密级：　　保密期限：

北京郵電大學

博士学位论文

题目：智能视觉监控中行人再识别技术研究

学　　号：2014010093

姓　　名：四建楼

专　　业：信息与通信工程

导　　师：张洪刚

学　　院：信息与通信工程学院

二〇一八年四月十六日

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：_____

关于论文使用授权的说明

本人完全了解并同意北京邮电大学有关保留、使用学位论文的规定，即：北京邮电大学拥有以下关于学位论文的无偿使用权，具体包括：学校有权保留并向国家有关部门或机构送交论文，有权允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，有权允许采用影印、缩印或其它复制手段保存、汇编学位论文，将学位论文的全部或部分内容编入有关数据库进行检索。（保密的学位论文在解密后遵守此规定）

本学位论文不属于保密范围，适用本授权书。

本人签名：_____ 日期：_____

导师签名：_____ 日期：_____

智能视觉监控中行人再识别技术研究

摘 要

行人再识别（Person Re-Identification, ReID）是智能视觉监控系统中的关键技术，其作用是基于视觉信息实现不同摄像头下行人目标的身份一致性关联，将单摄像头视频监控扩展为多摄像头协同视频分析。本文在大量国内外研究成果的基础上，结合对 ReID 所面临主要挑战的分析以及我们对特定场景下 ReID 问题的认识，分别从小数据集上的度量学习模型泛化能力、基于手工特征设计和多特征融合的分步处理模型、基于特征序列提取和序列匹配的端到端处理模型三个方面，提出不同的高精度 ReID 算法或者改进算法。

从小数据集上的度量学习模型泛化能力的角度，我们提出利用正则化的度量学习方法来增强小数据集上 ReID 模型的泛化能力。众所周知，度量学习的研究在 ReID 技术发展过程中扮演着重要的角色。然而，受某些应用场景的限制，研究人员往往无法获取充足的标记样本进行模型训练和学习，从而导致 ReID 算法的泛化能力较弱。为此，我们从限制模型复杂度的角度，提出利用正则化的度量学习算法实现 ReID 中的特征距离度量，从而提升小数据集上模型的泛化能力。具体来讲，我们分别从马氏距离学习、对称投影学习、以及非对称投影学习三个不同的角度理解度量函数，并构造了四种不同的正则化度量学习模型来实现 ReID。在数据集 VIPeR 和 CUHK01 上的实验验证了，正则化的模型约束，往往可以带来整体算法性能的提升。

从基于手工特征设计和多特征融合的分步处理模型的角度，我们提出了一种统一的局部统计特征提取框架，并结合多核学习在度量学习阶段实现 ReID 中的多特征融合。在监控场景中实现跨摄像头视域的行人匹配是一项极富挑战的任务，因为不同拍摄角度、不同行人姿态、不同光照条件、以及局部遮挡等都会引起行人外观的剧烈变化，从而增大匹

配难度。目前，大量的研究工作主要集中在，构造优秀的特征表示或者学习合理的特征匹配模型这两个方面。然而，由于影响行人外观的因素众多，很难构造单一特征来全面地刻画行人外观；而且，不同特征的提取过程相对独立，缺乏系统而详细地评估分析，很难启发研究人员充分发掘特征的性能或设计其他更有效的特征表示。为此，我们提出一种空间金字塔统计特征提取框架，在此框架基础上去实现多种常用统计特征的提取以及改进；同时，我们还利用基于多核学习的费舍尔判别分析方法实现 ReID 中的度量学习和多特征融合。实验结果证明，在 ReID 任务中，此框架下提取的改进局部统计特征性能要优于原始特征，并且结合多特征融合算法后可以进一步提升再识别的准确率。

从基于特征序列提取和序列匹配的端到端处理模型的角度，我们提出了一个上下文敏感的特征序列提取以及基于双重注意力机制序列匹配的深度 ReID 模型。传统的 ReID 算法在匹配行人图像或者行人跟踪序列之前，往往先将其表示成为一个单独的特征向量，然后在向量空间进行度量学习。然而在复杂的拍摄环境下，单一特征向量并不足以消除行人外观上的模糊性。为此我们提出，将每个行人表示成一系列包含细节信息的特征集合或者序列，并利用双重注意力机制进行序列匹配，从而实现高精度的行人再识别。模型中采用的双重注意力机制是整个算法的核心，其中序列内部的注意力机制用来进行特征序列的去噪精炼，而序列间的注意力机制用来实现序列对的语义对齐。借助这两种注意力机制，包含在特征序列对中的细节信息可以被自动地挖掘出来，并被合理地比较，从而得到恰当的行人距离或相似性度量。实验结果证明我们的模型在多个大规模数据集上都取得了同期最优的性能。

关键词：行人再识别　度量学习　特征表示　多核学习　深度神经网络　注意力机制　正则化约束

PERSON RE-IDENTIFICATION IN INTELLIGENT VISUAL SURVEILLANCE

ABSTRACT

Person Re-Identification (ReID) is one of the most vital tasks in an intelligent visual surveillance system, which aims at associating the same pedestrian across multiple camera views, and extending the visual surveillance from single camera to camera network. Based on the previous substantive research work, we propose to improve the ReID performance from three perspectives, including the generalization ability in metric learning on small dataset, the feature engineering and fusion model, and the feature sequence extraction and matching model.

From the perspective of the generalization ability in metric learning on small dataset, we propose to accommodate the regularization strategy to enhance the robustness of the methods. Metric learning plays a critical role in person re-identification problem. Unfortunately, due to the small size of training data, the metric learning used in this scenario suffers from over-fitting which leads to degenerated performance. In this paper, we investigate the effect of regularization in metric learning for person re-identification. Concretely we formulate the distance function from three perspectives and hence present four different regularized metric learning methods. Experiments on two popular benchmark data sets VIPeR and CUHK01 validate the effectiveness of our proposed regularization approaches.

From the perspective of feature engineering and fusion model, we propose a unified framework of statistical local feature extraction and combination for ReID. Re-identifying individual across non-overlapping camera views is one of challenging problems in surveillance video analysis. The difficulties mainly come from the large appearance variations caused by camera view angle, human pose, illumination, and occlusion. Recently, extensive efforts have been cast

into addressing this problem by developing invariant features or discriminative distance metrics. However, there is still a lack of systematic evaluations on the pipeline for feature extraction and combination. In this paper, we propose a spatial pyramid based statistical feature extraction framework as a unified pipeline of feature extraction and combination for person ReID, and systematically evaluate the configuration details in feature extraction and the strategies in feature combination. Extensive experiments on benchmark datasets demonstrate the critical components in feature extraction. Moreover, by combining multiple features, our proposed approach can yield state-of-the-art performance.

From the perspective of feature sequence extraction and matching model, we propose a novel end-to-end trainable framework for person ReID, which can jointly learn context-aware feature sequences and perform sequences comparison with dual attention mechanism. Typical ReID methods usually describe each pedestrian with a single feature vector and match them in a task-specific metric space. However, the methods based on a single feature vector are not sufficient enough to overcome visual ambiguity, which frequently occurs in real scenario. In this paper, we propose a novel end-to-end trainable framework, called Dual ATtention Matching network (DuATM), to learn context-aware feature sequences and perform attentive sequence comparison simultaneously. The core component of our DuATM framework is a dual attention mechanism, in which both intra-sequence and inter-sequence attention strategies are used for feature refinement and feature-pair alignment, respectively. Thus, detailed visual cues contained in the intermediate feature sequences can be automatically exploited and properly compared. We train the proposed DuATM network as a siamese network via a triplet loss assisted with a de-correlation loss and a cross entropy loss. We conduct extensive experiments on both image and video based ReID benchmark datasets. Experimental results demonstrate the significant advantages of our approach compared to the state-of-the-art methods.

KEY WORDS: Person Re-Identification Metric Learning Feature Rep-

resentation Multiple Kernel Learning Deep Neural Network Attention Mechanism Regularization

目 录

表格索引

插图索引

符号对照表

X	常数
x	标量
$\mathbf{x}$	向量
$\mathbf{X}$	矩阵
$\mathscr{X}$	张量
$(\cdot)^{\mathrm{T}}$	矩阵转置
$\{\cdots\}$	集合

第一章 绪论

本章首先介绍了智能视觉监控以及行人再识别技术的研究背景和意义（小节1.1）；接着，从行人再识别的研究进展、存在的挑战、以及算法性能评价指标和常用数据集等方面对课题的研究现状进行了概述（小节1.2）；然后，介绍了我们的研究内容、创新点、和主要研究成果（1.3）；最后，介绍了论文的组织结构（1.4）。

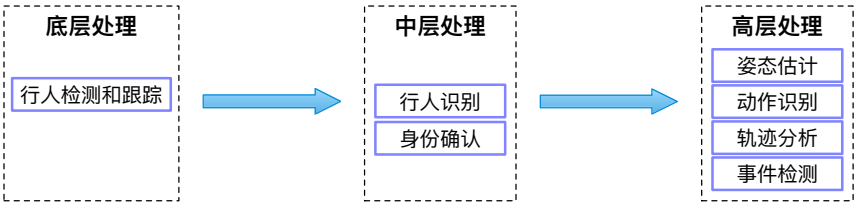
1.1 课题的研究背景及意义

随着计算机、通信等新技术的飞速发展，视频监控网络作为重要的基础公共设施已在全球范围内得到广泛应用，并且已经成为国内外各个公共管理部门进行社会安全管控及治理过程中不可或缺的重要力量。在国内，尤其是近几年，国家层面大力推进“平安城市”、“智慧城市”、“智能交通”等项目的建设，有力地促进了各级公安、政府及社会各界对视频监控产品的需求。随着前端监控摄像头安装数量的不断增多、视频监控网络的日趋完善，监控视频录像数据体量也开始呈爆炸式增长，致使监控系统的可用性及有效性受制于系统对海量图像视频数据的采集、分析及处理能力。目前主流的视频监控系统借助于成熟的硬件技术、传输技术、以及云存储技术，可以充分发挥异地监控、同步记录、延时存储等数据采集功能，而对于监控信息的筛选整理以及监控内容的分析理解往往还需要大量人工参与。同时，由于监控视频数据具有体量巨大、实时性要求高、价值密度低（例如，一小时的监控视频，有效数据可能仅仅只有一两秒）等特点，随着监控数据量的不断增长，单纯依靠人力进行数据的分析和处理变得代价高昂，甚至不切实际。如何借助先进的计算机技术、利用强大的机器算法，更高效地完成有效信息的挖掘是目前海量监控数据背景下亟待解决的难题。

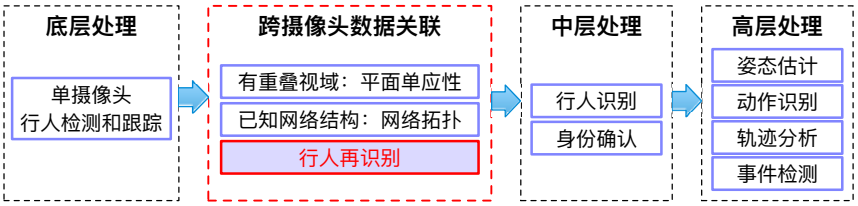
为解决以上问题，智能视觉监控（Intelligent Visual Surveillance，IVS）技术应运而生，并迅速发展成为一个工业界和学术界共同关注的研究热点。IVS 的核心是，利用计算机视觉、模式识别、以及机器学习等方法，实现自动地监控视频处理、内容分析和理解。图1-1展示了典型的IVS 系统所涉及的关键技术步骤。由于对于大部分监控应用来说，监控视野中的行人通常是其重点关注对象，因此IVS 系统往往需要具备以人为中心的视频理解的能力。具体来讲，一个典型的基于单摄像头的IVS 系统一般需要解决以下任务：

- 在底层处理上，实现对监控场景中行人目标的自动检测和跟踪；
- 在中层处理上，实现对感兴趣目标的身份识别和身份确认；
- 在高层处理上，实现对感兴趣目标的行为分析和理解。

不难发现，行人目标的自动检测和跟踪技术，是实现智能视觉监控的基础所在。



(a) 单摄像头智能视觉监控系统



(b) 多摄像头智能视觉监控网络

图 1-1　典型的智能视觉监控系统所涉及的关键技术步骤。

然而，在实际应用中，一个完整的监控网络通常由布置在不同地点的多个摄像头构成，受监控行人目标的活动往往会横跨多个摄像机视域，由此导致行人活动轨迹在时空上的不连续性，使传统的基于单一摄像头的目标检测和跟踪算法无法完成对感兴趣目标的持久跨区域追踪。因此，为了实现多摄像头协同监控，我们有必要探索高效的数据关联策略使不同摄像机下的行人检测和跟踪结果得以结合。受使用场景限制，一般有以下几种数据关联策略：

- 假如监控网络中两两摄像头之间存在视域重叠，可以利用重叠视域图像的平面单应性（Planar Homography）建立不同摄像头视点间的对应关系，从而实现跨摄像头数据关联；
- 假如已知监控网络中摄像头布置的空间拓扑关系（Topological Spatial Relationship），可以由此推断行人的运动路径，实现跨摄像头行人追踪；
- 直接利用行人外观特征（如衣着、外貌、体态等）进行基于视觉信息的相似性匹配，实现跨摄像头跨视域的行人图片或行人跟踪片段的关联，即行人再识别（Person Re-Identification）。

由于与其他跨摄像头数据关联策略相比，行人再识别技术无需对监控网络的布局结

构有特殊要求，也不依赖外界提供网络结构信息，因此在实际应用上更具可行性，同时也更具研究价值。

此外，行人再识别技术除了可以解决IVS 网络中的跨摄像头数据关联问题，还可作为一种通用的基于视觉的行人匹配算法，为多种计算机视觉任务提供解决方案。例如，在非受控的视频监控场景中，我们往往无法获取高清的人脸图像，因此无法依靠成熟的人脸识别（Face Recognition）技术进行有效的身份认证；而行人再识别所依赖的行人衣着、外貌等信息获取相对容易，以行人再识别技术作为补充，可以很大程度地提升复杂场景下的行人身份认证系统的稳定性和准确度；另外，行人再识别也提供了一种非侵犯式的、无需目标配合的，身份一致性认证替代解决方案。又如，在公安执法过程中，执法人员可以根据案发现场监控中捕获的嫌疑犯图像，利用行人再识别技术在公安系统的数据库或者全国各地的监控录像中进行嫌疑犯搜索，从而实现快速准确地对嫌疑人身份的识别、以及对嫌疑人所在位置的定位，帮助执法人员缉拿罪犯；即使在没有捕获嫌疑犯照片的情况下，执法人员也可以根据目击者的描述（如衣着、肤色、体态等），与行人再识别技术从数据库图像或监控网络中提取的行人外观特征描述进行比对，实现对嫌疑犯的快速排查和追踪。

除了广泛的应用价值，从学术研究上来看，行人再识别可以看做图像或者视频匹配技术在行人抓拍图像或者行人跟踪视频上的应用特例，因此高精度的行人再识别算法或者行之有效的算法设计思路可以很好地启发人脸匹配、衣物检索、视频检索等其他图像或者视频匹配热点技术的发展，从而推动计算机视觉技术的进步。正因为行人再识别在智能视觉监控、身份认证、行人检索、图像和视频匹配等方面具有重要的应用价值和研究价值，我们将本课题的主要研究内容定位于探索行人再识别中的关键技术点，并提出高精度的行人再识别算法。

1.2 课题的研究现状

1.2.1 研究进展

行人再识别 (Person Re-Identification，ReID)[6,7]，在国内也被称为行人重识别或者行人再辨识。该任务最初的目的是，基于行人的外观视觉特征（如衣着、肤色、体态等），将再次出现在不同时间不同地点的行人监控图像与已抓拍到的行人监控图像进行关联，以实现对行人的持久跨摄像头追踪。对于传统的ReID 任务，算法输入的原始数据往往是行人检测算法得到的行人边界框（Bounding Box）或者行人跟踪算法得到的行人轨迹片段（Tracklet），输出的是样本对的距离度量或者相似性度量结果。

