



# 行人重识别的领域现状与最新进展

2022 年《计算技术前沿》课程报告

讲座题目	智能视频监控关键技术：行人再识别		
主讲人	赵才荣		
所在学院	电子信息与工程学院		
专    业	计算机科学与技术系		
学生姓名	包广垠	学 号	2230771
联系方式	19921315546		

### 摘要

随着社会和经济的快速发展，城市公共安全问题受到越来越多的关注。视频监控作为保障城市安保的重要技术手段，被广泛应用于街道、学校、商场等人流密集的公共场所。城市视频监控网络每时每刻都在获取视频数据，监控视频的数据量十分庞大。然而，目前对于监控视频的处理主要由“人工分析”来完成，这导致了巨大人力资源的开销，也导致了诸如“视频在但找不到”、“找得到但找太久”、“有服务但不可靠”等视频监控技术的应用瓶颈。在这样的情况下，如何实现视频监控的智能处理，尤其是对大规模行人数据的智能处理和可疑行为的自动研判，是新时代公共安全领域的迫切需要。

行人重识别是指在多个时空不重叠的摄像头下寻找目标人物的过程，是为了解决跨摄像头跨场景下目标行人的关联与匹配问题，是智能监控系统的核心技术。随着深度神经网络的发展和智能视频监控需求的增加，它在计算机视觉领域引起了极大的兴趣。

本文是同济大学研究生课程《计算技术前沿》的课程报告，基于赵才荣教授的讲座，聚焦于行人重识别领域，梳理该领域的研究现状和关键技术。

**关键词:** 行人重识别, 目标检测, 计算机视觉, 智能信息处理

## ABSTRACT

With the rapid development of society and economy, the issue of urban public safety has received more and more attention. As an important technical means to ensure urban security, video surveillance is widely used in crowded public places such as streets, schools, and shopping malls. The urban video surveillance network acquires video data all the time, and the amount of surveillance video data is huge. However, at present, the processing of surveillance video is mainly done by "manual analysis", which leads to huge human resources overhead, and also leads to problems such as "the video is there but the person can't be found", "person can be found but it takes too much time", "some service but unreliable" and other video surveillance technology application bottlenecks. Under such circumstances, how to realize the intelligent processing of video surveillance, especially the intelligent processing of large-scale pedestrian data and the automatic judgment of suspicious behaviors, is an urgent need in the field of public security in the new era.

Pedestrian re-identification refers to the process of finding target persons under multiple cameras that do not overlap in time and space. The purpose is to solve the problem of correlation and matching of target pedestrians in cross-camera and cross-scene situations. It is the core technology of intelligent surveillance systems. With the development of deep neural networks and the increasing demand for intelligent video surveillance, it has attracted great interest in the field of computer vision.

This report is a course report of Tongji University's postgraduate course "Frontiers of Computing Technology". Based on Professor Zhao Cairong's lecture, I will focus on the field of pedestrian re-identification, and sorts out the research status and key technologies in this field.

**Keywords:** person re-identification, object detection, computer vision, intelligent information processing

## 目录

1	引言	1
1.1	行人重识别的任务描述	1
1.2	标注行人重识别系统的构建步骤	1
1.3	标准行人重识别与开放行人重识别	2
2	数据集与评价标准	3
2.1	标准行人重识别的数据集	3
2.2	开放行人重识别的数据集	3
2.3	行人重识别的评价标准	4
3	标准行人重识别算法	5
3.1	基于特征表示学习的行人重识别	5
3.1.1	全局特征的表示学习	5
3.1.2	局部特征的表示学习	5
3.2	基于深度度量学习的行人重识别	5
3.2.1	分类损失	6
3.2.2	验证损失	6
3.2.3	三元组损失	7
3.3	重排序优化	8
4	开放行人重识别	9
4.1	遮挡行人重识别	9
4.1.1	遮挡情况下的特征表示方法	9
4.1.2	遮挡情况下的特征相似度度量方法	9
4.2	多模态行人重识别	9

4.2.1	深度图像和 RGB 图像间的行人重识别 .....	9
4.2.2	文本和图像间的行人重识别 .....	10
4.2.3	红外图像和 RGB 图像的行人重识别 .....	10
4.2.4	交叉分辨率的行人重识别 .....	10
4.3	行人搜索 .....	10
4.4	噪声鲁棒的行人重识别 .....	11
5	总结 .....	12
	参考文献 .....	13

