摘要

随着视频监控设备的大规模使用与普及，对行人及车辆目标数据的处理、分析和利用是协助警方办案、维护社会安全与稳定的重要手段。目标再识别是监控视频技术领域的核心课题之一，蕴含着巨大的社会价值与科研价值。目标再识别旨在对无重叠视域的多摄像头中的行人或车辆目标进行匹配及检索，有助于追踪目标轨迹，实现对公共场所的智能安全监控。再识别模型通常应用在超大规模的数据库上，部署在不同的场景中，并面临着频繁的模型更新。这对模型表征的泛化力提出了巨大挑战。具体来说，在大规模数据库下，目标样本由于拍摄视角、姿态的变化呈现多样性，这要求表征具有泛化力和鲁棒性，以实现跨视角、跨摄像头搜索。 此外，多样的部署场景（域）间面临着背景、光照等采集环境的变化，这要求表征具有域间泛化力和适配性，以应对域间数据分布的差异。最后，模型的更新部署催生出模型表征间互操作的需求，这进一步要求表征具备模型间的泛化力和兼容性，以更好地进行跨模型协作。因此，本文为解决目标再识别在数据、应用和部署中所面临的挑战，提出了系统的表征学习方法，以提升表征在不同应用场景下的泛化力。本文的创新点主要包括以下几个方面：

1）针对目标样本多样性引发的表征泛化问题，本文提出了融入度量空间对抗机制的表征解耦学习方法。本方法首先引入对抗学习策略和注意力机制，解耦获得对目标视角变换鲁棒的视角共享表征和对相似目标极具区分力的视角特定表征。之后进一步设计了自适应的表征混合排序方法，针对不同的匹配对灵活地强化不同的表征，降低表征间的冗余性以提升表征泛化力。此外，本文构建了开放场景下的车辆再识别数据集 VERI-Wild 2.0，数据集采集自 274 个摄像头，包含了 686,525 张车辆图像，可以有效评估模型表征在大规模场景下的判别力和在不同摄像头间的泛化力。对比未施加解耦学习的基础模型，提出的方法在 VERI-Wild 2.0 数据集上取得了 5.13%mAP 的性能提升，并在跨视角检索任务中取得了 11.99% mAP 性能提升。

2）针对多样的部署场景引发的多域间泛化问题，本文提出了基于多重元学习的表征泛化提升方法，将元学习同时融入到训练策略和度量空间优化中，全面提升表征在不同场景下的泛化力和判别力。元学习训练策略在线地进行“先训练再泛化性评估”任务，模拟模型在未知场景下的表现，让模型“学会泛化”。进一步引入的元学习判别损失通过在度量空间优化中显式地模拟跨摄像头或跨域匹配场景，让模型“学会判别”。为了综合评估模型表征的泛化力，本文构建了极具多样性的 Person30K 数据集，数据集覆盖了超市和商场的室内外场景，拍摄自 6,497 个摄像头，包含了 3 万个行人的 138 万张图像。对比未加入元学习的基础模型，本方法在 Person30K Test-C 测试集上性能提升了 2.38% mAP。此外，数据集泛化性实验结果表明，相比 MSMT17 数据集，Person30K 上获得的模型表征具有更强的泛化力。

3）针对源域模型到目标域场景的无标签域适配问题，本文提出了高质量伪标签生成驱动的表征域适配方法，通过为无标签的目标域数据提供更可靠的伪标签，进而获得目标域适配力强、泛化力强的模型。本方法借助图结构，设计了层次化图卷积网络，逐步地获取样本间的关联性。首先使用基于顶点的图卷积

聚集视觉相似的样本并获得纯度较高的子簇，接着使用基于簇的图卷积关联同一目标极具多样性的子簇，最终获得高精度、高召回的伪标签。本方法进一步设计了一种关系表征，通过将源域数据作为基描述目标域图像，解决目标样本具有较大视觉差异时的关联问题。在DukeMTMC到Market1501的域适配实验中，基于MMT框架，当引入层次化图卷积后，伪标签质量和再识别模型性能分别提升了12.11% F-score和7.7% mAP。

4）针对模型间表征的泛化兼容问题，本文提出了一种基于迁移学习的表征兼容方法。再识别模型面临频繁的更新和部署，每当模型更新均需要重提数据库表征来保证表征的一致性，这带来了大量的时间和计算开销，是实际应用中的一个痛点问题。因此，本文设计了一种表征兼容学习方法，通过引入迁移学习，使得新模型表征可以直接与旧数据库表征匹配，实现模型表征间的互操作。本方法将迁移学习思想引入到了表征空间和模型空间，设计的代表特征兼容损失通过迁移代表特征显式地对齐新旧表征空间，进一步提出的网络组件互用的结构正则通过迁移网络组件中蕴含的规则信息隐式地实现表征兼容。在六个数据集上的实验表明，相比独立训练的模型，兼容模型可以取得相当的自测试性能（-0.3% ～+1.1% CMC@1），并能取得更好的交叉测试性能(+1.5% ～ +6.5% CMC@1)。

综上所述，本文针对不同的目标再识别应用场景，分析了所面临的表征泛化性挑战，并提出了系统的表征学习方法，用于提升表征的泛化能力。实验表明所提出的表征学习方法在不同的应用场景和任务中均表现出显著的性能优势，验证了提出方法的优越性。