因此，一个完整的ReID 算法通常由以下两个步骤构成，

- 特征表示：提取可靠的、鲁棒的、紧致的视觉特征，使其对于光照、视角、姿态、背景干扰、遮挡、以及图像质量等变化具有一定的不变性，同时对于不同身份的行人具有较高的可辨识度；
- 特征匹配：构造一种合适的距离度量或者相似性度量方式，使具有同一身份的行人特征距离近（相似度高），使具有不同身份的行人特征距离远（相似度低）。依据不同的特征匹配策略，有时需要利用数据驱动的训练过程来优化特征匹配模型中的参数，以获得特定任务专用的度量方式。

在实际应用中，一般给定某摄像头下的行人图像或者视频组成探测集（Probe Set，PSet），给定不同摄像头下的行人图像或者视频组成候选集（Gallery Set，GSet）；然后，提取PSet 和GSet 中样本元素的特征表示；最后，计算PSet 中的每一个测试样本（Probe）与GSet 中所有候选样本的特征距离，如果距离小于某一阈值则判定为匹配结果。图1-2对ReID 的基本流程进行了详细展示。

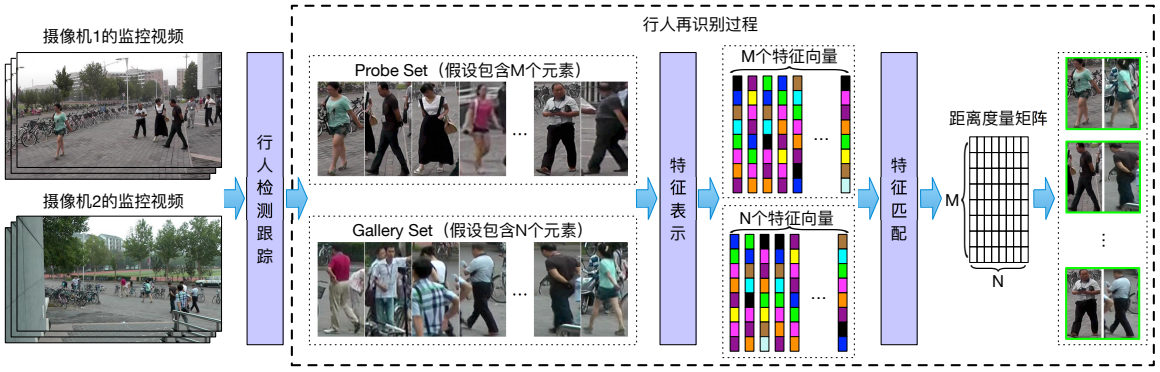


图 1-2　行人再识别基本框架。图中所用到的图像样本均来自于数据集 PRW[2]。

自 2006 年以术语 *Person Re-Identification* 在计算机视觉国际顶级会议 CVPR① 上被提出[8]，ReID 问题开始日益受到学术届和工业界的共同关注；尤其是 2007 年行人再识别数据集 VIPeR 的公开发布[4]，促使越来越多的研究学者将自己的研究成果发表在计算机视觉和机器学习的各大顶级国际会议（如 CVPR、ECCV②、ICCV③、

① CVPR: IEEE Conference on Computer Vision and Pattern Recognition
② ECCV: European Conference on Computer Vision
③ ICCV: IEEE International Conference on Computer Vision

BMVC④、ICIP⑤、AAAI⑥、IJCAI⑦、NIPS⑧等）以及各大顶级国际期刊（如 TPAMI⑨、TIP⑩、IJCV⑪等）上，大大地推动了ReID 技术的快速发展。在ReID 研究之初，很多学者将其构造成图像匹配问题，并提出了大量特征提取[9–22] 和特征匹配[23–53] 方法，带来了算法性能的逐步提升。同时，由于ReID 主要应用于视频监控场景，行人外观特征除了可以从静态抓拍图像中获取以外，还可以从大量的动态跟踪视频序列中获取，而且视频序列数据同时还包含了丰富的时空信息（如动作特征等）；因此，为了进一步提升算法准确度，许多学者也开始将ReID 构造成视频匹配问题进行处理[54–65]。近年来，随着深度学习（Deep Learning，DL）技术的发展，尤其是卷积神经网络（Convolutional Neural Network，CNN）在各种计算机视觉应用上的卓越表现（如图像分类、人脸识别等），一种基于深度神经网络的端到端优化的模型设计思路被引入到ReID 中[1,66–88]，使算法的性能开始逐渐能够满足实际应用的要求。

随着研究人员对ReID 的深入研究，在不断提升算法准确度的同时，也催生了大量与ReID 相关的新兴计算机视觉问题。比如，研究人员在[89–91] 中构造了一种行人搜寻（Person Search，PeSe）问题，将行人检测跟踪和行人再识别两个任务进行联合建模优化，实现给定待搜寻人物的查询图片，直接从全景监控图片或者视频中定位出目标人物的位置；研究人员在[92,93] 中提出了一种新的行人检索（Person Retrieval，PeRe）任务，实现给定查询人物的视觉属性特征文本描述，直接从图像或者视频库中检索出目标人物；研究人员在[94,95] 中提出一种跨模态行人检索（Cross-Modality Person Re-Identification，X-ReID）任务，用以解决白天拍摄到的 RGB 图像与夜间拍摄到的红外图像的相似性匹配问题。另外，由于目前大部分ReID 研究通常将其当做闭集（Closed-Set）上的匹配问题进行处理，即假设PSet 中所有的待查询行人都恰好被包含在GSet 中，而实际应用中该假设往往不能成立，因此研究人员在[96,97] 中定义了一种开集的行人再识别（Open-Set Person Re-Identification，OP-ReID）问题，将再识别过程分解为判断行人目标否存在以及检索存在的行人目标两个子任务。

总体而言，经过十几年的探索，ReID 技术无论在理论研究上，还是在应用场景

④ BMVC: British Machine Vision Conference
⑤ ICIP: IEEE International Conference on Image Processing
⑥ AAAI: AAAI Conference on Artificial Intelligence
⑦ IJCAI: International Joint Conference on Artificial Intelligence
⑧ NIPS: Annual Conference on Neural Information Processing Systems
⑨ TPAMI: IEEE Transactions on Pattern Analysis and Machine Intelligence
⑩ TIP: IEEE Transactions on Image Processing
⑪ IJCV: International Journal of Computer Vision

扩展上都已经取得了长足的进步。

1.2.2 存在的挑战

虽然部分ReID 算法的性能在很多公开数据集上已经取得了大幅度的提升，但是受复杂的监控环境的影响，如何在实际应用场景中实现稳定的、高精度的ReID 仍是一个充满挑战的任务。结合图1-2中所示的ReID 处理流程，我们主要从特征表示和特征匹配两个角度对ReID 在实际应用中所面临的主要挑战进行简要介绍。

1.2.2.1 特征表示上面临的挑战

（1）复杂的光照和成像条件

摄像机抓拍图像或者录制视频的过程是一个光电转化过程，拍摄时的光照条件、摄像机的内置感光元件、光电转换系统等的不同，都会引起成像结果的视觉差异。这些差异使得具有同一身份的行人，在不同时间、不同地点、不同摄像机下拍摄的外观特征，比如颜色信息等，产生不同程度的偏差，从而影响了ReID 系统的性能。图1-3(a) 为不同光照条件下拍摄的同一行人图像，颜色信息发生了明显失真；图1-3(b) 为不同设备拍摄的同一行人图像，清晰度不同导致外观不一致。

（2）局部的背景干扰和遮挡

在拥挤环境下或者低质量拍摄环境下，行人自动检测跟踪算法有时无法精确地定位行人在监控画面中的位置，致使得到的行人边界框可能包含部分背景画面或者感兴趣行人目标被局部遮挡，从而一定程度上造成行人外观特征的破坏和污染，进而降低了ReID 系统的准确度。图1-3(c) 展示了行人图像被局部遮挡或者存在背景干扰时所造成的外观特征的恶化。

（3）多变的拍摄角度和行人姿态

由于ReID 技术是一种非侵犯式的、无需用户配合的行人身份一致性匹配技术，行人目标可能从任何角度、以任何姿势进入摄像机视域；并且，随着目标的不断移动，目标与摄像机的相对空间位置也不断发生变化，使得摄像机拍摄行人的视角以及拍摄到的行人姿态都在不断变化；而且，在监控镜头下，行人除了进行类似刚体的空间移动外，还会进行有四肢、头部等参与的非刚体活动，使其外观特征发生剧烈地变化。图1-3(d) 和1-3(e) 分别展示了行人在不同拍摄角度、或者不同姿态下外观的显著差异。

（4）不固定的衣着外观

在ReID 中，行人的衣着、佩戴的饰物包裹等，很大程度上决定了行人的视觉外观特征，从而影响了系统的再识别准确度。然而，与传统的生物识别特征（如人脸、虹膜、指纹等）不同，由于衣着佩戴等视觉信息不具有独特性、稳定性、以及不可复制性等特点，因此致使ReID 系统的稳定性比较差。比如，现阶段的ReID 算法一般假设行人目标在监控网络中没有发生大幅度地衣着变化，而只有短时跨摄像头运动才能满足这项假设，大大限制了ReID 在长时行人跟踪问题上的应用；另外，当监控网络布置在学校、工厂、政府机关等场景时，由于被监控人员一般会穿戴统一制服，使具有不同身份的行人间的可辨识度降低，大大增加了ReID 的难度。图1-3(f) 展示了不同行人，当衣着极其相似时，所造成的视觉外观上的混淆。

鉴于ReID 技术在实际应用时所面临的诸多特征表示方面的挑战，启发我们在设计高精度算法时应考虑以下特征提取策略：采用光照归一化、颜色空间转换等图像预处理技术，以降低复杂的光照影响；学习跨摄像机的特征投影，探索跨摄像头共享的特征空间，以降低摄像机成像条件不同造成的视觉外观影响；进行局部特征筛选，以减少背景干扰或者局部遮挡的影响；提取跨视角、跨姿态不变的高层语义特征，以刻画行人内在的、稳定的、可辨识的外观特点；一方面通过融合多种互补的外观特征来弥补单纯依靠衣着特征的不足，一方面通过增强对细节特征的关注来提升相似衣着情况下的行人可辨识度；等等。

(a) 光照不同　(b) 清晰度不同　(c) 局部遮挡和背景干扰　(d) 拍摄角度不同

(e) 行人姿态不同　(f) 衣着外观相似　(g) 空间位置不对齐问题

图 1-3　行人再识别中的主要问题和挑战。绿色矩形框代表样本对身份一致，红色矩形框代表样本对身份不一致，红色叉号“×”代表空间位置不对齐现象，所用图像均来自数据集 Market1501[3]。

1.2.2.2 特征匹配上面临的挑战

（1）类间混叠和类内偏移

如前面介绍的，受复杂的监控场景的影响，ReID 系统所提取的基于外观信息的特征表示在鲁棒性和可辨识度上往往不够理想，因而容易导致类间混叠和类内偏移问题。其中，类间混叠的产生通常是由于具有不同身份的行人却拥有相似的外观，比如身着类似的衣服等等；而类内偏移的产生通常是由于具有同一身份的行人以不同角度、不同姿态、或者在不同环境下被不同摄像机拍摄所致，而且这种跨视域的外观特征的改变往往是极其复杂且多模态的，因此也是很难被单一模型所学习刻画的。

（2）局部不对齐

与传统的生物特征匹配过程不同，如人脸匹配在进行相似度计算之前会进行基于标定点的人脸图像的对齐（保证嘴和嘴、鼻子和鼻子等空间位置的相对一致），由于行人姿态变化更为复杂且不易获取稳定的标定点，ReID 中处理的往往是空间位置不对齐的行人图片、或时间戳不对齐的行人跟踪序列，增加了计算可靠距离或者相似度的难度。图1-3(g) 展示了由于行人定位框不精准造成的待匹配图像对间的局部空间位置不对齐现象。

（3）训练样本缺乏

首先，由于ReID 要解决的是行人匹配或者身份确认的问题，在实际应用中算法所处理的行人目标通常不会被包含在训练样本之中，因此一般没有足够的样本将ReID 构造成传统的多分类模型（即“判断是谁”）进行学习，而只能当做二分类问题（即“区分一对样本是属于同类，还是属于不同类”）进行处理。其次，在监督学习的策略下，如果要使特征匹配模型有能力刻画行人外观在各摄像机对间的相对变化，那么必须保证目标行人在所有摄像机视域内出现并被标记，显然这种样本标记策略对于大型的监控网络来说是不切实际的。因此，ReID 任务通常要面对的是，不充足数据集上的模型训练和优化问题。

（4）模型泛化能力

大部分ReID 算法通常可以在给定训练数据的摄像机对之间学习到性能良好的跨视域匹配模型，然而却无法很好地将模型推广到监控网络中的其他摄像机对之间，即模型泛化能力差。然而，在实际应用，受训练样本匮乏的限制，往往更需要泛化能力强的算法和模型。

鉴于ReID 技术在实际应用时所面临的诸多特征匹配方面的挑战，启发我们在设计高精度算法时应考虑以下模型构造及训练策略：模型训练过程中，在克服类间混

叠的同时，还要减小类内偏移；在距离度量或者相似度计算过程中，要引入局部对应性约束，以保证特征匹配机制的合理性；由于训练样本匮乏，在模型训练过程中，要对有限训练样本进行充分利用，比如联合利用样本对的身份确认标签（Verification Label）与样本个体的身份标签（Identity Label）；对于小规模数据集，通过降低模型复杂度或者借助无标签数据的方式，提升模型的泛化能力；等等。

1.2.3 性能评估及常用数据集

1.2.3.1 性能评估指标

ReID 通常可以被当做匹配检索或者排序任务，其基本目标是在给定待查询目标的基础上，利用算法计算待查询目标与被查询集合中所有元素的距离，并按照距离从近到远对被查询集合进行排序，从而使与待查询目标身份一致的元素排在第一位。目前有多种的性能评估指标来量化不同ReID 算法的有效性，其中首位准确度（Rank-1 Accuracy，Rank-1）和累计匹配性能（Cumulative Matching Characteristic，CMC）曲线是两种最常用的评估指标。

Rank-1 可以被看做是传统概念上的分类准确度，即ReID 算法返回的排序列表首位元素与待查询目标身份一致的概率。然而，在实际应用场景中，单纯依靠ReID 算法往往无法取得很高的Rank-1 值，也就没办法真正地独立完成跨摄像机的行人身份一致性匹配任务。因此，现阶段更实际的ReID 应用场景是，算法返回一个排序列表，用户从列表前 N 个对象中选取真正的匹配结果。假如 N 远远小于被查询集合的大小，ReID 算法就可以大大地降低人力成本，提升匹配任务的效率。而CMC 曲线就可以刻画排序列表前 N 个对象包含与待查询目标身份一致的元素的准确度，其定义如下：