装

订

线

## 1 引言

### 1.1 行人重识别的任务描述

行人重识别是针对时空非重叠摄像机下的特定人物的检索问题。给定一个待查询的人，行人重识别的目标是判断待查询人是否在不同时刻出现在不同相机中，或者是在同一时刻出现在不同相机里，或者是在同一个相机在不同的时刻出现。待查询人可以由图像、视频甚至文本描述来表示。由于公共安全的迫切需求和监控摄像头数量的不断增加，行人重识别技术在智能监控系统中愈发重要，因而该技术具有重大的研究意义和实用价值。

行人重识别的任务如图1.1所示。给定一个待查询人（通常叫做 **Query**），我们需要从搜索集（通常叫做 **Gallery**）中找出与 **Query** 表示的是同一个人的那些图片，并按照相似度进行排序。



图 1.1: 行人重识别的任务描述

由于在模型的训练过程中，行人重识别会遇到诸多问题，例如相机的视角不同、图像或视频的分辨率不同、光照条件不同、行人姿态无约束、遮挡、人物更换衣物、多模态异构、复杂的相机环境、背景杂波和不可靠的边界框生成等问题，因而这是一项极具挑战性的任务。此外，模型部署过程中还会遇到摄像机网络的动态更新、检索集规模庞大、域偏移、测试场景的不可见性、模型在线更新等问题，这进一步增加了问题的难度。

### 1.2 标注行人重识别系统的构建步骤

通常而言，构建行人身份重识别系统包括以下五个步骤：

1. 数据收集：从监控摄像机获取原始视频数据。
2. 边界框生成：提取 **Bounding Box** 边界框，每个框中的内容是原始视频数据中的人物图像。

3. 数据注释：对训练数据进行标注。
4. 模型训练：使用边界框和标注训练一个具有较高性能和鲁棒性的行人重识别模型。
5. 行人检索：进行行人检索，即给定一个待查询人和一个查询集，使用前一阶段训练的模型提取特征表示，然后计算出待查询人与图库中图片的相似度，并进行排序。

## 1.3 标准行人重识别与开放行人重识别

行人重识别根据任务类型的细微差异主要可以分为两类——标准行人重识别和开放行人重识别，它们的主要区别如下：

1. 单模态与多模态：标准行人重识别中的所有数据都由单模态可见摄像机在封闭世界环境中拍摄的图像或视频。而开放行人重识别还需要处理多模态异构数据，包括红外图像、草图、深度图像甚至是文本描述。这使得人们研究多模态行人重识别。
2. 是否有边界框：标准行人重识别通常基于生成的边界框来进行训练和测试。而开放行人重识别需要在原始图像或视频上进行端到端的行人重识别。这使得人们研究行人搜索问题。
3. 数据标注的完整性：标准行人重识别有足够的被标注的训练数据用于模型的监督训练。而开放行人重识别假设没有足够的标注数据，甚至没有任何标注信息。这使得人们研究半监督和无监督行人重识别。
4. 数据标注的正确性：标准行人重识别的所有标注都是正确的。而开放行人重识别假设标注数据中存在错误标注的样本，即噪声样本。这使得人们研究噪声鲁棒行人重识别。
5. 待查询人是否存在于查询集：标准行人重识别的所有待查询均在查询集中。然而开放行人重识别的待查询人不一定出现在查询集中，即查询集是开放集。这使得人们研究开放集行人重识别。

## 2 数据集与评价标准

算法和模型的搭建都基于数据集来进行，而评价标准是对算法好与坏的评估，因此我先调研了行人重识别领域的常用数据集和评价标准。

### 2.1 标准行人重识别的数据集

目前常用的标准行人重识别的图像数据集有 VIPeR, CUHK03, Market-1501, DukeMTMCreID, MSMT17, 如2.1所示。

数据集	行人个数	图像个数	提出时间
VIPeR	632	1264	2007
CUHK03	1467	28192	2014
Market-1501	1501	32688	2015
DukeMTMCreID	1404	34183	2017
MSMT17	4101	126441	2018

