

基于大模型的公共安全监控文本到图像行人检索研究

摘要

随着智慧城市建设的不断发展，文本到图像行人检索在公共安全监控中的应用价值日益凸显。针对现有行人检索方法对大量标注数据的依赖及其在跨场景部署中的泛化瓶颈，本研究提出一种基于图结构的跨域知识蒸馏方法（GCKD），实现大模型驱动下的无监督跨域检索能力。该方法通过图结构多域传播与对比式动量蒸馏模块，解决了跨场景语义迁移与模态差异问题。实验在多个行人检索基准数据集上进行了验证，平均 Rank-1 检索精度提升超过 4%。研究成果已被人工智能领域国际顶级会议、中国计算机学会推荐 A 类会议 AAAI 2025 接收。

关键词

文本到图像行人检索；多模态大模型；智慧城市；公共安全

Abstract

With the advancement of smart city development, text-to-image person retrieval has shown increasing value in public security surveillance applications. To address the reliance of existing retrieval methods on large-scale annotated data and their limited generalization across diverse deployment scenarios, this study proposes a Graph-Based Cross-Domain Knowledge Distillation (GCKD) method, enabling large-model-driven unsupervised cross-domain retrieval. The proposed approach leverages a graph-based multi-domain propagation module and a contrastive momentum distillation module to tackle the challenges of semantic transfer and modality discrepancy across scenarios. Experiments on multiple benchmark person retrieval datasets demonstrate that our method achieves an average improvement of over 4% in Rank-1 retrieval accuracy. This work has been accepted as an oral presentation at AAAI 2025, a top-tier international conference in artificial intelligence and a CCF-A recommended venue.

Keywords

Text-to-Image Person Retrieval; Multimodal Large Model; Smart City; Public Security

Author: 骆, 炳君

Presenter: 骆, 炳君

Session Classification: AI+

Track Classification: 03 口头报告: AI+