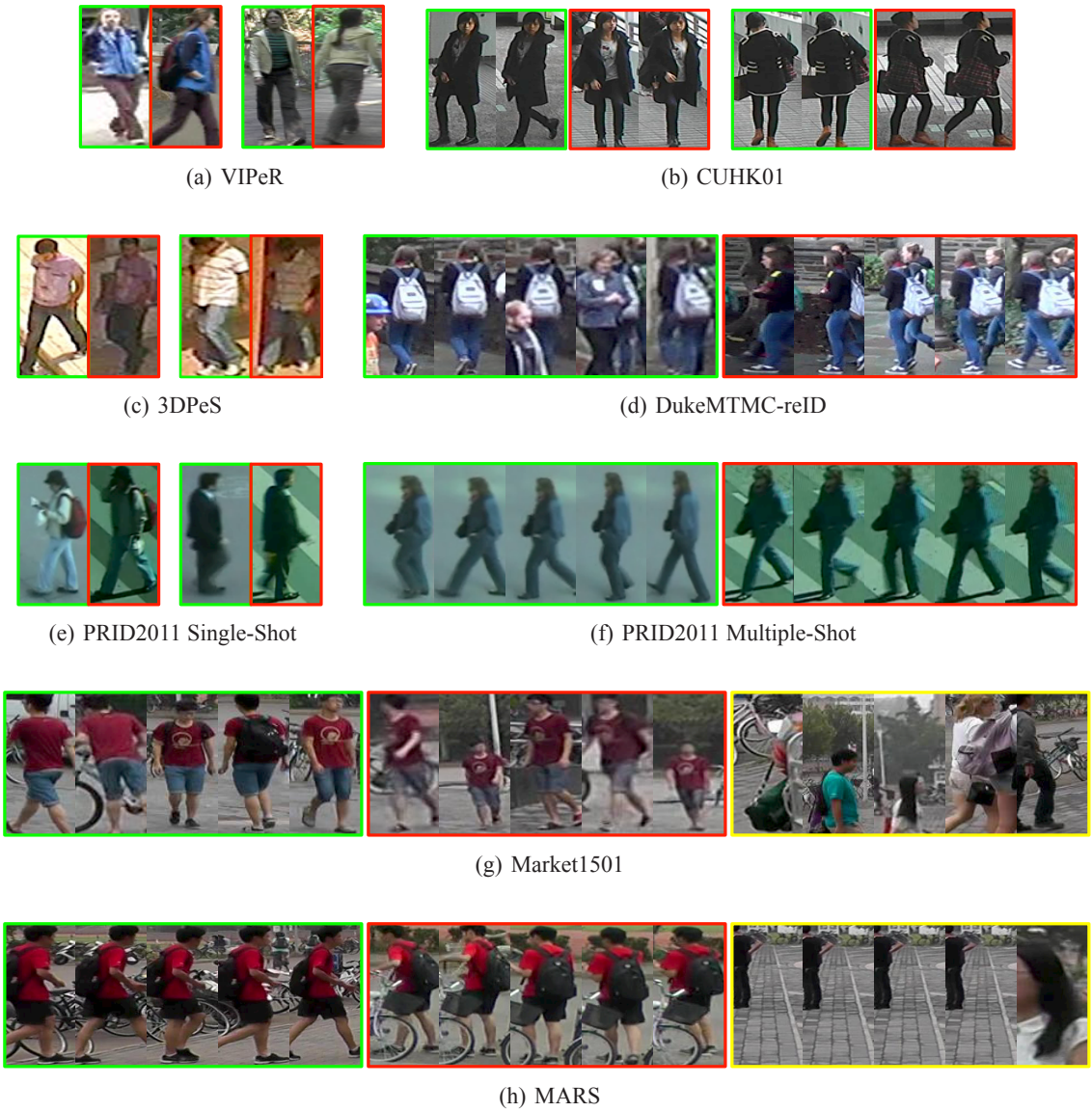
$$cmc(N) = \sum_{n=1}^{N} r(n), \tag{1-1}$$

其中 $r(n)$ 表示排序列表第 n 个元素与待查询目标身份一致的概率。以 N 为横坐标，以 $cmc(N)$ 为纵坐标，即可绘制出CMC 曲线；而且容易发现，$cmc(1)$ 与Rank-1 相同。

当被查询集合中存在且仅存在一个与待查询目标身份一致的元素时，CMC 曲线可以很好地刻画ReID 算法的性能，因为此时排序结果的准确度和召回率是一致的。然而，当被查询集合中存在多个与待查询目标身份一致的元素时，CMC 曲线因为没有考虑到召回率指标而会有所偏颇。这种情况下，往往会引入均值平均精度（Mean Average Precision，mAP）作为ReID 算法性能评估指标的补充。mAP 的计算过程如下：

- 对于每一个待查询目标，利用ReID 算法得到排序结果，计算其对应的平均精度 (Average Precision，AP），即精确率-召回率曲线 (Precision-Recall Curve，PRC) 底部的面积；
- 求所有待查询目标对应AP 的均值，即为mAP。

1.2.3.2 常用数据集



(a) VIPeR (b) CUHK01

(c) 3DPeS (d) DukeMTMC-reID

(e) PRID2011 Single-Shot (f) PRID2011 Multiple-Shot

(g) Market1501

(h) MARS

图 1-4 行人再识别常用数据集样本举例。绿色和红色矩形框代表图片来自不同摄像头，黄色矩形框代表干扰项。

在计算机视觉或者机器学习领域，如果说算法是引擎的话，数据就是驱动引擎

的燃料。优质的公开数据集既可以帮助研究人员快速构建并训练先进的机器学习模型学习，也可以促使研发人员对各自的算法进行公平、公开地比较，从而推动算法和技术的进步。随着近几年ReID 研究领域的蓬勃发展，已经有多个数据集被公开发表出来，这些数据集一般由拍摄于不同摄像头下的行人图片或视频序列构成。为了消除行人检测或行人跟踪算法对再识别准确度的影响，行人图像或者视频序列通常由相同的自动检测跟踪算法或者直接由人工进行边界框标定。我们在表1-1中对几种常用的ReID 数据集做了对比说明，并将在下文对每个数据集的特点一一介绍，图1-4也展示了每个数据集的若干样例。

表 1-1 常用行人再识别数据集。

数据集	# 摄像机	# 行人	# 图片	图片尺寸	单图	多图	序列	标记方式
VIPeR	2	632	1,264	128×48	√	×	×	手工
CUHK01	2	971	3,884	160×60	×	√	×	手工
3DPeS	8	192	1,011	不固定	×	√	×	手工
DukeMTMC-reID	8	1,812	36,441	不固定	×	√	×	手工
PRID2011	2	934	24,541	128×64	√	√	√	手工
Market1501	6	1,501	32,217	128×64	×	√	×	自动
MARS	6	1,261	1,191,003	256×128	×	√	√	自动

（1）VIPeR[4]

VIPeR 是ReID 领域早期研究中应用最广的数据集，由 Gray 等人收集并于 2007 公开发布。数据集中的行人图片采集于校园场景，跨摄像头的行人图片对之间通常具有比较大的拍摄视角和光照变化。

（2）CUHK01[5]

CUHK01 由 Li 等人于 2012 年发布，行人图像采集于校园场景，其特点是同一身份行人在每个摄像头下都拥有 2 张外观明显差异的图像，增大了ReID 中的类内偏移问题。

（3）3DPeS[98]

3DPeS 由 Davide 等人于 2011 年公布，行人图像采集于 8 个视域不重叠的室外场景。其特点是强烈的光照明暗对比以及不固定的图像尺寸，同时数据集还提供了摄像头的标定信息。

（4）DukeMTMC-reID[99]

DukeMTMC-reID 是由大规模多目标多摄像头跟踪 (Multi-Target Multi-Camera

Tracking，MTMCT）数据集的子集组成的集合，由 Zheng 等人于 2017 年发布。数据集提供了由手工裁剪得到行人图片，而且其中的 408 个身份的行人只出现在一个摄像头下。

（5）PRID2011[100]

PRID2011 由 Hirzer 等人于 2011 年发布，数据集包含了采集于摄像机 A 的 385 个行人和采集于摄像机 B 的 749 个行人的监控图片，其中只有 200 个人同时出现在了两个摄像头。此外，作者分别提供了基于单幅图与基于跟踪视频序列的两个ReID数据集版本。

（6）Market1501[3]

Market1501 是目前应用最为广泛的一个大规模数据集，由 Zheng 等人采集于校园场景并于 2015 年发布。数据集提供了由可变形组件模型（Deformable Parts Model，DPM）行人检测算法得到行人图片，而且还包含了 2793 张虚警照片作为干扰项。

（7）MARS[101]

MARS 是由 Zheng 等人于 2016 年发布的最大规模的基于视频的行人再识别数据集，行人跟踪序列由行人检测算法DPM 和行人跟踪算法广义最大多团问题跟踪器（Generalized Maximum Multi Clique Problem，GMMCP Tracker）得到，而且数据集中同一身份的行人在每个摄像头下通常拥有多个跟踪序列。

1.3 论文的主要工作和研究成果

本文将在国内外大量研究成果的基础上，结合ReID 所面临的主要挑战以及我们对ReID 问题的认识，分别从小数据集上的度量学习模型泛化能力、基于手工特征设计和多特征融合的分步处理模型、基于特征序列提取和序列匹配的端到端处理模型这三个角度提出不同的优化的ReID 方案。本文的主要研究内容及成果总结如图1-5。

（1）提出利用正则化的度量学习方法来增强小数据集上ReID 模型的泛化能力

众所周知，度量学习的研究在 ReID 技术发展过程中扮演着重要的角色。然而，受某些应用场景的限制，研究人员往往无法获取充足的标记样本进行模型训练和学习，从而导致 ReID 算法的泛化能力较弱。为此，我们从限制模型复杂度的角度，提出利用正则化的度量学习算法实现 ReID 中的特征距离度量，从而提升小数据集上模型的泛化能力。具体来讲，我们分别从马氏距离学习、对称投影学习、以及非对称投影学习三个不同的角度理解度量函数，并构造了四种不同的正则化度量学习模型来实现 ReID。在数据集 VIPeR 和 CUHK01 上的实验验证了，正则化的模型约束，往

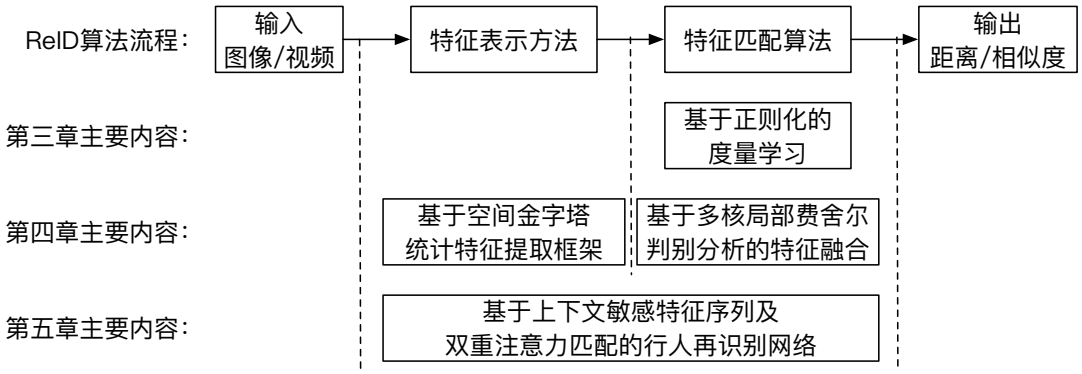


图 1-5 论文的主要研究内容及成果概要。

往可以带来整体算法性能的提升。

（2）提出了一种统一的局部统计特征提取框架，并结合多核学习实现ReID 中的多特征融合

在监控场景中实现跨摄像头视域的行人匹配是一项极富挑战的任务，因为不同拍摄角度、不同行人姿态、不同光照条件、以及局部遮挡等都会引起行人外观的剧烈变化，从而增大匹配难度。目前，大量的研究工作主要集中在，构造优秀的特征表示或者学习合理的特征匹配模型这两个方面。然而，由于影响行人外观的因素众多，很难构造单一特征来全面地刻画行人外观；而且，不同特征的提取过程相对独立、缺乏系统而详细地评估分析，很难启发研究人员充分发掘特征的性能或设计其他更有效的特征表示。为此，我们提出一种空间金字塔统计特征提取框架，在此框架基础上去实现多种常用统计特征的提取以及改进；同时，我们还利用基于多核学习的费舍尔判别分析方法实现 ReID 中的度量学习和多特征融合。实验结果证明，在 ReID 任务中，此框架下提取的改进局部统计特征性能要优于原始特征，并且结合多特征融合算法后可以进一步提升再识别的准确率。

（3）提出了一个端到端的上下文信息敏感的特征序列提取以及基于双重注意力机制的序列匹配ReID 模型

传统的 ReID 算法在匹配行人图像或者行人跟踪序列之前，往往先将其表示成为一个单独的特征向量，然后在向量空间进行度量学习。然而在复杂的拍摄环境下，单一特征向量并不足以消除行人外观上的模糊性。为此我们提出，将每个行人表示成一系列包含细节信息的特征集合或者序列，并利用双重注意力机制进行序列匹配，从而实现高精度的行人再识别。模型中采用的双重注意力机制是整个算法的核心，其中序列内部的注意力机制用来进行特征序列的去噪精炼，而序列间的注意力机制

用来实现序列对的语义对齐。借助这两种注意力机制，包含在特征序列对中的细节信息可以被自动地挖掘出来，并被合理地比较，从而得到恰当的行人距离或相似性度量。实验结果证明我们的模型在多个大规模数据集上都取得了同期最优的性能。

1.4 论文的结构安排

本论文的主要内容结构如下:

第一章，绪论。主要介绍了课题的研究背景及意义、研究现状、面临的主要挑战、以及算法的评价指标和常用数据集，并列出了本文的主要研究内容、创新点、以及主要研究成果。

第二章，行人再识别的相关算法。主要对目前行人再识别研究领域所涉及的重要算法进行概要介绍，以加深对行人再识别问题的全面理解。

第三章，基于正则化度量学习的行人再识别算法。从研究背景和意义、相关工作、算法细节等方面对我们所提出的基于正则化度量学习的行人再识别算法进行介绍，并通过大量实验验证算法的有效性。

第四章，基于空间金字塔统计特征及多核学习的行人再识别算法。从研究背景和意义、相关工作、算法细节等方面对我们所提出的基于空间金字塔统计特征及多核学习的行人再识别算法进行介绍，并通过大量实验验证算法的有效性。

第五章，基于上下文敏感特征序列及双重注意力匹配的行人再识别网络。从研究背景和意义、相关工作、算法细节等方面对我们所提出的基于上下文敏感特征序列及双重注意力匹配的行人再识别网络进行介绍，并通过大量实验验证算法的有效性。

第六章，总结与展望。总结了全文的研究工作，对行人再识别问题将来的工作和研究内容进行了展望。

第二章 行人再识别的相关算法

为了更清楚、更全面地理解行人再识别问题以及本课题的研究动机，本章分别从特征表示方法（小节2.1）、特征匹配算法（小节2.2）、以及深度神经网络算法（小节2.3）这三个方面对目前常用的ReID 算法进行简要概述和分析。

2.1 特征表示方法

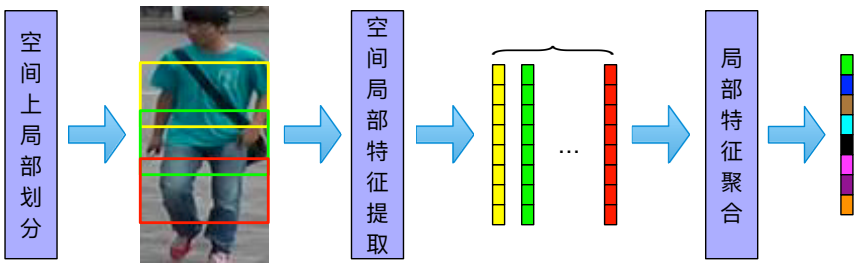
ReID 中常用的特征表示方法大致可以归纳为基于底层视觉特征提取的方法和基于高层语义特征学习的方法两种，前者主要涵盖了图像颜色直方图、梯度方向直方图、局部二值模式、纹理滤波器（如 Gabor、Schmid 滤波器等）、及其变种等特征表示，后者主要包括属性特征和深度特征等。下面我们将分别对这两种类型的特征表示方法进行逐一介绍。