表 2.1: 标准行人重识别的数据集

VIPeR 是较早公开的行人重识别数据集，创建于 2007 年，数据集中涉及大幅度姿态变化以及轻微光照变化。CUHK03 是香港中文大学于 2014 年发布的数据集，由 6 个不同监控摄像机拍摄收集，每个行人都至少出现在两个摄像视角下，该数据集中存在光照、遮挡、身体部分不匹配等环境因素造成的影响。Market-1501 数据集采集自清华大学的校园中，包括 6 个摄像头（5 个高清摄像头和 1 个低清摄像头），数据集中的行人边界框由 DPM 自动检测得到。DukeMTMC-reID 是大规模多目标跟踪数据集 DukeMTMC 中的一个子集，采集自杜克大学校园内，于 2016 年公布。MSMT17 是目前最大规模图像行人重识别数据集之一，由 12 个室外摄像机和 3 个室内摄像机在不同时段、不同气候条件下拍摄得到，该数据集中的行人边界框由 Faster RCNN 检测得到。

### 2.2 开放行人重识别的数据集

开放行人重识别数据集主要是为了解决遮挡、行人搜索、跨模态等问题。如2.2所示。

针对遮挡问题，有 Partial-REID, Occluded-REID, Occluded-Duke, P-ETHZ, P-DukeMTMC 等数据集。其中，Partial-REID 采集自大学校园中，每个行人有 5 张全身图片和 5 张遮挡图片，该数据集存在复杂的视角、背景变化和严重遮挡情况。Occluded-REID 由校园内的移动摄像设备拍摄，每个行人由 5 张全身图片和 5 张不同遮挡图片组成。Occluded-Duke 由 DukeMTMC-reID 中的遮挡图像组成，是迄今为止最大的遮挡行人数据集，该数据集存在多种视角变化和丰富的遮挡类型。

针对行人搜索，常用的两个数据集是 CUHK-SYSU 和 PRW 数据集。CUHK-SYSU 收集自手持摄像机拍摄的街景和电影中的场景。PRW 采集自大学校园内的 6 台摄像设备。

在通过自然语言描述查询图库行人的文本-图像行人重识别方面，CUHK-PEDES 是该领域唯一



数据集	行人个数	图像个数	提出时间	特性
Partial-REID	60	600	2015	遮挡
Occluded-REID	200	2000	2018	遮挡
Occluded-Duke	1812	35489	2019	遮挡
P-ETHZ	85	3897	2008	遮挡
CUHK-SYSU	8432	18184	2017	行人搜索
PRW	932	11816	2017	行人搜索
CUHK-PEDES	13003	40206	2017	多模态（文本）
SYSU-MM01	491	303430	2017	多模态（红外）
RegDB	432	8640	2017	多模态（红外）

表 2.2: 存在遮挡的行人重识别的数据集

的数据集, 相关研究人员从 CUHK03, Market-1501, SSM, VIPER 和 CUHK01 数据集中挑选图像, 每张图像都有两段自然语言描述。

可见光-红外行人重识别的两个常用数据集分别是 SYSU-MM01 和 RegDB。SYSU-MM01 包含 6 个摄像头, 其中 4 个是在白天工作的可见光摄像头, 2 个是在黑暗环境中工作的红外摄像头。RegDB 数据集中每个行人包含 10 张可见光图像和 10 张红外图像, 且只有正面和背面两种拍摄视角。

## 2.3 行人重识别的评价标准

### (1) mAP

mAP 是行人重识别领域最常用的评价指标。

对于每张图片中的每个预选框, 将其与 IOU 值最大的 GT 框相对应起来, 若 IOU 值大于 0.5 且没有 IOU 更大的框与之对应, 则该预选框是一个正样本; 若 IOU 值小于 0.5 或有其他 IOU 值更大的预选框与之对应, 则该预选框是一个负样本。根据候选框分类置信度与给定阈值的关系进行预测:

TP——置信度大于阈值的正样本

TF——置信度大于阈值的负样本

FN——置信度小于阈值的样本或者没有候选框对应的 GT

然后, 计算准确率  $P$  与召回率  $R$ :

$$P = \frac{TP}{TP+TF} \quad R = \frac{TP}{TP+FN}$$

在不同的阈值下得到不同的准确率和召回率, 从而绘制 P-R 曲线, 曲线下的面积对应的类别的 AP 值。计算所有类别的 AP, 取平均数就得到 mAP 值。

### (2) top-k 准确率

top-k 准确率是行人重识别领域另一个常用的评价指标。top-k 指的是将待匹配图片与 k 个最有把握的图片相匹配, 只要 k 个中有一个匹配成功, 则认为匹配成功, 一般而言 top-1 使用较多。

### 3 标准行人重识别算法

标准行人重识别算法是在理想情况下进行的重识别，仅考虑少量光照、姿态、视角变化因素，而将更多注意力放在图像自身的特征表达和相似性度量上。这一领域的研究主要包含三个方面：特征表示学习、深度度量学习和重排序优化。

#### 3.1 基于特征表示学习的行人重识别

特征表示学习聚焦于对行人特征的表示，希望能够提取出对光照、姿态、视角等变化不敏感的特征。

##### 3.1.1 全局特征的表示学习

全局特征提取是指将输入的图像作为一个整体进行特征的提取。在行人重识别算法发展之初，这种全局特征提取的方法比较多。它是将整个图片不加处理输入特征提取网络，得到图片对应的特征向量，然后通过比较不同图片所提取的特征向量的相似度来进行行人的匹配与重识别。

由于全局特征提取的方法提取到的特征粒度比较粗，对于行人重识别任务而言，这样的粗粒度特征不利于精细的行人匹配任务，因而早期的行人重识别算法的精度并不高。

##### 3.1.2 局部特征的表示学习

局部特征的表征学习旨在提取图像的细粒度特征。由于卷积神经网络中的卷积操作本质上是提取图像的局部相关性，因此细粒度的局部特征提取会更加有效。此外，进行局部特征的表示学习会降低模型对于遮挡、视角和姿态的敏感度，也有利于提高标准行人重识别的准确度。

目前基于局部特征表征学习的行人重识别算法主要有三类，分别是基于水平分块对齐、基于姿态估计、基于注意力机制。如3.1所示。

#### 3.2 基于深度度量学习的行人重识别

度量学习方法一般通过网络学习出多张图片的相似度，并根据图像标签与相似度的差异构建损失。在行人重识别问题上，度量学习表现为同一行人的不同图片间的相似度大于不同行人的不同图片。度量学习常用的损失有三类，分别是分类损失、验证损失和三元组损失，对应分别使用单张图像、图像对、图像三元组作为网络的输入来构建损失。

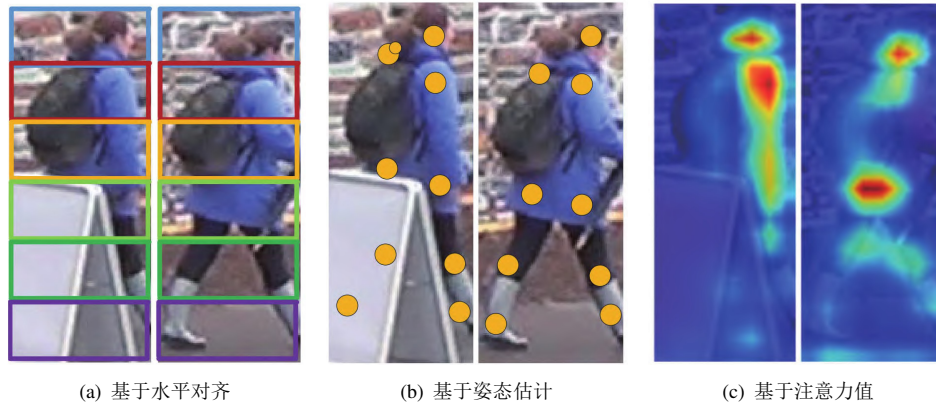


图 3.1: 局部特征的表示学习

### 3.2.1 分类损失

分类损失如图3.2所示，它将行人重识别任务视为图像分类问题，通过将网络提取的高层语义特征输入全连接层得到分类结果，根据该分类结果计算交叉熵损失：

$$L_{id}(i) = \frac{1}{n} \sum_{i=1}^n \log(p(y_i)|x_i)$$

该类损失通过最大化样本真实类别的后验概率以聚类方式对样本的间距进行优化, 以簇的形式对样本进行聚类分割, 从而优化不同样本之间的距离。该损失在训练过程中会自动挖掘出难分类样本, 且易于与其他损失函数结合使用从而进一步优化判别空间，因此被广泛应用于行人重识别任务中。

