

## 摘要

行人重识别旨在摄像机网络中查找特定目标人。通过给定目标人的查询图像，系统自动地在海量数据库中进行匹配查询，以返回该目标人在何时何摄像头下出现过的信息。行人重识别具有广阔的应用前景，能够帮助寻找丢失儿童、老人，追踪犯罪嫌疑人等，是构建智慧城市中智能安防的关键技术之一。由于其重要性，行人重识别在学术界和工业界受到越来越多的关注。本文主要围绕行人表征学习及场景迁移技术进行探索，以提升行人重识别在实际应用上的性能。本文的主要工作可以总结如下：

首先，本文提出了基于部件检测的全局一局部特征联合学习方法。该方法提出了部件检测技术来对图像进行区域对齐。然后，通过设计全局一局部特征联合学习网络来有效地提取并融合全局信息和局部信息，使得生成的全局一局部特征表达力更强、鲁棒性更高。实验证明，基于部件检测的全局一局部特征联合学习方法能够有效地应对行人图像不对齐、行人表观信息变化大等挑战，大幅提高行人重识别的准确率。

其次，本文提出了面向行人重识别的检索框架。针对全局一局部特征的特征维度高，以及现有重识别方法都是将查询图像与数据库图像一一比对等导致行人检索过程时间开销大的问题，本文提出了快速高效的检索框架。该检索框架通过离线索引和由粗到细的在线检索方案能够在保持准确率的同时，有效地降低线上检索时间开销。

然后，本文提出了基于生成对抗网络的场景迁移技术。当前数据集规模小、场景单一导致行人重识别模型无法在跨场景任务下取得良好性能。此外，由于数据标注比较困难，不可能对每一新场景都构造新的训练数据集。为了解决上述问题，本文借鉴生成对抗网络的思想，提出了面向目标场景的迁移方法。该方法能够将行人图像的风格转换为目标场景的数据风格，同时在风格迁移过程中保持行人信息不发生改变。实验结果证明，在迁移数据上训练的行人重识别模型要远远优于基准方法，在任意新的目标场景中都能表现出良好性能。然而，目标场景中各个摄像机所拍摄的图片也会具有较大的风格差异。为了更好地模拟目标场景的数据特点，本文进一步提出了面向目标场景摄像机的迁移方法。该方法能够将任意一张原始图片迁移至目标场景的各个摄像机下。

最后，本文搭建了行人图像快速检索演示系统。该系统利用了本文所提基于部件检测的全局一局部特征联合学习方法以及快速的检索框架，有效地将本文提出的算法用于实际系统中，在实践中验证了本文算法的有效性。

关键词： 行人重识别，全局一局部特征，场景迁移，检索演示系统



# Study on Pedestrian Descriptor Learning and Scene Transfer for Person Re-Identification

Longhui Wei (Computer Applied Technology)

Directed by Prof. Shiliang Zhang

## ABSTRACT

Person Re-Identification(ReID) is designed to find the specific target person from camera networks. Given a query image of the target person, ReID system will automatically match it with gallery images, and return the information about when and where the target person has appeared. Person ReID has a wide range of applications, *e.g.*, helping find the lost children and the elderly, tracking suspects. Thus, person ReID is one of the key technologies to build intelligent security ability for smart city. Because of its importance, person ReID has attracted more and more attention from both academia and industry. This paper focuses on descriptor learning and scene transfer for person ReID, and tries to improve the ReID performance in real scenarios. The contributions of this paper can be mainly summarized as follows:

Firstly, this paper proposes a global-local descriptor learning method based on human part detection. This method utilizes region detection related technique to align person regions. Moreover, a global-local descriptor jointly learning framework is designed to effectively extract and fuse the global and local cues. In the end, the learned global-local descriptor is expressive and robust. Extensive experiments show our proposed approach can cope with the challenges of misalignment and huge variance on person image. Finally, the person ReID performance on public datasets is hugely improved by our method.

Secondly, this paper proposes a retrieval framework for person ReID. For the high dimension of learned global-local descriptor and the costly retrieval process that match the query image with gallery images one by one, person ReID is very time-consuming. Target to solve this issue, this paper proposes an effective and efficient retrieval framework. Through utilizing offline indexing and coarse-to-fine online retrieval approaches, the retrieval framework can hugely reduce the online retrieval time while maintaining the ReID accuracy.

Thirdly, this paper proposes scene transfer related approaches based on generative adversarial networks. The ReID model trained on small-scale and single-scene dataset cannot perform well under other new scenes. Moreover, due to the difficulty of data annotation, it is

impossible to construct a training set for each target scene. Target to address it, this paper proposes a scene-oriented transfer related approach based on generative adversarial networks. It can transfer the style of source images to be more similar with target scene, while keeping the person information unchanged. Experimental results show the proposed method is superior to the benchmark methods, and performs well under each new target scene. However, person images captured by different cameras in target scene also show different styles. To better model the style variance inside target scene, camera-oriented transfer technologies are further designed. These methods can transfer one source image to each camera on target scene.

Finally, this paper builds an efficient pedestrian retrieval system. This system is constructed based on the proposed global-local descriptor jointly learning method and efficient retrieval framework. Thus, the effectiveness of the proposed approaches in this paper is finally evaluated in real ReID system.

**KEY WORDS:** Person re-identification, Global-local descriptor, Scene transfer, Retrieval system