2.1.1 底层视觉特征提取

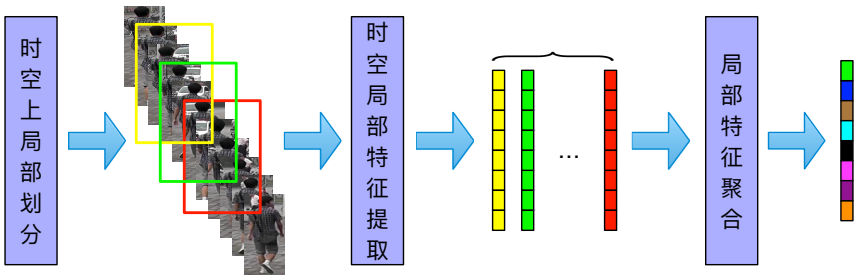
与高层语义特征相比，底层视觉特征往往可以更细致地刻画图像或者视频的细节信息。在特征提取过程中，对于行人图像，通常需要先将图像在空间上划分出若干个感兴趣区域，然后设计局部区域的手工特征表示，最后再根据每个区域所包含的有效信息量进行关键特征的选择或者局部特征的聚合；而对于行人跟踪序列，往往需要将其在时域上将划分出若干最小完整视频片段，然后再基于视频片段进行局部时空特征的提取和聚合。图2-1分别展示了，输入为行人图像或者行人跟踪序列时底层视觉特征的提取过程。

2.1.1.1 局部划分

由于在监控场景下的行人再识别任务中，行人图像往往都是以相对水平的拍摄角度被抓拍得到的，因此常假设图像中处在同一水平位置区域中的像素集合刻画了行人某一特定的身体部位，因而图像可以在垂直方向划分为若干个包含了不同身体部位的局部区域（比如从上到下可能依次包含了，头、肩膀、胸部、腹部、臀部、大腿、小腿、脚等信息）。按照这一经验知识，论文[9,16,18,21–23,25,37,38,42,44,49,102] 将行人图片在垂直方向划分成了若干图像条（Image Stripe），并分别提取基于图像条的局



(a) 行人图像的底层视觉特征提取流程



(b) 行人跟踪序列的底层视觉特征提取流程

图 2-1　底层视觉特征提取基本流程。

部特征。然而，实际上由于行人的某些身体部分无法简单地在垂直方向上进行分割，如双臂与躯干一般在水平位置上存在重叠，因此论文[3,13,14,17,24,26–30,32,33,39,40,103,104]中采用一种颗粒度更精细的图像分割方法将行人图像划分成稠密的图像块（Image Patch），分别提取基于图像块的局部特征。为了消除在局部特征提取过程中，局部区域分割线附近像素引起的信息模糊，在图像条或者图像块划分过程中，往往采取有部分重叠的分割策略。除此之外，为了更好地显式地利用人体的结构信息，研究人员也尝试按照人体的结构将图像划分成若干人体组件模块，并分别提取基于人体组件的局部特征。例如，Farenzena 等人[10,12] 先将行人从背景图像中分割出来，然后基于行人外观的对称反对称性及人体结构的比例知识，将前景图片划分成头部、躯干、和腿部三部分；Gheissari 等人[8] 基于可分解三角图，先将人体划分为若干三角形区域，然后利用颜色一致性对三角区域进行融合，最后得到六个身体子区域。

对于行人跟踪序列，由于序列长度不一，很难直接提取全局稳定的、有效的视频特征，因此也需要先将其分割成若干稳定视频段，再提取基于视频段的局部特征。论文[38,57,105] 中采取了最简单的视频段分割方法，即将每一视频帧当做最小视频段单元，直接提取视频帧的图像特征。这种方法的优点是直观且易实现，可以直接借鉴行人图像特征提取上的大量研究成果，而缺点是会造成视频中全部时域信息的丢失。因此，更合理的方式是划分出包含完整时域信息的小视频段单元。例如，Wang

等人[54,60]基于视频序列的光流能量分布图（Flow Energy Profile，FEP）将每个序列划分为若干个完整步行运动周期视频段，从而保证每个小视频段单元都包含了完整步态信息；Liu 等人[55,59]除了利用FEP 划分视频段以外，还在空间上将每个视频段分割成六个不同种身体组件（即头、躯干、左右手、和左右腿），从而在时空上将原始视频划分成若干最小动作视频单元。

2.1.1.2 局部特征提取

下文中，我们将分别从图像中的颜色特征、纹理特征、组合特征、以及视频中的时空特征这四个类型对行人再识别中常用的局部特征进行简要介绍。

（1）颜色特征

由于监控场景下行人的外观受行人衣着影响较大，而颜色往往又是衣着服饰的显著信息，因此颜色特征是行人外观描述的重要组成。颜色特征通常对于跨视域、跨姿态等变化不敏感，因此也比较适合行人再识别场景。最常用的颜色特征是颜色直方图（Color Histogram，Hist），如论文[13,14,17,21,23–26,28,30–33,38–40,42,44,46,103,104,106]中采用Hist 描述行人外观。它的基本原理是将颜色空间（如 HSV、RGB、LAB 等）等分成若干个取值区间，并在感兴趣区域内统计落在每个特征区间内的像素个数。在实际应用中，为了消除统计区域大小对统计结果的影响，往往需要将结果进行归一化，使其转化成为概率分布；而对于三通道的颜色空间，为了降低量化后统计特征的维数，往往是先单独在每一个通道中计算Hist 特征，然后再将其串联成三通道的Hist 特征。另外，当图像含有复杂背景干扰时，背景处的像素会污染直方图统计结果，因此 Farenzena 等人[10,12]提出利用像素到人体对称轴的距离为每个像素赋予不同权重，从而降低了远离人体的像素对最终统计结果的影响。由于Hist 属于一种均匀量化的颜色特征，当图像的真实颜色信息分布不均匀时，会导致位于某些颜色区间内的像素数目稀少，因而使直方图统计特征的某些维度区分能力变差，为此 Yang 等人[16]提出利用 16 种显著的常用颜色值作为划分区间中心，代替原始的均匀划分区间对颜色信息的进行量化统计。

（2）纹理特征

尽管颜色特征因良好的角度不变性以及丰富的衣着外观表达能力被广泛使用在ReID 应用中，但是由于其容易受光照变化影响的特点，限制了ReID 在自然场景下的系统稳定性。而物体的纹理特征因其具有光照不变性的特点，可以很好的补充颜色特征的不足，同时纹理特征也是衣着服饰外观描述的重要组成。论文[9,21,23,29,38]中

将不同参数配置的 Gabor、Schmid 等纹理滤波器作用在行人图像上，以生成对应的纹理信息通道，并计算相应的纹理信息统计直方图特征。局部二值模式 (Local Binary Pattern，LBP)[107] 特征是一种具有旋转和灰度不变性的的特征描述子，通过比较中心像素与周围像素的灰度值大小来刻画图像的纹理信息，并通过一定的规则对比较结果进行直方图统计，论文[17,24–26,32,42,46,103,104] 中将其应用到ReID 任务中。梯度方向直方图 (Histogram of Oriented Gradient，HOG)[108] 特征是一种广泛应用在行人检测中的特征描述子，通过计算和统计图像局部区域的梯度或边缘的方向密度分布来刻画局部目标的外观和形状，论文[17,29,32,46,104] 中将其引入到ReID 中以刻画行人的形状外观。除了以上基于局部区域的纹理特征以外，基于关键点的特征表示也常被应用到各种计算机视觉任务当中，如论文[13,14,30,39,40] 将密集采点的 SIFT 特征，即稠密的尺度不变特征变换 (Dense Scale-Invariant Feature Transform，DSIFT) 特征，应用到ReID 之中。

（3）组合特征

由于颜色和纹理信息分别从不同角度刻画了行人的外观，而且颜色和纹理特征往往具有各自不同的变换不变性，因此能够合理组合两种信息的特征表示通常可以进一步提升ReID 的准确度。

Liao 等人在[18] 中提出局部最大共生表示 (Local Maximal Occurrence Representation，LOMO)，该特征结合颜色和纹理两种信息，并且具有光照、角度、尺度等多种不变性。具体来说，为了使其具有光照不变性，作者采用 Retinex 算法[109] 进行图像预处理，使行人图像的颜色信息与人的感知相一致而不受外界环境干扰，并在此基础上提取图像局部区域的颜色直方图以及尺度不变的局部三元模式 (Scale Invariant Local Ternary Pattern，SILTP)[110] 特征；为了使其具有角度不变性，作者将基于图像块的局部特征在水平方向进行最大汇总操作，得到基于图像条的局部特征；同时为了使其具有尺度不变性，作者在不同尺度上进行了特征提取操作，并加以融合。目前，LOMO 特征已经被广泛应用到了ReID 研究之中，如[2,37,46,47,49–51,51,52,101,105,111–116]。

Matsukawa 等人在[22] 中提出的针对行人再识别的层次化的高斯描述子 (Gaussian Of Gaussian，GOG)，同样融合了颜色和纹理特征。GOG 的提取过程可以归纳为三个步骤：第一，将图像中的每个像素表示成包含了垂直坐标、方向梯度、以及颜色值等信息的多维向量；第二，在图像块区域内基于像素的向量表示进行高斯建模，得到基于图像块的高斯模型，并将高斯模型的均值和方差作为图像块的特征表示；第三，在图像条范围内基于图像块的特征表示进行高斯建模，得到基于图像条的高斯

模型，并将高斯模型的均值和方差作为图像条的特征表示。目前，GOG 也已经越来越多的受到研究人员的关注，如[105,114]。

（4）时空特征

上文介绍的颜色和纹理特征，往往都只是关注于图像中行人外观的描述，而对于视频序列来说，丰富的光流、运动等时空特征也有助于行人的身份一致性匹配。基于三维梯度的时空特征描述子（Spatio-Temporal Descriptor based on 3D Gradients，HOG3D）[117] 是动作识别领域广泛采用的一种时空特征，它是HOG特征在视频特征提取中的扩展。HOG3D 同时包含了视频中像素点在空间上和时间上的梯度变化信息，因此可以有效的捕获目标物体的光流信息或者运动信息等时空特征。论文[54,57]将HOG3D 引入到基于视频的行人再识别任务当中，用以弥补单纯依靠行人静态外观特征的不足。此外，研究人员 Liu 等人[55,59]，针对基于视频的ReID 任务，提出了一种联合纹理、颜色、运动信息的时空特征，在特征提取过程中先将每个像素点表示为包含了空间坐标、时间坐标、像素值、空间梯度、时间梯度等信息的多维向量，然后计算局部区域的费舍尔向量（Fisher Vector，FV）作为局部时空特征表示。

2.1.1.3 局部特征聚合

这里的特征聚合包含两个方面的含义，其一，是指在局部区域内或者局部时域内的不同类型的特征的融合，如颜色和纹理特征、或者静态空间特征和时空特征等；其二，是指将不同区域或者时域内的局部特征聚合为一个全局特征，从而将行人图像或者行人跟踪视频映射到特征向量空间，以便在向量空间内计算行人的特征距离。

对于多种不同类型特征的融合来说，目前ReID 中使用最多的方法是直接将不同类型的特征归一化后再串联起来，如[21,23,25,26,32,33,39]。然而，不同类型的特征往往对于ReID 任务有不同程度的作用，比如当行人衣着颜色相似时，纹理特征就具有更高的辨识度，更有助于进行行人身份一致性匹配；即使同一类型特征，在不同场景一下对ReID 任务的重要性也不同，比如光照条件变化剧烈时颜色信息的可信度会受到一定程度的影响。因此直接将不同特征串联往往并不能得到性能最优的特征组合。为此，Gray 等人[9] 提出局部特征集成（Ensemble of Localized Features，ELF）多特征融合策略，利用自适应增强（Adaptive Boosting，AdaBoost）算法以数据驱动的方式自适应地对局部特征进行筛选；而 Paisitkriangkrai 等人[35] 提出基于距离度量融合的多特征融合策略，即先以每一种单独特征为行人外观描述计算行人间的距离度量，然后学习最优的距离度量组合方式，使最终匹配结果与训练数据尽量一致。

对于将局部特征聚合为全局特征来说，目前ReID 中使用最多的方法也是直接将局部特征归一化后再串联起来，如[21,23,25,26,54]。同时，为了降低全局特征的维数，得到紧致的特征向量，有时也会引入投影降维的后处理操作，如论文[24,33,44,106] 利用主成分分析（Principal Component Analysis，PCA）算法降维、论文[29] 利用典型关联分析（Canonical Correlation Analysis，CCA）算法降维、而论文[38] 利用局部费舍尔判别分析（Local Fisher Discriminant Analysis，LFDA）算法实现降维。除此之外，研究人员Zheng 等人[3] 提出利用词袋模型（Bag-of-Words，BoW）算法实现局部特征到全局特征的融合，具体来讲，先提取行人图片局部区域的局部特征，接着利用 k-means 算法在训练数据上学习码本词典，然后利用码本词典对局部特征进行量化编码，最后对编码特征进行汇总得到图像全局特征。另外，由于在实际应用中计算机最终是依据样本对的距离度量来进行再识别判断，因此也有研究人员将局部特征的融合转化为对局部特征对距离度量结果的融合来进行处理。这类方法的关键点是，找到合理的局部特征对之间的对应关系，我们将在下文小节2.2.3对相关算法进行详细概述。

2.1.2 高层语义特征学习

由于底层特征往往刻画的是目标物体的视觉信息，因此更容易受到外界环境，如光照、拍摄视角等影响；而高层语义特征天然地对环境干扰不敏感。另外，类似于人类的感知系统，大脑通过对初级的视觉信号进一步加工，产生感知信号，从而产生行为决策，如识别种类、判断相似度等；图像的高层特征往往也是对底层特征的进一步抽象和提炼，产生语义相关的信息，从而更有助于计算机自动完成决策任务，如图像识别、图像匹配等。目前，在ReID 常用算法中主要涉及属性特征和深度特征这两种高层特征表示。

图像的属性特征一般是对图像中物体的一组语义描述，比如对于某张行人图片来说可能的属性描述为“男性、白人、短发、贝雷帽、红色外套、灰色裤子、等等”。属性特征一般无法直接通过设计特征提取器从图像中直接提取，而是基于图像的底层视觉特征，通过学习一个属性分类器来获取。在论文[19] 中，由于缺乏大规模的监控场景下行人属性标注数据集，Shi 等人利用弱监督的学习策略，先在大型的时装摄影数据集上学习属性分类器，再利用转移学习的方法将分类器转移到ReID 数据上，从而获取行人的属性特征；同样基于转移学习的策略，Su 等人[71] 将在辅助属性标签数据集上训练好的深度神经网络模型，微调到ReID 数据集上，以提取行人图像属性特征；而 Li 等人为了提升监控视频行人属性预测的结果，在[118] 中，收集并发布

了大型监控行人属性数据集 RAP。另外，受训练数据、分类器模型、学习算法等的限制，利用属性分类器提取的图像属性特征可能包含了错误的属性标签或相关性太高的属性对，从而限制了基于属性特征的ReID算法的有效性。为此，Su 等人[20] 提出通过学习一个低秩的投影矩阵，将原始属性特征映射到一个更加紧致、更加完整、连续的特征空间，实现对属性标签的去相关和去噪。除了直接利用图像的属性特征，Lin 等[119] 人利用多任务学习的策略，将属性预测分类器与行人再识别模型联合起来进行优化，从而进一步提升了ReID 的性能。