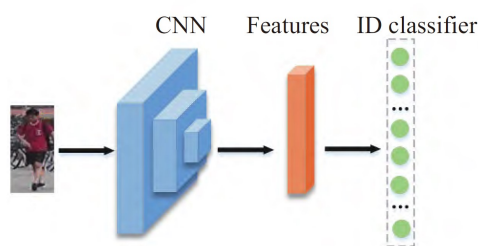


图 3.2: 度量学习中的分类损失

### 3.2.2 验证损失

验证损失如图3.3所示，它基于相似样本在特征空间中仍然相似的假设，利用图像间的成对关系将检索问题转化为验证问题。

当两个输入样本相似时，对比损失可以有效地对细节进行建模，即对这两个样本的差异性进行度量，主要用于处理行人重识别中的光照及几何变换问题。基本定义如下：

$$L_{con}(i, j) = (1 - \delta_{ij}) \{ \max(0, p - \delta_{ij}) \}^2 + \delta_{ij} d_{ij}^2$$

在行人重识别任务中，对输入样本得到的特征作差的平方结果进行分类，判断该结果的正负。结合交叉熵的二分类验证损失公式表示如下：

$$L_{veri}(i, j) = -\delta_{ij} \log p(\delta_{ij}|f_{ij}) - (1 - \delta_{ij}) \log(1 - p(\delta_{ij}|f_{ij}))$$

其中， $\delta_{ij}$  定义与对比损失中相同，为相似度标签，而  $f_{ij} = (f_i f_j)^2$ ，其中  $f_i$ ， $f_j$  分别代表样本  $i$  和  $j$  的特征，而  $p(\delta_{ij}|f_{ij})$  代表  $f_{ij}$  被分类为  $\delta_{ij}$  的概率。

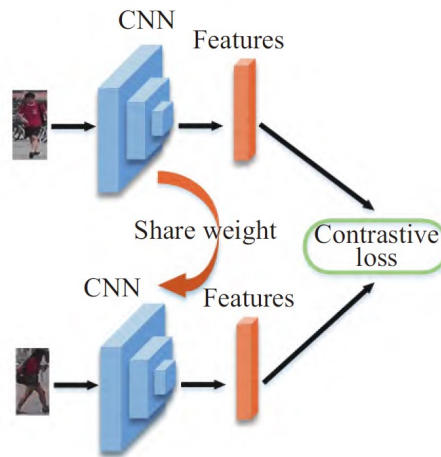


图 3.3: 度量学习中的验证损失

### 3.2.3 三元组损失

三元组损失如图3.4所示，它将行人再识别任务当作检索排序问题进行处理，是一种被广泛应用于行人重识别领域的损失函数。其构造的三元组中包含锚样本  $X_i$ ，与  $X_i$  类别相同的正例样本  $X_i^+$ ，以及与  $X_i$  类别不同的负例样本  $X_i^-$ ，即一个三元组包含一对正样本对及一对负样本对，进一步引入预定义的距离阈值参数后，三元损失的函数形式如下：

$$L_{tri} = (d_{a,p} - d_{a,n} + \alpha)_+$$

将三元组损失引入行人重识别任务，并且将阈值引入三元组损失的定义中，使得三元组损失不

仅仅依赖于类内距离小于类间距离这一简单约束，并针对随机映射产生三元组造成生成较多无用数据的情况，提出在每次迭代过程中，随机选取一定数量的行人身份，根据该身份信息从数据集中选取对应类别的特定数量的行人图像作为锚样本，进行随机三元组构造，解决了模型优化效率过低的问题。

