet-50默认架构在SYSU-MM01数据集上的跨模态匹配性能对比

Table A2 Performance comparisons of cross-modality matching for ResNet-50 (default architecture) on SYSU-MM01

/%

| 方法 | 全搜索 | | | 室内搜索 | | |
|------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | mAP | Rank-1 | Rank-10 | mAP | Rank-1 | Rank-10 |
| LBA (unreal) (Park 等, 2021) | 54.6 | 55.4 | 92.2 | 68.8 | 60.9 | 96.5 |
| CM-NAS (unreal) (Fu 等, 2021) | 50.6 | 52.4 | 89.2 | 65.3 | 57.7 | 94.4 |
| CAJL (unreal) (Ye 等, 2021) | 67.8 | 70.5 | 95.5 | 79.9 | 75.9 | 97.3 |
| MPANet (unreal) (Wu 等, 2021) | 68.3 | 71.3 | 95.7 | 79.7 | 75.5 | 97.9 |
| 本文方法 (三通道模型) | 66.5 | 69.6 | 96.3 | 78.2 | 73.2 | 98.2 |
| 本文方法 (单通道模型) | 65.1 | 68.8 | 95.8 | 78.0 | 73.4 | 97.5 |
| 本文方法 (双模型融合) | 67.6 | 70.9 | 96.6 | 79.5 | 74.9 | 98.2 |

表A3 使用ResNet-50默认架构在RGBNT201数据集上的跨模态匹配性能对比

Table A3 Performance comparisons of cross-modality matching for ResNet-50 (default architecture) on RGBNT201

/%

| 方法 | 热成像-可见光 | | 可见光-热成像 | | 近红外-可见光 | | 可见光-近红外 | |
|------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 |
| LBA (unreal) (Park 等, 2021) | 18.6 | 18.1 | 18.4 | 17.2 | 31.4 | 37.3 | 30.7 | 31.9 |
| CM-NAS (unreal) (Fu 等, 2021) | 26.8 | 25.1 | 28.7 | 28.0 | 36.8 | 38.4 | 36.6 | 41.1 |
| MPANet (unreal) (Wu 等, 2021) | 28.8 | 27.6 | 25.0 | 23.4 | 34.5 | 32.1 | 34.2 | 37.8 |
| 本文方法 (三通道模型) | 34.5 | 36.2 | 34.3 | 36.1 | 44.1 | 45.5 | 45.7 | 45.9 |
| 本文方法 (单通道模型) | 33.2 | 35.3 | 33.0 | 34.0 | 38.6 | 39.7 | 41.5 | 45.2 |
| 本文方法 (双模型融合) | 36.3 | 37.3 | 36.4 | 39.1 | 43.4 | 46.0 | 45.9 | 47.4 |

表A4 使用ResNet-50默认架构在RegDB数据集上的跨模态匹配性能对比

Table A4 Performance comparisons of cross-modality matching for ResNet-50 (default architecture) on RegDB

/%

| 方法 | 热成像-可见光 | | 可见光-热成像 | |
|-----------------|-------------|-------------|-------------|-------------|
| | mAP | Rank-1 | mAP | Rank-1 |
| LBA (unreal) | 67.0 | 72.1 | 67.7 | 72.8 |
| CM-NAS (unreal) | 83.3 | 87.3 | 82.0 | 86.8 |
| CAJL (unreal) | 76.3 | 84.4 | 78.9 | 87.0 |
| MPANet (unreal) | 82.2 | 84.7 | 83.4 | 85.2 |
| 本文方法 (三通道) | 84.9 | 87.5 | 87.1 | 90.3 |
| 本文方法 (单通道) | 83.9 | 87.0 | 85.9 | 88.9 |
| 本文方法 (双模型) | 86.2 | 88.3 | 88.3 | 90.9 |

参考文献

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Wang G S, Yuan Y F, Chen X, Li J W and Zhou X. 2018. Learning discriminative features with multiple granularities for person re-identification//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea: ACM: 274-282 [DOI: 10.1145/3240508.3240552]