图像的深度特征一般是利用一个层次化的CNN，通过对原始输入图像进行逐层抽象而得到；而且，由于在训练特征提取CNN 模块时，经常将其当做识别或者分类问题对待，因此CNN 输出的深度特征一般也具有一定的语义相关性。由于早期ReID 研究中缺乏大规模的标记数据集，因此许多研究人员直接将在 ImageNet 分类数据集上训练好的CNN 模型拿来当做深度特征提取器。如 Paisitkriangkrai 等人[35] 将预训练的 AlexNet[120] 的最后全连接层输出作为深度特征；Chen 等人[121] 将预训练的 AlexNet 的前两个卷积层输出作为特征通道，提取每个通道的统计直方图作为深度特征；Cheng 等人[122] 将预训练的 ResNet152[123] 的全连接层的 2048 维输出作为深度特征；等等。此外，随着多个大规模ReID 数据集的发布，研究人员也开始针对再识别任务训练端到端的深度特征提取网络，如 Zheng 等人[101] 利用在 MARS 数据集上训练的 CaffeNet[120] 提取行人图片的具有身份区分度的嵌入特征 (ID-discriminative Embedding，IDE）。更多的关于端到端深度神经网络的ReID 算法模型将在小节2.3中被详细介绍。

2.1.3 特征表示方法总结

行人再识别中的特征表示方法，分别已经在底层视觉特征设计以及高层语义特征学习等方面都取得了显著的研究成果，为设计高精度的ReID 算法提供了坚实的基础。然而，部分特征表示方法中仍存在一些不足，值得我们进一步深入研究：

- 其一，在底层视觉特征设计上，虽然多种不同的特征表示方法已经被提出，而且也都带来了ReID 性能的提升，但是目前仍然缺少对底层视觉特征设计和提取过程的全面而详细地剖析和评估。我们发现，ReID 中多数常用底层视觉特征具有类似的性质或者提取过程（大部分属于统计特征，如颜色直方图、HOG、LBP、协方差特征等），因此可以利用统一的提取流程进行特征提取；从而也可以借助统一框架对特征提取细节进行合理评估，帮助研究人员更好的理解每个

特征设计环节的作用，启发大家设计性能更好的特征表示。

- 其二，目前主流的特征表示方法，包括底层视觉特征设计和高层语义特征学习，都利用单一特征向量对行人外观进行整体表示。我们认为，这种单一向量表示的方法容易导致细节信息的丢失，而外观细节往往是实现精准行人匹配的关键；因此，我们需要设计能更好保留行人图像或者视频中关键信息的特征表示方法，比如基于特征集合或者特征序列的特征表示。
- 其三，由于监控场景下行人外观变化的复杂性，多数特征表示方法都采用多特征融合的方式进行行人外观刻画；然而，目前常用的特征融合方法多数基于启发式的融合策略，比如直接串联不同特征等，无法保证获得最优的组合特征。因此，我们有必要设计基于数据驱动的、自适应的特征融合策略，进一步提升特征表示的有效性。

2.2 特征匹配算法

特征匹配是行人再识别中的关键步骤，特征间的距离度量或者匹配相似度直接决定着系统判定样本对为“同一行人”，还是“不同行人”。为了做出正确的判断，ReID系统对特征匹配算法的基本要求就是，保证具有同一身份的行人特征之间的距离（或相似度）小于（或大于）具有不同身份的行人特征之间的距离（或相似度），如图2-2所示。在这一基本原则的基础上，研究人员主要从度量学习、投影学习、以及局部对应关系学习三个方面对特征匹配算法进行探索。下文我们将从这三各方面对涉及到的主要算法进行一一概述。

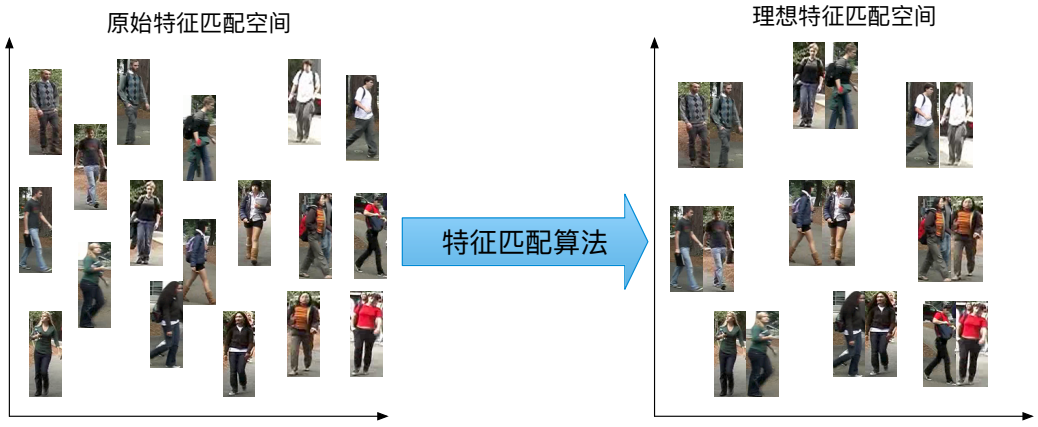


图 2-2 特征匹配算法效果示意图。

2.2.1 度量学习

度量学习是一类最直接的特征匹配算法，直接学习任务专用的距离度量方式。如果定义距离度量函数为 $d(\cdot,\cdot;\boldsymbol{\theta})$，度量学习的目标是，对于任意的输入样本对 $(\mathbf{x},\mathbf{y})$ 或 $(\mathbf{x},\mathbf{z})$ 都满足如下公式：

$$\begin{aligned} d(\mathbf{x},\mathbf{y};\boldsymbol{\theta}) &< \varepsilon_1, \\ d(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) &> \varepsilon_2, \end{aligned} \tag{2-1}$$

其中 $\mathbf{x}$ 和 $\mathbf{y}$ 为具有相同身份标签的样本对的特征表示，$\mathbf{x}$ 和 $\mathbf{z}$ 为具有不同身份标签的样本对的特征表示，ε_1 和 ε_2 分别为依据经验设定的阈值，$\boldsymbol{\theta}$ 为距离度量函数的参数。除了以上判别性的约束以外，$d(\cdot,\cdot;\boldsymbol{\theta})$ 还往往必须满足距离度量的非负性与对称性等约束。

经典的线性马氏距离（Mahalanobis Distance）是一种最为常用的距离度量模型，其数学模型为：

$$d(\mathbf{x},\mathbf{y};\mathbf{M}) = \sqrt{(\mathbf{x}-\mathbf{y})^T\mathbf{M}(\mathbf{x}-\mathbf{y})}, \tag{2-2}$$

其中 $\mathbf{M}$ 是要学习的马氏矩阵，为了满足距离度量的非负性和对称性，$\mathbf{M}$ 通常为（半）正定对称矩阵。在马氏距离的标准型中，$\mathbf{M}$ 为协方差矩阵的逆，此时马氏距离是一种去耦合的且量纲无关的距离度量；当 $\mathbf{M}$ 为单位阵时，马氏距离退化为欧氏距离（Euclidean Distance）。然而，在监控场景行人匹配问题上，标准马氏距离和欧氏距离往往并不足以刻画复杂的特征空间分布，因此无法合理地度量特征之间的距离。大量研究人员基于对行人再识别问题的理解、以及对行人特征空间的先验认识，尝试以数据驱动的方式，在不同约束条件下学习适合当前任务的马氏距离模型。

一般情况下，研究人员会假设存在一个全局的最优马氏矩阵，并适用于所有的输入样本。Zheng 等人[23,31] 提出概率相对距离比较（Probabilistic Relative Distance Comparison，PRDC）算法，从概率的角度去优化马氏距离模型，使正样本对距离比负样本对距离小的概率在训练数据上最大化。Li 等人[27] 提出局部自适应决策函数（Locally-Adaptive Decision Functions，LADF）模型，将马氏距离引入到决策函数学习中，并将固定的决策阈值替换为参数化阈值函数，构建基于样本对的二分类问题（即判断输入样本对是否为同一类）。考虑到常规的度量学习模型，在训练过程中往往需要利用迭代的梯度下降算法进行模型优化，因此模型学习收敛速度较慢。Kostinger 等人[24] 提出了一种简单直观的距离度量（Keep It Simple and Straightforward Metric，

KISSME）模型，通过假设样本对的特征差异服从高斯分布，并最大化样本对二分类的似然比检验，得到马氏矩阵 **M** 的一个最优快速解。由于在基于视频的ReID 中，不同身份的行人除了可能具有相似的外观以外，经常还具有相似的运动特征，因此与基于图像的ReID 相比往往更难克服类间混淆问题。You 等人[57] 提出在度量学习过程引入 Top-Push 约束，即在缩小类内差异增大类间距离的同时，对距离最小的负样本对增大惩罚力度。除此之外，也有大量基于前人工作的改进模型被提出。比如，由于在分类或者匹配任务中，低维紧致的特征向量往往更有效，Liao 等人[18] 将特征降维融合到KISSME 模型中，并将其转化为一个最大化广义瑞利商（Generalized Rayleigh Quotient，GRQ）问题；Yang 等人[44] 为了增加距离度量模型的判别力，除了在KISSME 模型中对样本对的特征差异进行高斯建模外，还对样本对的特征共性同样进行高斯建模，引入了额外的样本对相似性信息；等等。

为了更好的利用数据分布的多样性，一些研究人员将局部多距离度量学习方法引入到ReID 任务中。Zhou 等人[52] 提出一种在线的局部自适应距离度量算法，为每个查询样本生成专用的距离度量模型。其基本思路是先学习一个离线的全局距离度量模型，然后对于每一个查询样本快速地从数据中采样若干样本组成负样本对，最后基于这些负样本对调节全局模型成为专用的严格半正定的局部距离度量矩阵。Sun 等人认为由于行人外观特征受行人姿态、拍摄角度等多种因素影响，因此在特征匹配过程中需要为不同的影响因素学习不同的距离度量模型。基于这些发现和动机，研究人员在[51] 中提出了含有隐变量的度量学习模型，使具有不同特点的样本对可以利用不同的局部度量模型进行距离计算。

由于数据分布的复杂性，线性的距离度量如马氏距离等有时无法很好地刻画数据的相似性，因此也有研究人员通过学习非线性的距离度量模型来解决ReID 中的匹配问题。Huang 等人[124] 提出非线性局部度量学习（Nonlinear Local Metric Learning，NLML）算法，利用局部多距离度量学习来刻画数据的局部分布，利用深度神经网络来刻画数据的非线性特点。

2.2.2 投影学习

由于在ReID 中，待匹配行人图像往往采集于不同的摄像头，受拍摄环境的影响，不同摄像头下的行人外观通常具有不同的特征分布，因此无法直接进行有效匹配；为了使样本对具有更强的可比性，研究人员一般会先将原始特征投影到统一的公共特征空间中，然后再进行特征匹配。另外，经典的马氏距离也可从特征投影的角度

理解。由于半正定参数矩阵 $\mathbf{M}$ 等价于 $\mathbf{L}^T\mathbf{L}$，因此马氏距离可以等价于如下特征投影形式：

$$d(\mathbf{x},\mathbf{y};\mathbf{L}) = \sqrt{(\mathbf{x}-\mathbf{y})^T\mathbf{L}^T\mathbf{L}(\mathbf{x}-\mathbf{y})} = \sqrt{(\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{y})^T(\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{y})}, \tag{2-3}$$

其中 $\mathbf{L}$ 为投影矩阵。

与度量学习方法类似，大量研究工作分别从学习全局投影[25,26,28,42,46,59,125]、局部投影[29]、以及非线性投影[33] 等多个角度探究适合于ReID 任务的特征投影方法。Mignon 等人[25] 提出成对约束成分分析 (Pairwise Constrained Component Analysis，PCCA) 算法学习全局投影矩阵，将原始高维特征投影到低维紧致特征空间内，并保证投影后的正样本对距离小于某个阈值，同时负样本对距离大于某个阈值。Li 等人[29] 提出局部对齐特征变换 (Locally Aligned Feature Transforms，LAFT) 算法，通过学习局部投影来克服跨视域特征投影函数的多峰特点。具体来讲，研究人员先将行人图像划分为若干个子类别，同时学习每个子类别内部的跨视域特征投影函数；给定待匹配样本对之后，先确定样本所属的子类别，然后利用对应的局部投影函数将样本对投影到统一的公共特征空间进行匹配。Xiong 等人[33] 利用核方法将PCCA、LFDA、边缘费舍尔分析 (Marginal Fisher Analysis，MFA) 等线性投影模型转化为非线性模型，进一步提升了ReID 的准确度。

在ReID 中，采集于不同摄像头下的行人图像通常会因拍摄环境的不同，而在外观形态上呈现不同的特点，因此很难通过学习一个通用的投影矩阵将跨摄像头的行人特征投影到共享的特征子空间中。为此，Li[29]、Yu[125] 等人提出非对称的特征投影学习，为每个摄像头学习不同的投影矩阵，从而更自然地发掘不同摄像下行人的共享特征空间。对于采集于不同摄像下的样本对 $(\mathbf{x},\mathbf{y})$，非对称的特征投影学习可以表示为：

$$d(\mathbf{x},\mathbf{y};\mathbf{L},\mathbf{H}) = \sqrt{(\mathbf{L}\mathbf{x}-\mathbf{H}\mathbf{y})^T(\mathbf{L}\mathbf{x}-\mathbf{H}\mathbf{y})}, \tag{2-4}$$

其中 $\mathbf{L}$ 和 $\mathbf{H}$ 分别为不同摄像头下对应的投影矩阵。

在ReID 中，除了通过直接学习特征投影矩阵来建立跨摄像头的特征关联外，有些研究人员也尝试利用字典学习与稀疏表示 (Dictionary Learning and Sparse Representation，DL-SR) 的思想来建立跨摄像头下行人的联系。由于，DL-SR 可以看做将原始特征投影到所学字典张成的特征空间，因此我们也可以将此类算法归为特征投影学习。利用稀疏表示分类 (Sparse Representation Classification，SRC) 原理，可以很容易地实现跨摄像头的行人身份识别。具体来讲，首先可以将某一摄像头下的

具有同一身份的行人图像组成子字典 $\mathbf{D}_c$，并将所有子字典组成完整字典 $\mathbf{D}=[\mathbf{D}_c]_{c=1}^{C}$；接着将采集于另一摄像头下的待查询行人图片表示成词典 $\mathbf{D}$ 中稀疏个元素的加权组合；最后根据其在每个子字典上的重构误差大小确定待查询行人图片所属的身份类别。该过程可表示为如下数学公式，

$$\begin{aligned}&\text{构建字典：}\quad \mathbf{D}=[\mathbf{D}_1,\mathbf{D}_2,...,\mathbf{D}_C],\ \text{其中}\ \mathbf{D}_c=[\mathbf{x}_{c_1},\mathbf{x}_{c_2},...,\mathbf{x}_{c_{k_c}}];\\&\text{稀疏表示：}\quad \boldsymbol{\beta}=\arg\min_{\boldsymbol{\beta}}||\mathbf{y}-\mathbf{D}\boldsymbol{\beta}||_2^2+\lambda_1||\boldsymbol{\beta}||_1+\lambda_2 r(\boldsymbol{\beta});\\&\text{身份匹配：}\quad \mathbf{y}\text{的身份标签}=\arg\min_{c}||\mathbf{y}-\mathbf{D}\delta_c(\boldsymbol{\beta})||_2^2;\end{aligned}\tag{2-5}$$