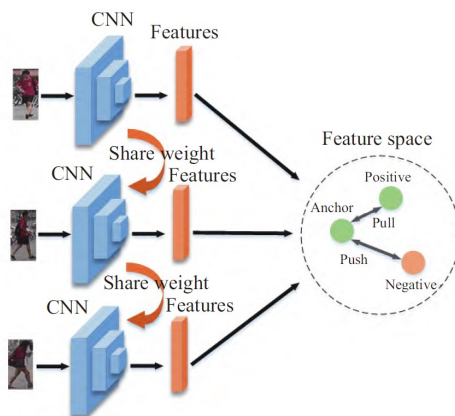


图 3.4: 度量学习中的三元组损失

### 3.3 重排序优化

如图3.5所示，重排序优化是行人重识别的一种后处理手段,它利用图片数据集的上下文信息和相似度信息对检索结果进行优化，从而进一步提升检索精度。

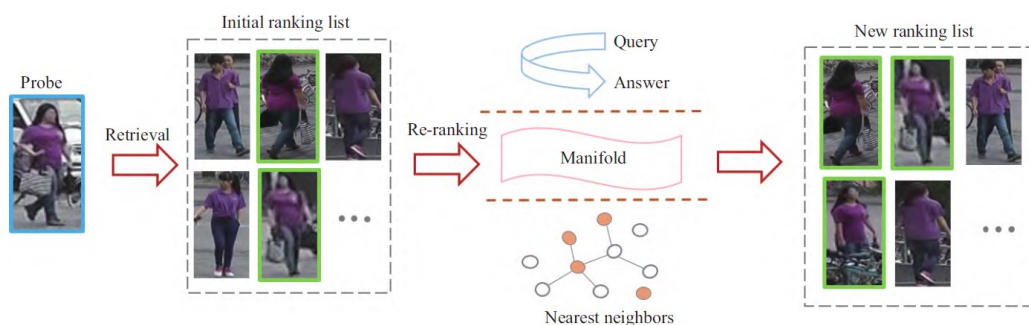


图 3.5: 行人重识别中的重排序优化

现有的重排序优化有三种方法，分别是基于用户反馈的重排序、基于候选集流形结构的重排序和基于近邻信息挖掘的重排序。

重排序算法是一种提升检索精度的有效手段，但是其往往大量的参数计算，使得模型的推理速度下降，甚至影响模型的稳定性。

## 4 开放行人重识别

开放环境下的行人重识别旨在解决真实情况中的行人重识别问题，在这个过程中会遇到诸多挑战，例如遮挡问题、多模态问题、噪声问题等。解决好开放世界的行人重识别问题是将算法运用于实际生产中所必须完成的挑战。

### 4.1 遮挡行人重识别

遮挡行人重识别主要是在行人部分信息被遮挡的情况下，研究有效的特征表示方法以及合理的相似性度量方法。

#### 4.1.1 遮挡情况下的特征表示方法

遮挡行人的特征描述有以下几种方法：

基于行人身体注意力的特征表示方法通过遮挡模拟器自动为一般行人图像生成人工遮挡，然后设计多任务损失训练网络，强化其对遮挡行人图像特征的学习和行人身份的识别。基于姿态信息的方法设计姿态引导网络来提取人体有效区域特征，即姿态引导特征对齐。

#### 4.1.2 遮挡情况下的特征相似度度量方法

考虑到不同图像遮挡区域的差异性，实现特征之间有效的对齐和度量十分重要。

局部-局部特征匹配与全局-局部特征匹配优化从多粒度的角度实现遮挡行人之间的有效度量；隐式对齐的深度空间特征重构方法利用字典学习模型的重构误差计算遮挡图像不同空间特征图之间的相似性，从而实现鲁棒表征；基于半监督学习的可见性感知的局部模型通过预测行人不同部位的可见性分数判断对应部位是否被遮挡，实现不同遮挡行人的区域特征自适应对齐和度量等。

目前对遮挡行人重识别的研究工作仍处在探索阶段，现有的方法在遮挡行人数据集上的性能与实际应用仍有一定差距。因此，遮挡场景下的行人重识别问题仍需系统性研究，为实际场景中行人身份识别问题提供解决方案。

### 4.2 多模态行人重识别

多模态行人重识别主要包含四个方面：深度图像和 RGB 图像间的行人重识别，文本和图像间的行人重识别，红外图像和 RGB 图像的行人重识别以及交叉分辨率的行人重识别。