Abstract

With the widely used video surveillance equipment, the processing, analysis, and utilization of person/vehicle data contain enormous social and scientific value. Object re-identification (ReID) is one of the core subjects of surveillance video technology, which is valuable for the industry and the research community. Object re-identification aims to match and retrieve pedestrian or vehicle targets in multi-cameras without overlapping fields of view. It helps to obtain target trajectories and realize intelligent and safe monitoring of public places. ReID models usually need to be applied to super-large databases, deployed in various application scenarios, and face frequent model updates. This poses significant challenges to the generalization ability of model representations. Specifically, in large-scale databases, target samples present diversity due to changes in shooting angles and postures, requiring representations to have generalization ability and robustness to achieve cross-view, cross-camera retrieval. In addition, different deployment scenarios (domains) face environment changes, requiring representations to have inter-domain generalization ability and adaptability to cope with different data distributions. Finally, the model faces frequent updates, therefore the representations need to have generalization ability and compatibility between models for better cross-model collaboration. In this paper, we propose a systematic generalizable representation learning solution for different re-identification

application scenarios. The main contributions of this paper include the following aspects.

1) For the representation generalization problem caused by the sample diversity, a disentangled representation learning with metric-based adversarial scheme is proposed. This method introduces an adversarial learning strategy and an attention mechanism to decouple common representations and specific representations. The common representations are robust against different variations, and the specific representations are highly discriminatory to similar targets. Moreover, to effectively use these two types of representations, we further design a hybrid ranking strategy to flexibly strengthen different representations for different matching pairs, reduce redundancy between representations, and enhance representation generalization. In addition, this article constructs a vehicle re-identification dataset, VERI-Wild 2.0, which is collected from 274 cameras and contains 686,525 vehicle images. This dataset can effectively evaluate the discriminative power and generalization ability of model representations in large-scale scenes. Compared with the basic model without disentangle learning, the proposed method achieves 5.13% mAP performance improvement on the VERI-Wild 2.0 dataset and 11.99% mAP performance improvement in the cross-view retrieval task.

2) For the domain generalization challenge caused by diverse deployment scenarios, this paper proposes a representation generalization method based on multiple meta-learning. It integrates meta-learning into both training strategy and metric space optimization, comprehensively improving the generalization and discriminative power of representations in different scenarios. The meta-learning training strategy performs an online "train-then-evaluate" task to simulate the performance of the model in unknown scenarios, enabling the model to "learn to generalize". The introduced meta-learning discrimination loss explicitly simulates cross-camera or cross-domain matching scenarios in metric space optimization, enabling the model to "learn to discriminate". To comprehensively evaluate the generalization power of the model's representations, we construct a highly diverse dataset, Person30K, which covers indoor and outdoor scenes in supermarkets and shopping malls, and contains 1.38 million images captured from 6,497 cameras. Compared with the baseline model without meta-learning scheme, this method achieves 2.38% mAP improvement on the Person30K Test-C dataset. Furthermore, in the dataset diversity analysis experiment, the model representations obtained on Person30K exhibit stronger generalization power than those on the MSMT17 dataset.

3) For the unsupervised domain adaptation problem, this paper proposes a domain adaptation method driven by high-quality pseudo label generation. By providing reliable pseudo labels for the unlabeled data of target domain, we can obtain a model with adaptation and generalization abilities on the target domain. This method leverages graph structures and designs a hierarchical graph convolutional network to gradually obtain the correlations between samples. Firstly, vertex-based graph convolution is used to aggregate visually similar samples and obtain high-purity sub-clusters, then cluster-based graph convolution is used to associate sub-clusters with diverse targets,

and finally high-precision and high-recall pseudo labels can be obtained. Besides, a robust relation representation is proposed for clustering. It leverages the labeled source domain samples as references, which can alleviate the impact of intra-person variation in target domain. In the domain adaptation experiment from DukeMTMC to Market1501, by adding hierarchical graph convolution to the MMT framework, the quality of pseudo labels and the performance of the re-identification model are improved by 12.11% in F-score and 7.7% in mAP, respectively.

4) To address the problem of representation generalization between models, this paper proposes a representation compatibility method based on transfer learning. Re-identification models face frequent updates and deployments. It is a heavy workload to re-extract representations of the whole database every time. Therefore, this paper designs a representation compatibility learning method by introducing transfer learning, which enables the new model representation to directly match the old database representation and achieve interoperability between model representations. This method introduces transfer learning into the representation space and model space. The designed prototype-based compatible loss uses prototypes to bridge and align the new and old embedding representations explicitly. The mutual structure regularization is further designed to implicitly achieve representation compatibility by transferring the rule information contained in the network components. The experiments on six datasets show that compared with independently trained models, the compatible models can achieve comparable self-testing performance (-0.3% ~ +1.1% in CMC@1) and better cross-testing performance (+1.5% ~ +6.5% in CMC@1).

To sum up, this paper analyzes the challenges of representation generalization for different re-identification application scenarios, and proposes a systematic representation learning method to improve the generalization ability of representation. Experiments demonstrate that the proposed representation learning method shows significant performance advantages in different application scenarios and tasks, which verifies the superiority of the proposed method.