其中，k_c 为身份标签为 c 的行人图片数目，$r(\boldsymbol{\beta})$ 代表某些结构化的约束条件，δ_c 函数将 $\boldsymbol{\beta}$ 中与标签为 c 相关的系数提取出来。在此基础上，Zheng 等人[126] 提出模糊性敏感的匹配分类器（Ambiguity-Sensitive Matching Classifier，AMC）算法，以基于行人图像的局部图像块的特征为输入样本，实现不完整图片的行人再识别；Lisanti 等人[127] 提出迭代稀疏系数重分配排序（Iterative Re-weighted Sparse Ranking，ISR）算法，将SRC 从行人身份识别，推广到基于身份一致性匹配的行人检索排序。利用对偶字典学习（Coupled Dictionary Learning，CDL）原理，可以很容易地实现跨摄像头的共享特征空间的发掘。具体来讲，假设，每个摄像头下都可以学习到一个刻画了该摄像头下行人特征分布的字典，而且每个行人特征都可以被所属的摄像头下的字典进行稀疏表示；那么，如果使不同摄像头下学习到的词典保持一致，或者使同一身份标签的行人特征的稀疏表示的距离比不同身份标签的行人特征的稀疏表示的距离小，那么可以认为所学的词典刻画了跨摄像头的共享特征空间。对于采集于不同摄像头的样本对 $(\mathbf{x},\mathbf{y})$，该过程可表示为如下数学公式：

$$\min_{\mathbf{D}_\mathbf{x},\mathbf{D}_\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta}}||\mathbf{x}-\mathbf{D}_\mathbf{x}\boldsymbol{\alpha}||_2^2+||\mathbf{y}-\mathbf{D}_\mathbf{y}\boldsymbol{\beta}||_2^2+\lambda_\mathbf{x}||\boldsymbol{\alpha}||_1+\lambda_\mathbf{x}||\boldsymbol{\beta}||_1+\lambda r(\mathbf{D}_\mathbf{x},\mathbf{D}_\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta}),\tag{2-6}$$

其中 $r(\mathbf{D}_\mathbf{x},\mathbf{D}_\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta})$ 为对词典或者稀疏表示的一致性约束条件。在此基础上，Liu[32]、Jing[103] 等人通过对稀疏表示的一致性约束，分别实现了跨摄像头和跨分辨率的行人再识别；Karanam 等人[38] 通过同时对词典和稀疏表示的一致性约束，实现了基于多张图的行人再识别。

2.2.3 局部对应关系学习

对于图像特征的匹配，如果先将空间局部特征聚合成全局特征，再在向量空间内进行匹配，会造成部分空间位置信息的丢失；类似的，对于视频序列特征的匹配，

如果先将时空局部特征聚合成全局特征，再进行匹配，同样会造成部分时空信息的丢失。因此，在ReID 中，为了更好地刻画行人细节，一些研究学者提出直接基于局部特征向量集合进行相似度匹配。然而，由于待匹配的行人图像对之间通常无法直接进行位置对齐，而待匹配的行人跟踪视频之间通常也无法直接进行时序对齐，因此局部特征集合之间往往也是语义不对齐的，无法直接进行合理的距离计算或相似度匹配。因而，基于局部特征集合的行人匹配，关键是局部对应关系学习。

有些局部对应关系学习算法，在训练过程中，单独学习每个局部特征的最优匹配。Zhao 等人[13,30] 提出利用带有邻域约束的贪婪搜索方法，为每一个图像块在对应的被匹配图像的邻域位置搜索外观最相似的图像块组成匹配对，并将图像块的显著性引入到图像块匹配对的相似度计算过程中；最终两幅行人图像的相似度，即为所有相匹配的图像块对相似度之和。Wang 等人[54,60] 提出可判别视频排序（Discriminative Video Ranking，DVR）算法，将基于视频的行人再识别构造成排序问题，从视频序列中选择最具辨别力的跨视域子视频段对，以此来计算视频序列之间的距离度量。

另外，有些局部对应学习算法，在训练过程中，全局考量所有局部特征之间的最优匹配。Shen 等人[40,53] 提出基于助推机制（Boosting）学习ReID 中行人图像间的局部对应结构，该结构反映了图像块之间的匹配概率。Zhou 等人[128] 将局部对应学习过程构造成图匹配模型，学习若干典型匹配结构。

2.2.4 特征匹配算法总结

大量研究人员已经分别从度量学习、投影学习、局部对应性学习等方面，对ReID 中的特征匹配算法进行了深入研究，并提出了许多行之有效的算法，大幅度提升了ReID 的准确度。然而，针对特定复杂应用场景进行特征匹配算法的优化，仍是高精度ReID 算法设计的必要工作。

- 其一，目前大部分特征匹配算法都是基于特征向量间的距离度量进行模型优化，如果行人的特征向量中丢失了某些关键细节信息，会直接造成匹配算法的失效。因此，我们认为可以利用特征序列来尽可能全面地刻画行人外观细节，因而也需要设计合理的序列匹配算法来进行行人匹配。
- 其二，目前大部分先进的特征匹配算法，通常需要在大量标注数据集驱动下，才能训练和学习到性能优异的匹配模型；而实际应用中，跨摄像头跨视域的行人样本对采集往往非常困难，因而样本不足限制了高性能特征匹配算法的应用。因此，我们需要针对样本不足问题，设计合适的匹配模型，使其在小数据

集上仍能具有良好的泛化能力。

2.3 深度神经网络算法

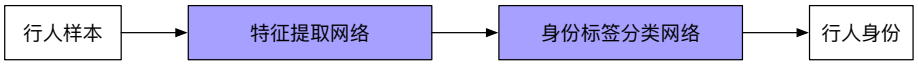
经典的ReID 算法，一般将再识别过程分解成特征表示和特征匹配两个级联的步骤，分别进行算法设计和优化。这种分步骤的算法设计流程有可能无法达到ReID 系统的整体最优性能，而且其中的特征表示方法往往基于手工特征，过于依赖设计者的个人经验。随着近几年深度神经网络技术的日益成熟、在计算机视觉中应用的日益广泛、以及大规模ReID 数据集的普及，大量研究人员开始提出数据驱动的端到端的ReID 深度模型。下文我们将从深度模型的网络结构和损失函数设计这两个方面对相关算法进行详细概述。

2.3.1 网络结构

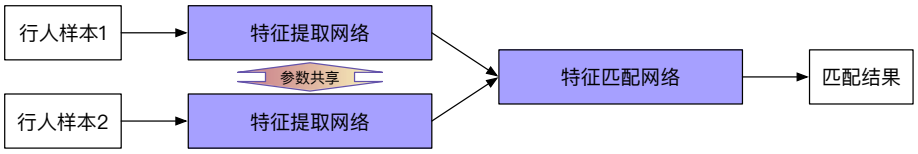
ReID 中的神经网络在训练阶段与测试阶段（或者推断阶段）往往具有不同的网络结构。在训练阶段，如果将ReID 当做分类任务处理，其网络结构往往只包含一个分支，用于特征提取及身份标签分类，如图2-3(a) 所示；如果将ReID 当做匹配任务处理，其网络结构往往包含多个参数共享的子网络分支进特征提取，并连接一个匹配子网络实现特征匹配，如图2-3(b) 和 (c) 所示。在测试或者推断阶段，ReID 通常需要输入待匹配的样本对，并返回匹配结果，因此其网络结构通常包含两个参数共享的特征提取子网络，并连接一个特征匹配子网络，如图2-3(d) 所示。由于特征表示和特征匹配是ReID的主要组成部分，因此在进行网络结构设计时，研究人员往往也主要从这两个方面进行考量。

2.3.1.1 从特征表示角度设计神经网络

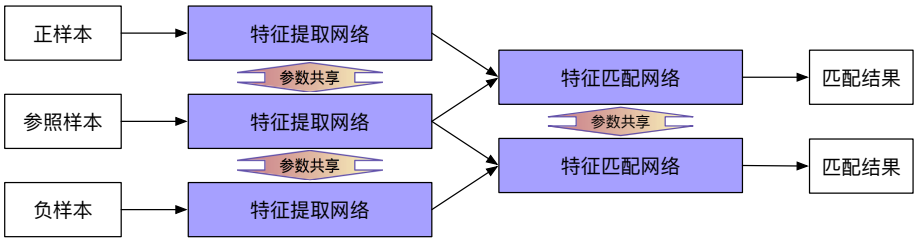
CNN 通常被用作行人图像特征的提取器，将网络的全连接层输出或者卷积层输出作为特征表示。在[45,66–70,72] 中，受行人数据集规模的限制，研究人员一般为ReID 定制网络层数较浅、参数较少的专用CNN 结构，实现行人图像特征表示的学习。为了更好地利用深层网络强大的学习和表达能力，也有研究人员将成熟的深度神经网络结构借鉴到ReID 任务中，比如[71,77,79] 以 AlexNet[120] 为基础、[73,83,86] 以 GoogLeNet[129] 为基础、[1,87] 以 ResNet-50[123] 为基础等，实现特征提取模块。需要注意的是，一般深度的神经网络会先在 ImageNet 等大规模图像分类数据集上与训练后，再细调到ReID 任务中。



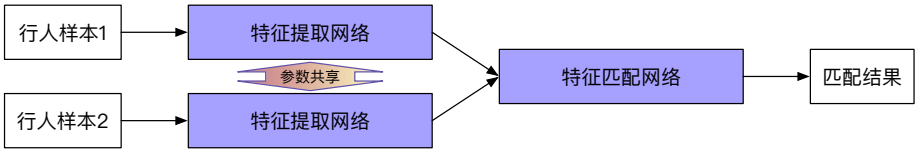
(a) 训练阶段，基于目标分类的网络结构



(b) 训练阶段，基于样本对匹配的网络结构



(c) 训练阶段，基于三元组匹配的网络结构



(d) 测试阶段网络结构

图 2-3 行人再识别中常用网络结构。三元组中的正样本为与参考样本身份一致的行人，负样本为与参考样本身份不同的行人。

循环神经网络（Recurrent Neural Network，RNN）结构也被引入到行人再识别中实现对序列信号的特征提取。在[56,62–64]中，研究人员先利用CNN 从行人跟踪视频帧中提取图像特征，再利用RNN 提取帧特征序列中的时序特征，从而更完整的刻画行人视频的信息；而[58]中直接提取帧图像中的手工特征，再利用RNN 提取时序信息。在[49]中，研究人员先将行人图像在空间上从上到下划分为若干图像条序列，再利用RNN 序列化地提取图像条的特征，从而更好地利用了图像的空间上下文信息，提升了所学特征的判别能力。

然而受行人检测框精度、以及拍摄环境遮挡等影响，行人图像或跟踪序列往往会因含有局部干扰，导致所学全局特征受局部噪音污染，降低特征质量；另外，对于衣着外观相似的目标，图像或者视频序列中的局部细节信息往往是区分不同行人的关键。为此，有些研究人员在设计特征提取网络结构时，往往会更关注于对高质量局部细节的发掘。在[70,80,88]中，研究人员将行人图像划分为若干图像条，利用多个特征

提取子网络提取局部图像条和全局完整图像的深度特征，并将其融合作为最终的特征表示。与直接强制对图像进行分割不同，在[82,86]中，研究人员利用辅助的人体分割网络将行人图像自适应地划分为若干身体部分图像块，并学习和融合基于身体部分的特征表示。除了显式按照身体部分对图像进行区域划分以外，在[61,62,64,76,83–85]中，研究人员利用深度学习中的注意力机制，通过数据驱动的优化学习策略，自动地发掘行人图像或者跟踪视频中的关键局部空间特征或时空特征。

2.3.1.2 从特征匹配角度设计神经网络

在利用CNN 提取行人图像的特征表示时，由于中间层输出的特征图谱（Feature Map）不仅包含了图像的抽象语义特征，同时还保留着局部特征间的空间位置关系，因此常被用来直接计算图像间的距离度量或相似度；同样，RNN 提取视频的特征表示时，输出的序列特征同样保留了局部特征的时空关系，因此也常被用来直接计算视频间的距离度量或相似度。此时，如何设计匹配网络，保证特征图谱或者序列特征间合理的局部对应关系成为关键。

在[68,73]中，研究人员直接将待比较的两幅图像的特征图谱或特征图谱的差异矩阵堆叠在一起，在此基础上利用卷积操作学习基于特征图谱的相似度；在[62]中，研究人员为了合理地比较视频序列的相似度，将视频中待比较的帧图像对的特征图谱的差异矩阵堆叠在一起，并利用空间的RNN 网络按照不同方向对差异矩阵集合进行汇总。Li 等人[66]在设计特征匹配网络时，提出图像块匹配（Patch Matching）模块，将特征图谱中每个位置的局部特征与待比较特征图谱中的所有处在同一水平位置的局部特征一一比较，并将结果汇聚成最终相似度；类似的，Varior 等人[72]设计了匹配门（Matching Gate）机制，为每个局部特征从与其比较的特征图谱的同一水平位置上选择对应的局部特征。Ahmed[67]、Subramaniam[74]等人提出跨输入邻域差异（Cross-Input Neighborhood Difference）计算过程，将特征图谱中每个位置的局部特征与待比较特征图谱中的邻域位置的局部特征一一比较，并将结果汇总成最终匹配结果。

2.3.2 损失函数

在端到端的深层神经网络模型中，我们除了需要设计合理的网络结构来从原始数据中抽取有效信息，还需要设计合理的损失函数来指导模型的学习和训练。下文我们将对ReID 中神经网络模型常用的损失函数进行详细概述。

2.3.2.1 Cross Entropy Loss

考虑到在分类任务上性能良好的特征表示，往往具有很强的抗干扰能力和很高的可区分度，因此研究人员经常将ReID 深度模型当做分类器进行训练，并将学习到的模型作为特征提取器提取用于行人匹配的特征表示。为了使分类误差最小，通常以身份标签为监督信号，利用 Cross Entropy Loss 作为损失函数进行模型训练。训练阶段，模型的网络结构如图2-3(a)。假设对于某行人图像或者跟踪序列，其类别标签的 One-Hot 编码为 $\mathbf{l}$，深度模型得到的分类打分向量为 $\mathbf{s}$，则 Cross Entropy Loss 定义如下：

$$f = -\sum_{c=1}^{C} l_c \ln(\hat{s}_c), \tag{2-7}$$

其中 $\hat{s}_c = \frac{\exp(s_c)}{\sum_{c=1}^{C}\exp(s_c)}$，$s_c$ 和 l_c 分别为标签向量 $\mathbf{l}$ 和打分向量 $\mathbf{s}$ 中的第 c 个元素。论文[58,80,82,84,86–88,101] 中都采用了这种损失来进行模型训练。

2.3.2.2 Binary Classification Loss

将ReID 模型当做分类器进行训练，虽然可以使所学模型提取到性能良好的特征表示，但是由于没有直接优化行人匹配问题，因此无法保证行人匹配的结果。一种直接优化行人匹配性能的方式，是将深度网络当做二分类模型，对样本对进行正负样本分类。此时一般采用基于样本对的 Binary Classification Loss，监督信号就是标识样本对是否为同一类的 0/1 标签。训练阶段，模型的网络结构如图2-3(b)。假设某输入样本对的标签为 l，模型预测输入为正负样本对的打分分别为 s_1 和 s_2，则 Binary Classification Loss 定义如下：

$$f = -l\ln(\hat{s}_1) - (1-l)\ln(\hat{s}_2), \tag{2-8}$$

其中 $\hat{s}_i = \frac{\exp(s_i)}{\sum_{i=1}^{2}\exp(s_i)}$。论文[66,67,69,73,74] 中都采用了这种损失函数或者其等价变形来进行模型训练。

2.3.2.3 Contrastive Loss

另一种更直接的，优化模型在样本对匹配任务上性能的策略是，使模型预测的正样本对间的距离尽量小，同时使模型预测的负样本对间的距离大于某一阈值。这种优化策略可以通过基于样本对的 Contrastive Loss 实现。此时，训练阶段模型的网

络结构如图2-3(b)。假设输入样本对为 $(\mathbf{x},\mathbf{y})$，样本对的标签为 l，模型预测样本对的距离表示为 $d(\mathbf{x},\mathbf{y})$，则 Contrastive Loss 定义为：

$$f = ld(\mathbf{x},\mathbf{y}) + (1-l)\max(M - d(\mathbf{x},\mathbf{y}),0), \tag{2-9}$$