#### 4.2.1 深度图像和 RGB 图像间的行人重识别

深度图像捕捉身体形状和骨骼信息，这使得在照明不同或者换衣服的情况下仍可能进行行人重识别。

基于递归注意力的模型将卷积神经网络和递归神经网络相结合，以识别人体的有区别的小型局部区域；RGB 到深度图像的转换方法弥补了深度图像和 RGB 图像之间的差距；RGB 和深度信息的组合提高了深度行人重识别的性能，解决了换衣挑战。

## 4.2.2 文本和图像间的行人重识别

文本和图像间的行人重识别解决文本描述和 RGB 图像之间的匹配。当无法获得查询人的视觉图，并且只能替代地提供文本描述时，这是有必要的。

具有递归神经网络的门控神经注意力模型学习文本描述和人物图像之间的共享特征，实现了文本到图像的端到端搜索；全局鉴别图像语言关联学习方法在重建过程中捕获身份鉴别信息和局部重建图像语言关联；交叉投影学习方学习具有图像到文本匹配的共享空间。

然而，文本描述和视觉图像之间的巨大语义差距仍然给问题提供了巨大的挑战性。同时，如何将文本与手绘素描图像相结合也是值得研究的问题。

## 4.2.3 红外图像和 RGB 图像的行人重识别

红外图像和 RGB 图像的行人重识别旨在处理日间 RGB 图像和夜间红外图像之间的跨模态匹配问题，这在低光照条件下非常重要。

深度零填充框架可以自适应地学习模态共享特征，是解决这个问题的第一次尝试；双流网络以对模态可描述和特定信息进行建模，同时处理模态内和跨模态变化；GAN 技术生成跨模态人物图像，以减少图像和特征水平上的跨模态差异；双重注意聚合学习方法可以捕获多层次关系，从而提高红外跨模态问题的精度。

## 4.2.4 交叉分辨率的行人重识别

交叉分辨率行人重识别旨在低分辨率图像和高分辨率图像之间进行匹配，要解决大分辨率变化问题。

级联 SR-GAN 以级联的方式生成高分辨率人物图像，其中包含身份信息，从而减小模态间的差异；抗性学习技术可以获得分辨率不变的图像表示，从而使得匹配能够更加准确。

## 4.3 行人搜索

行人搜索将行人检测和行人重识别任务联合起来，该任务只需输入原始场景图像，即可完成对目标行人的查找和匹配。与基于裁剪图片的行人重识别相比，行人搜索更符合现实场景下的视频监控需求，即从原始的监控场景中查询目标行人。行人搜索的方法大致可分为两类：将检测和识别联合起来的端到端方法和先检测再识别的两阶段方法。

端到端的行人搜索是基于 Faster R-CNN，在此基础上在线实例匹配机制被提出，它代替 Softmax 来计算识别过程中的损失。图卷积学习框架提出了上下文扩展模块，使用相对注意力机制来搜索和过滤场景中的上下文信息。NAE 网络考虑在极坐标系中分解行人向量，通过向量范数区分背景和行人，通过向量角度实现行人身份识别。序列端到端网络将检测和重识别任务视为一个渐进的过程，并依次处理两个子网络。此外，由于基于 Faster R-CNN 虽然拥有很高的性能，但是计算开销太大，基于无锚框行人搜索框架 AlignPS 被提出，它利用可变形卷积和特征融合来改进特征金字塔网络，来克服行人区域和尺度不对齐问题。

两阶段行人搜索不同于联合行人检测和行人重识别的端到端方法，该方法认为：检测和识别两个任务之间存在冲突，即检测关注行人的共性，而识别关注行人之间的差异，因此端到端框架并不是一个好的解决办法。通过计算检测器检测出的行人框和查询图片的相似度，去除相似度低的框，可以为后续行人识别任务提供数量更少、精确度更高的行人框。此外，将人工裁剪的行人图像和检测器检测出的行人图像混合，共同用于行人识别网络的训练，可以使行人识别网络更加适应检测器。

现有的行人搜索方法中，端到端行人搜索注重于检测识别一体化实现，将行人重识别网络嵌入检测网络中，再使用各自的损失函数进行联合优化。相比之下，两阶段行人搜索先做检测重做识别，这种两阶段方案可以克服两种不同优化目标带来的冲突，即行人检测关注行人之间的共性而行人重识别关注行人之间的差异。