其中 M 为依据经验设定的阈值。论文[49,72] 中采用了这种损失函数进行模型优化。

2.3.2.4 Triplet Loss

在实际应用中，ReID 也可以被看做基于匹配的排序问题。ReID 算法往往需要返回一个与查询样本匹配距离从近到远的排序列表，而且只需保证匹配距离最近的元素为正样本即可。为此，可以通过约束正样本对的距离小于负样本对的距离，来进行模型的训练和学习。这种优化策略可以通过基于三元组的 Triplet Loss 实现。此时，训练阶段模型的网络结构如图2-3(c)。假设三元组由参照样本 **x**、正样本 **y**、和负样本 **z** 组成，则 Triplet Loss 定义如下：

$$f = \max(d(\mathbf{x},\mathbf{y}) - d(\mathbf{x},\mathbf{z}) + M,0), \tag{2-10}$$

其中 M 为依据经验设定的阈值。论文[1,45,62,76,83] 中采用了这种损失函数进行模型优化。

2.3.2.5 Multiple Loss

联合使用多种损失函数进行模型训练，往往可以进一步提升ReID 算法的性能。为了提升模型对输入样本对的区分辨别能力，论文[68] 联合了基于样本对的 Contrastive Loss 和基于三元组的 Triplet Loss，而论文[77] 联合了基于样本对的 Binary Classification Loss 和基于三元组的 Triplet Loss，进行模型训练。为了保证网络学习到性能良好的深层特征表示，同时提升模型对样本对的区分能力，论文[85] 将 Cross Entropy Loss 和 Binary Classification Loss 联合使用，论文[56,63,64] 将 Cross Entropy Loss 和 Contrastive Loss 联合使用，论文[61,71] 将 Cross Entropy Loss 和 Triplet Loss 联合使用。

2.3.3 深度神经网络算法总结

得益于深度神经网络模型强大的表达能力、以及大规模的行人再识别数据集，基于深度学习的ReID 算法使在实际应用中实现高精度再识别成为可能。根据以上对

近年来深度ReID 算法的总结和对比，我们认为针对ReID 任务，从以下几个方面进行网络设计，可以进一步提升深度模型的性能。

- 其一，由于行人再识别数据集规模相对较小，因此我们应尽量限制深度模型的复杂度，其中一个策略就是利用参数共享。比如，对于基于多分支的深度匹配模型，特征提取模块可以进行参数共享，从而减少模型参数。
- 其二，由于细节信息对高精度的行人再识别至关重要，因此可以引入基于中间层特征图谱（对于 CNN 结构特征提取网络来说）或者中间层特征序列（对于 RNN 结构特征提取网络）的特征集合匹配模块，直接在细节特征上进行距离度量或者相似度匹配;
- 其三，在有限的标注数据集上，要充分利用标注信息。比如，将ReID 任务同时当做分类和确认问题进行优化，同时利用样本身份标签和样本对分类标签。

2.4 本章小结

本章我们主要从特征表示方法、特征匹配算法、以及深度神经网络算法这三个方面对目前常用的ReID 算法进行简要概述和分析。另外，通过对常用算法的总结，进一步阐述了我们的研究动机。

第三章 基于正则化度量学习的行人再识别算法

本章分别从引言、相关工作、不同形式度量函数、正则化度量学习算法、以及相关实验结果和分析等方面，对我们提出的基于正则化度量学习的行人再识别算法进行详细介绍。

3.1 引言

VIPeR CUHK01

图 3-1 采集于数据集 VIPeR[4] 和 CUHK01[5] 中的行人图像样例。

目前，视觉监控网络已被广泛布置于城市、公路、机场、火车站等公共区域，大量的监控视频往往包含了重要的公共安全有关的视觉线索。因此，多摄像头协同监控，尤其是跨摄像头视域的行人身份一致性认证越来越成为基于内容的视频分析的关键步骤。这也就是我们本文中研究的ReID 任务。尽管，ReID 问题可以很直接地被当作图像或者视频匹配问题进行处理。然而，由于拍摄环境中光照或者行人姿态的变化、以及复杂的背景干扰或遮挡等，使高精度匹配变得非常困难。在图3-1中，我们展示了几对待处理的行人图像，以便直观地了解和认识ReID 任务中的困难和挑战。

为了实现高精度的再识别，ReID 过程通常分为了特征提取和基于度量学习的特征匹配两个部分，其中度量学习至关重要。然而不幸的是，由于在实际应用中，用于度量学习训练的数据集往往比较小，因此很容易使所学模型趋于过拟合的状态，因此降低了实际应用中ReID 算法的稳定性和准确度。本章中，我们提出利用向度量学习模型中引入正则化约束条件的方式，来限制模型的复杂度，从而提升模型在小数据集上的泛化能力。具体来说，我们将从马氏距离学习、对称投影学习、以及非对称投影学习三个角度理解度量函数，并结合对具体度量学习算法实例的正则化，来证明正则化度量学习的有效性。

3.2 相关工作

目前，大部分的ReID 算法仍然从特征表示和度量学习这两个角度进行探究，而提取到具有良好表达能力的特征表示往往是成功进行行人匹配的基础。在进行特征设计和提取时，利用人体的结构信息被证明往往可以得到更好的行人外观表示。比如，论文[10,12] 利用人体结构的对称非对称性，对行人图像中的每个像素或者局部区域分配不同的权重，可以一定程度克服背景等无效信息的干扰。然而，这种通过设计精巧特征表示来提升再识别准确率的方法，往往需要研究人员或者技术开发者具有丰富的专业领域的先验知识，这也就大大限制了算法的推广能力。除了特征设计，好的距离度量方法也是成功匹配的关键，而且距离度量的结果往往直接决定了再识别的结果。在ReID 中，研究人员一般利用度量学习算法进行距离度量方式的选择。度量学习的基本原理是，针对训练数据，学习任务专用的距离度量函数或者度量矩阵，并保证在学习到的度量空间内正样本对的距离尽量小，负样本对的距离尽量大。比如，论文[23,31] 将ReID 中的度量学习构造成相对距离学习任务，在概率上去保证同一行人的样本对的距离比不同行人样本对的距离小。由于度量学习往往是数据驱动的、自动学习的过程，因此这类算法可以比较容易的推广到其他任务上。

然而，在ReID 任务中，大规模地采集跨摄像头的成对行人图像或者视频往往是很耗时耗力，而且难以操作的。因此度量学习算法通常要基于小规模数据集进行训练和学习。而训练样本的不足，往往会致使所学得的度量函数更容易发生过拟合，从而使模型的性能无法满足实际的需求。在本章中，我们提出借助于正则化策略，限制度量学习模型的复杂度，从而提升模型在小规模数据集上的泛化能力，进一步提升ReID 的准确度。具体来说，我们从马氏距离学习、对称投影学习、和非对称投影学习三个方面研究正则化对度量学习模型的影响，并在两个中小规模基准数据集VIPeR 和 CUHK01 上验证正则化度量学习的有效性。

3.3 我们的方法

在本小节，我们将简要回顾小节2.2中总结的度量函数的三种不同形式，并在此基础提出基于正则化的度量学习算法。为清楚起见，先对本章将使用的数学符号进行简要说明：粗体小写字母 $\mathbf{x},\mathbf{y} \in \mathbb{R}^d$ 代表从行人图像中提取的特征向量；$\ell_i \in \{1,2,3,...,C\}$ 第 i 个行人特征向量所对应的行人身份标签；g_{ij} 表示样本对 $(\mathbf{x}_i,\mathbf{x}_j)$ 或者 $(\mathbf{x}_i,\mathbf{y}_j)$ 对应的匹配标签，如果 $\ell_i = \ell_j$ 则 $g_{ij} = +1$，否则 $g_{ij} = -1$。

3.3.1 不同形式的度量函数

一般情况下，我们可以将度量函数定义为最经典的马氏距离形式，如下：

$$d^2_{\mathbf{M}}(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y})^T\mathbf{M}(\mathbf{x}-\mathbf{y}), \tag{3-1}$$

其中，$\mathbf{M} \in \mathbb{S}^d_+$，$\mathbb{S}^d_+$ 代表对称半正定(Positive Semi-Definite，PSD) 矩阵集合。在 $\mathbf{M} \in \mathbb{S}^d_+$ 约束下，可以将 $\mathbf{M}$ 表示为 $\mathbf{L}^T\mathbf{L}$，其中 $\mathbf{L} \in \mathbb{R}^{k\times d}$，因此我们可以进一步将公式3-1表示为对称投影形式，如下：

$$d^2_{\mathbf{M}}(\mathbf{x},\mathbf{y}) = (\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{y})^T(\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{y}). \tag{3-2}$$

用公式3-1和公式3-2我们可以建立起度量学习与线性投影之间的联系。考虑到，在行人再识别任务中，待匹配的样本对 $\mathbf{x}$ 和 $\mathbf{y}$ 通常采集于跨视域的不同摄像头下，因此我们需要可以刻画各自视域内特点的专用投影矩阵对样本对进行映射，为此我们将公式3-2进一步重新定义为非对称投影形式，如下：

$$d^2_{\mathbf{M}}(\mathbf{x},\mathbf{y}) = (\mathbf{L}\mathbf{x}-\mathbf{H}\mathbf{y})^T(\mathbf{L}\mathbf{x}-\mathbf{H}\mathbf{y}), \tag{3-3}$$

其中 $\mathbf{L},\mathbf{H} \in \mathbb{R}^{k\times d}$。

3.3.2 正则化度量学习算法

正则化是一种可以解决病态问题（Illposed Problem）或者缓解过拟合问题（Overfitting Problem）的有效方法。在本章中，我们尝试将正则化方法引入到ReID 中的度量学习算法当中。具体来说，我们将包括最大间隔最近邻 (Large Margin Nearest Neighbors，LMNN)[130]、费舍尔判别分析 (Fisher Discriminant Analysis，FDA) 、以及决策函数学习 (Decision Function Learning，DFL)[27] 在内的三种不同的度量学习方法，分别改进为相应的正则化版本，并且给出相应的模型优化算法。

3.3.2.1 正则化的最大间隔最近邻算法

对于样本 $\mathbf{x}$，我们将样本空间中理应与其保持较小距离或者具有相同身份标签的样本称为目标邻域样本，将与 $\mathbf{x}$ 之间的距离小于目标领域样本与 $\mathbf{x}$ 之间的距离，并且具有不同身份标签的样本称为冒名邻域样本。LMNN 算法试图通过拉近目标邻域样本，同时推远冒名邻域样本的方式，学习到一个最优的PSD 矩阵 $\mathbf{M}$。该过程可以

总结为如下优化问题：

$$\begin{aligned}\min_{\mathbf{M}\in\mathbb{S}_+^d}\ &(1-\mu)\sum_{i,j\rightsquigarrow i} d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)+\mu\sum_{i,j\rightsquigarrow i,l}(1-g_{il})\xi_{ijl},\\ \text{s.t.}\ \ &d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_l)-d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)\geq 1-\xi_{ijl},\ \ \xi_{ijl}\geq 0,\end{aligned} \tag{3-4}$$

其中，$0\leq\mu\leq 1$ 是用于平衡拉近目标样本和推远冒名样本作用的权重参数，$j\rightsquigarrow i$ 表示样本 $\mathbf{x}_j$ 是样本 $\mathbf{x}_i$ 的目标邻域样本，ξ_{ijl} 为松弛变量，而度量函数 $d^2_{\mathbf{M}}(\cdot,\cdot)$ 采用公式3-1中的形式。

尽管公式3-4表示的LMNN 算法，可以利用成熟的半正定规划（Semi-Definite Programming，SDP）求解器，进行优化求解，但是由于ReID 中训练样本的不足，使优化结果很容易趋于过拟合。为了克服这个问题，我们将参数的核范数约束 $\|\mathbf{M}\|_*$ 作为一种正则项，引入到原始LMNN 算法中，并提出核范数正则化的 LMNN（nuclear norm Regularized LMNN，nuLMNN）算法，数学模型如下：

$$\begin{aligned}\min_{\mathbf{M}\in\mathbb{S}_+^d,j\rightsquigarrow i}\ &\sum_{i,j\rightsquigarrow i}(1-\mu)d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)+\sum_{i,j\rightsquigarrow i,l}\mu(1-g_{il})\xi_{ijl}+\lambda\|\mathbf{M}\|_*\\ \text{s.t.}\ \ &d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_l)-d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)\geq 1-\xi_{ijl},\ \ \xi_{ijl}\geq 0,\end{aligned} \tag{3-5}$$

其中，λ 是正则化参数。基于核范数的正则化约束，可以促使LMNN 学习到一个低秩的PSD 矩阵 $\mathbf{M}$；而在利用低秩矩阵 $\mathbf{M}$ 计算距离度量时，同时隐含了特征选择操作，因此会使距离度量计算更稳定。另外，需要指出的是，对于一个PSD 矩阵 $\mathbf{M}$，存在 $\mathrm{tr}(\mathbf{M})=\|\mathbf{M}\|_*$，因此我们也可将公式3-5称为迹范数正则化的 LMNN（trace norm Regularized LMNN，trLMNN）算法。除此之外，另一种正则化约束方式，是直接促使模型学习到更简单的距离度量方式。为此，我们引入论文[131] 中的 LogDet 散度项作为我们的正则化约束条件，其定义如下：

$$D_{ld}(\mathbf{M},\mathbf{M}_0)=\mathrm{tr}(\mathbf{M}\mathbf{M}_0^{-1})-\log\det(\mathbf{M}\mathbf{M}_0^{-1})-d, \tag{3-6}$$

其中，$\mathbf{M},\mathbf{M}_0\in\mathbb{S}_+^d$。按照论文[131] 中的证明，可以知道，利用 LogDet 散度测量两个马氏矩阵相似性的物理意义是，测量两个马氏矩阵所代表的超球体数据分布的重合性。因此当令 $\mathbf{M}_0$ 为单位矩阵时，基于 LogDet 散度的正则化约束，可以促使模型学习到的距离度量方式更接近于欧氏距离，使距离计算更加简单。通过将 $\lambda D_{ld}(\mathbf{M},\mathbf{M}_0)$ 引入到公式3-4中，我们可以到相应的LogDet 散度正则化的 LMNN（LogDet regularized

LMNN，ldLMNN）模型，如下：

$$\begin{aligned}\min_{\mathbf{M}\in\mathbb{S}^d_+}&\sum_{i,j\rightsquigarrow i}(1-\mu)d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)+\sum_{i,j\rightsquigarrow i,l}\mu(1-g_{il})\xi_{ijl}+\lambda D_{ld}(\mathbf{M},\mathbf{M}_0)\\ \text{s.t. }& d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_l)-d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)\geq 1-\xi_{ijl},\ \ \xi_{ijl}\geq 0.\end{aligned} \tag{3-7}$$

正则化的LMNN 优化求解过程也可以采用原始LMNN 问题中的SDP 算法，而只需要在梯度计算过程中进行微小改动。在第 t 次优化迭代时，trLMNN 和ldLMNN 的梯度计算过程分别如公式3-8和公式3-9所示：

$$\mathbf{G}_t=(1-\mu)\sum_{i,j\rightsquigarrow i}\mathbf{C}_{ij}+\mu\sum_{i,j\rightsquigarrow i,l}(\mathbf{C}_{ij}-\mathbf{C}_{il})+\lambda\mathrm{I}, \tag{3-8}$$

$$\mathbf{G}_t=(1-\mu)\sum_{i,j\rightsquigarrow i}\mathbf{C}_{ij}+\mu\sum_{i,j\rightsquigarrow i,l}(\mathbf{C}_{ij}-\mathbf{C}_{il}),+\lambda(\mathbf{I}-\det(\mathbf{M}_t)\mathbf{M}_t^{-1}) \tag{3-9}$$

其中 $\mathbf{C}_{ij}=(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T$。此外，由于矩阵 $\mathbf{M}_t$ 有时可能是不可逆的，因此我们常用 $\hat{\mathbf{M}}_t=(1-\alpha)\mathbf{M}_t+\frac{\alpha}{N}\mathrm{tr}(\mathbf{M}_t)\mathbf{I}$ 代替原始 $\mathbf{M}_t$，其中 $0\leq\alpha\leq 1$，N 为样本数。