#### 4.4 噪声鲁棒的行人重识别

由于数据收集和注释困难，行人重识别通常会遇到噪声。主要包含两个方面：检测或跟踪错误导致的样本噪声、注释错误导致的标签噪声。

为了处理人物图像内的外围区域或背景杂波，姿态估计和注意力线索被广泛运用。其基本思想是抑制噪声区域在最终整体表示中的贡献。

由于注释错误，标签噪声通常是不可避免的。标签平滑技术可以避免标签过度填充问题，针对特征不确定性进行建模的分布网络可以用于针对标签噪声的鲁棒模型学习，减少了具有高特征不确定性的样本的影响。

此外，未知的新身份增加了噪声鲁棒的行人重识别模型的学习难度。



## 5 总结

本次学术报告以行人重识别技术作为切入点，介绍了这一问题的研究背景、数据集与评价标准和研究现状。在研究现状方面，本文将行人重识别分为标准行人重识别和开放行人重识别来分别介绍。在标准行人重识别问题上，主要介绍表征学习和度量学习的一般方法，还介绍了重排序优化的方法来提升重识别性能；在开放行人重识别问题上，本文关注于遮挡、跨模态、行人搜索和噪声鲁棒性四个问题，阐述了这些问题的定义，并且回顾了针对这些问题的一些解决方案。

目前，标准行人重识别研究日趋成熟，其性能在多个行人重识别数据集上已经很高。但是开放行人重识别研究仍然面临诸多挑战。行人重识别领域的未来研究方向可以从以下三个方面来进行：

1. 开放世界行人重识别：目前，开放场景下的行人重识别面临诸多困难，如光照、遮挡、衣着改变等。而现有的研究都是只针对单一的问题，这和实际应用还有很大的差距。因此，将不同开放问题的行人重识别方法集成到同一个模型中是一项待解决的任务。
2. 多模态信息融合：目前，行人重识别大多是建立在 RGB 图像上的，而跨模态问题也只是为了解决跨模态的匹配问题。然而单一模态数据易受环境因素干扰，因此如何进行多模态的数据融合还有待深入研究。
3. 半监督与无监督：在实际的视频监控系统中，图片数据往往没有行人标签信息，因此很难直接用来进行监督训练。近年来，研究利用无标签数据的无监督和半监督方法逐渐兴起，促进了行人重识别实际应用，然而现有的方法训练出来的模型性能距离有监督方法还存在较大差距。

## 参考文献

- [1] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.
- [2] LEE K, JANG I S, KIM K J, et al. Reet: Region-enhanced transformer for person re-identification[C]// 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2022: 1-8.
- [3] WANG X, CORDOVA R S. Global and part feature fusion for cross-modality person re-identification [J]. IEEE Access, 2022, 10: 122038-122046.
- [4] LODDO A, MELONI G, PES B. Using artificial intelligence for covid-19 detection in blood exams: A comparative analysis[J]. IEEE Access, 2022, 10: 119593-119606.
- [5] LI Y, ZHANG T, LIU X, et al. Visible-infrared person re-identification with modality-specific memory network[J]. IEEE Transactions on Image Processing, 2022, 31: 7165-7178.
- [6] ZHENG X, CHEN X, LU X. Visible-infrared person re-identification via partially interactive collaboration[J]. IEEE Transactions on Image Processing, 2022, 31: 6951-6963.
- [7] YANG S, KANG B, LEE Y. Sampling agnostic feature representation for long-term person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 6412-6423.
- [8] NGUYEN H A, NGUYEN H Q, NGUYEN T B, et al. Exploiting matching local information for person re-identification[C]//2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR). 2022: 1-6.
- [9] RAO H, LI Y, MIAO C. Revisiting  $\chi^2$ -reciprocal distance re-ranking for skeleton-based person re-identification[J]. IEEE Signal Processing Letters, 2022, 29: 2103-2107.
- [10] HUI Y, WU X, HU X, et al. Unsupervised cross-domain person re-identification method based on attention block and refined clustering[J]. IEEE Access, 2022, 10: 105930-105941.
- [11] YU F, JIANG X, GONG Y, et al. Conditional feature embedding by visual clue correspondence graph for person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 6188-6199.
- [12] MAO Z, WANG X, XU X, et al. Continuous and unified person re-identification[J]. IEEE Signal Processing Letters, 2022, 29: 1983-1987.