3.3.2.2 正则化的费舍尔判别分析算法

FDA 是一种应用最为广泛的有监督投影降维方法，可以被看做一种基于公式3-2的对称投影度量学习算法。FDA 以及它的变种也被广泛应用到ReID 任务中。这一类算法的一个优势是，算法的优化求解过程可以被转化为求解广义特征值问题，因此可以快速高效地得到最优模型。FDA 的优化目标函数定义如下：

$$\max_{\mathbf{L}\in\mathbb{R}^{k\times d}}\ \frac{\mathrm{tr}(\mathbf{L}\mathbf{S}^b\mathbf{L}^T)}{\mathrm{tr}(\mathbf{L}\mathbf{S}^w\mathbf{L}^T)}, \tag{3-10}$$

其中，$\mathbf{S}^w\in\mathbb{R}^{d\times d}$ 为类内散度矩阵，$\mathbf{S}^b\in\mathbb{R}^{d\times d}$ 为类间散度矩阵。为了求解上述问题，我们通常希望矩阵 $\mathbf{S}^w$ 必须是可逆的，以保证优化问题的可行性。然而，当训练样本数据集大小 N 比特征的维数 d 小时，$\mathbf{S}^w$ 通常是奇异的。为了保证优化过程的稳定性，我们将原始的 $\mathbf{S}^w$ 代替为正则化的形式 $\hat{\mathbf{S}}^w$，如下：

$$\hat{\mathbf{S}}^w=(1-\alpha)\mathbf{S}^w+\frac{\alpha}{N}\mathrm{tr}(\mathbf{S}^w)\mathbf{I}, \tag{3-11}$$

其中，$0\leq\alpha\leq 1$。我们将修改后的方法称为稳定费舍尔判别分析（stable Fisher Discriminant Analysis，sFDA），其求解过程依然可以利用广义特征值分析。

3.3.2.3 正则化的决策函数学习算法

对于一个线性可分的二分类问题，我们通常可以找到一个最优超平面 $f(\cdot)=0$ 将样本空间划分为两部分，使所有正样本落在超平面的一侧，而所有负样本落在另一侧。在ReID 任务中，算法的终极目标是判断行人图像或者视频对 $(\mathbf{x}_i,\mathbf{x}_j)$ 是否具有同一身份。因此ReID 也可以构造成以样本对为输入的二分类问题，同时可以设计算法去寻找最优的决策超平面 $f(\cdot,\cdot)=0$。决策函数通常定义如下：

$$f(\mathbf{x}_i,\mathbf{x}_j)=d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)-t, \tag{3-12}$$

其中 t 为决策阈值。

前面所介绍的度量学习方法，往往假设采集于跨视域的不同摄像头下的行人特征向量来自于相同的特征空间，可以表示为 $(\mathbf{x}_i,\mathbf{x}_j)$；然而，由于每个摄像头受拍摄环境或者硬件条件的影响，拍摄到的行人图像往往具有各自的特点，因此对应的行人特征向量更有可能来自于不同的特征空间，可以表示为 $(\mathbf{x}_i,\mathbf{y}_j)$。基于以上的动机，在利用投影变换发掘跨摄像头共享特征空间时，针对每个摄像头更应采用各自不同的投影矩阵，可表示为 $\mathbf{L},\mathbf{H}\in\mathbb{R}^{k\times d}$。为此我们可以将度量学习构造成如公式3-3所表示的非对称投影形式。

在我们的算法中，我们不仅利用了样本 $\mathbf{x}_i$ 和样本 $\mathbf{y}_j$ 采集于不同摄像头下的特点，还将固定决策阈值替换为类似于论文[27] 中的参数化阈值函数。阈值函数被定义为二次型，如下：

$$t(\mathbf{x}_i,\mathbf{y}_j)=\frac{1}{2}\mathbf{x}_i^T\tilde{\mathbf{A}}\mathbf{x}_i+\frac{1}{2}\mathbf{y}_j^T\tilde{\mathbf{B}}\mathbf{y}_j+\mathbf{x}_i^T\tilde{\mathbf{C}}\mathbf{y}_j+\mathbf{w}^T(\mathbf{x}_i+\mathbf{y}_j)+\mathbf{w}_0, \tag{3-13}$$

其中，$\tilde{\mathbf{A}}\in\mathbb{S}^d$，$\tilde{\mathbf{B}}\in\mathbb{S}^d$，$\tilde{\mathbf{C}}\in\mathbb{R}^{d\times d}$，$\mathbf{w}\in\mathbb{R}^d$，$\mathbf{w}_0\in\mathbb{R}$，并且 $\mathbb{S}^d$ 代表对称矩阵。基于以上假设，我们将决策函数重新构造成：

$$\begin{aligned} f(\mathbf{x}_i,\mathbf{y}_j)&=d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{y}_j)-t(\mathbf{x}_i,\mathbf{y}_j)\\ &=\frac{1}{2}\mathbf{x}_i^T\mathbf{A}\mathbf{x}_i+\frac{1}{2}\mathbf{y}_j^T\mathbf{B}\mathbf{y}_j-\mathbf{x}_i^T\mathbf{C}\mathbf{y}_j-\mathbf{w}^T(\mathbf{x}_i+\mathbf{y}_j)-\mathbf{w}_0, \end{aligned} \tag{3-14}$$

其中，$d^2_{\mathbf{M}}(\mathbf{x}_i,\mathbf{y}_j)$ 采用公式3-3中的非对称投影形式，$\mathbf{A}=2\mathbf{L}^T\mathbf{L}-\tilde{\mathbf{A}}$，$\mathbf{B}=2\mathbf{H}^T\mathbf{H}-\tilde{\mathbf{B}}$，并且 $\mathbf{C}=2\mathbf{L}^T\mathbf{H}+\tilde{\mathbf{C}}$。很显然，参数 $\mathbf{A}\in\mathbb{S}^d$ 和 $\mathbf{B}\in\mathbb{S}^d$ 是对称矩阵，而参数 $\mathbf{C}\in\mathbb{R}^{d\times d}$ 没有对称性的约束。

最优决策面可以通过优化以下正则化的决策函学习 (regularized Decision Function

Learning，rDFL）问题得到，如下所示：

$$\min_{\Theta} \quad \ell(\Theta) = \sum_i \sum_j h_\beta \left(g_{ij} f(\mathbf{x}_i, \mathbf{y}_j) \right) + r(\Theta), \tag{3-15}$$

其中，$\Theta = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{w}\}$，$r(\Theta) = \lambda_1 ||\mathbf{A}||_F^2 + \lambda_2 ||\mathbf{B}||_F^2 + \lambda_3 ||\mathbf{C}||_F^2$，并且 $h_\beta(x) = \frac{1}{\beta} \log(1 + e^{\beta x})$ 是 Hinge Loss 的平滑近似函数。该优化问题可以利用梯度下降法，通过迭代优化参数 **A**、**B**、**C**、以及 **w** 寻得最优解。

3.4 实验结果及分析

3.4.1 数据集和实验设置

（1）特征提取

我们在实验中，借助滑动窗提取基于局部区域的手工特征向量。具体来说，首先将图片缩放到 128×48，然后将大小为 16×16 的滑动窗口以步长 8 从左到右从上到下滑动，每一次滑动都可以在局部区域内提取出相应的特征向量。每个局部特征包含提取于 HSV 和 YUV 颜色空间的 3×8 维的颜色直方图、3 维的颜色矩、以及 10 维的旋转不变均匀LBP 特征；而全局特征通过将每个局部特征串联后，降维到 250 维得到。

（2）数据集

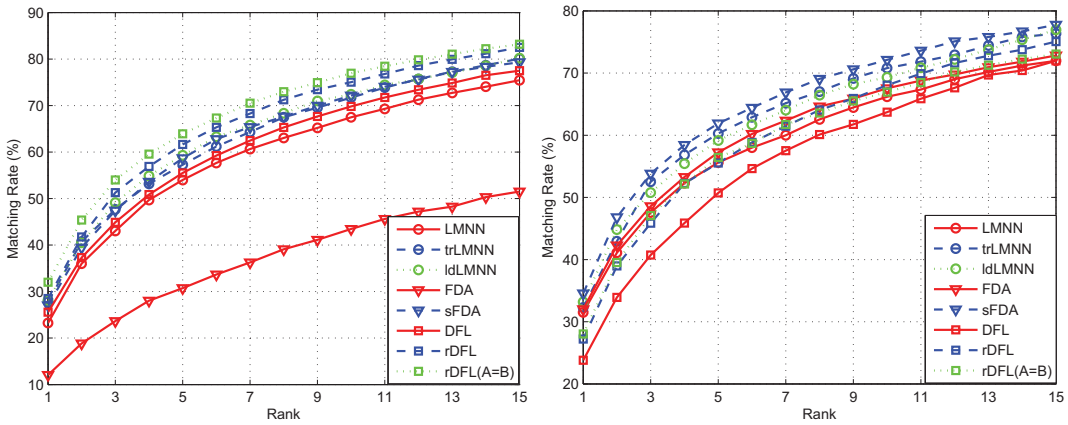
我们选用了两个基准数据集，分别是 VIPeR 和 CUHK01。关于数据集的详解介绍，可以参考小节1.2.3.2。

（3）性能评价指标

每个数据集都被随机平均分成不重叠的两部分，一部分用来模型训练，一部分用来测试。每个数据集上，分组实验进行了 10 次，以CMC 曲线的平均结果作为最终实验结果。

（4）参数设置

基于经验，对于trLMNN 和ldLMNN 算法来说，我们将正则化参数 λ 设置为 10^4；对于sFDA 和ldLMNN 算法来说，我们将参数 α 设置为 0.5；而对于rDFL 算法来说，我们分别将参数设置为 $\lambda_1 = 0.6$，$\lambda_2 = 0.6$，$\lambda_3 = 0.9$。



(a) VIPeR (b) CUHK01

图 3-2 基于正则化度量学习的方法与基准模型的性能对比。

3.4.2 与基准模型的性能对比

为了证明基于正则化度量学习方法的有效性，我们以不包含正则化的原始度量学习方法作为基准模型，并将他们在两个ReID 基准数据集上的性能对比结果展示在图3-2中。通过对比发现，在度量模型基础上，引入正则化约束，可以进一步提升行人再识别的准确度。比如，在数据集 VIPeR 上，trLMNN 和ldLMNN 将LMNN 的 Rank-1 准确度分别提升了 4.6% 和 5.0%，sFDA 将FDA 的 Rank-1 准确度提升了 14.9%，rDFL 将DFL 的 Rank-1 准确度提升了 2.9%。

3.4.3 与同期其他先进 ReID 模型的性能对比

我们将基于正则化的度量学习ReID 算法，与 kMFA[33]、LFDA[28]、以及 LADF[27] 这三种度量学习算法进行了性能对比。由于实验中提取的行人特征表示与论文[33] 一致，因此我们直接从论文[33] 中读取以上三种算法的CMC 曲线。除此之外，我们也对比了 SCNCD[16]、SDALF[10]、以及 SDC[13] 这三种常用的基于特征表示的ReID 算法。所有对比结果，被展示在表格3-1中。对比发现，基于正则化的度量学习的ReID 算法可以到达同期的先进水平。同时，对于算法rDFL 来说，当我们进一步约束模型复杂度，令参数 $\mathbf{A} = \mathbf{B}$ 时，算法在 VIPeR 上的再识别准确率几乎超过了所有对比方法。

表 3-1 在数据集 VIPeR 上，基于正则化度量学习方法与同期其他先进 ReID 模型的性能对比。

方法	Rank1	Rank5	Rank10	Rank20
SCNCD[16]	**33.7%**	62.7%	74.8%	85%
SDALF[10]	19.9%	38.9%	49.4%	65.7%
SDC[13]	26.7%	50.7%	62.4%	76.4%
kMFA[33]	31.1%	**65.2%**	**79.6%**	**90.2%**
LFDA[33]	21.5%	49.6%	64.6%	79.1%
LADF[33]	30.1%	63.2%	**77.4%**	88.1%
LMNN	23.2%	54.0%	67.5%	80.8%
trLMNN	27.8%	57.3%	71.7%	84.0%
ldLMNN	28.2%	59.3%	72.4%	84.6%
FDA	12.1%	30.7%	43.5%	57.7%
sFDA	27.0%	58.7%	72.2%	83.8%
DFL	25.6%	55.5%	69.8%	81.9%
rDFL	28.5%	61.7%	75.0%	87.0%
rDFL(A=B)	**32.0%**	**63.9%**	77.0%	**88.6%**

3.5 本章小结

本章分别从引言、相关工作、不同形式度量函数、正则化度量学习算法、以及相关实验结果和分析等方面，对我们提出的基于正则化度量学习的行人再识别算法进行详细介绍。我们提出的基于正则化度量学习的行人再识别算法，可以进一步提升度量学习模型在小数据集上的泛化能力。具体来说，我们从三种度量函数形式上探究正则化对度量学习的影响，并提出了四中不同的正则化度量学习算法；而且，算法的有效性在两个中小规模基准数据集 VIPeR 和 CUHK01 上得到了验证。

第四章 基于空间金字塔统计特征及多核学习的行人再识别算法

本章分别从引言、相关工作、基于空间金字塔的统计特征提取框架、基于多核局部费舍尔判别分析的特征融合、以及相关实验结果和分析等方面，对我们提出的基于空间金字塔统计特征及多核学习的行人再识别算法进行详细介绍。

4.1 引言

为了解决ReID 任务，大部分研究人员会将其构造成为一个图像匹配或者图像确认问题进行处理。具体来讲，将采集于某一摄像头下的行人图像当做 Probe，与采集于另一个跨视域摄像头下的所有行人图像一一比对，并返回匹配排序列表，如图4-1所示。然而，不幸的是，严重的拍摄角度变化、行人姿态变化、光照变化、以及背景遮挡等，使高精度地行人再识别变得困难重重。大量的学者尝试通过设计跨视域不变的特征表示、或者通过学习合适的距离度量方法，来提升ReID 的准确度。

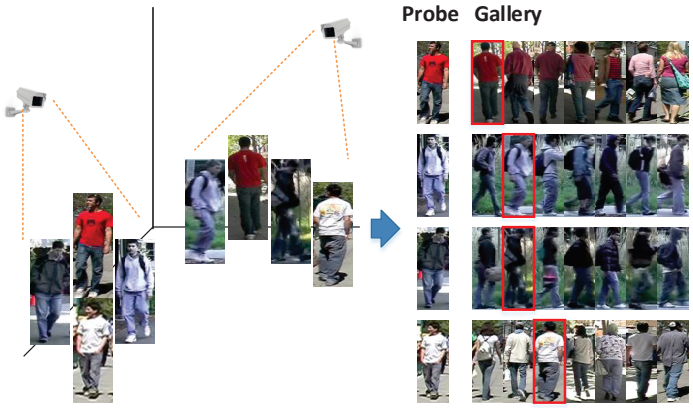


图 4-1 行人再识别任务示意图。

经过多年的研究，研究人员发现局部统计特征（如颜色直方图、方向梯度直方图[108] 等）是一类非常有效的可用于行人再识别的特征表示。例如，论文[13,30] 中通过将局部统计特征聚合成全局特征向量来刻画行人外观信息，论文[29,32] 中通过学习跨视域局部统计特征间的映射关系来关联不同摄像头下的行人。另外，也有研究人员开始尝试通过融合多种类型的局部统计特征（如颜色、纹理、空间结构信息等）来更全面地刻画行人外观。例如，论文[24,28] 中将不同类型的局部统计特征串联起来作

为行人的最终外观表示，论文[10,15] 通过融合不同局部统计特征的距离度量来获取最终的行人匹配结果。然而，尽管这些工作带来了ReID 性能的稳步提升，但是目前仍然缺乏对局部统计特征提取过程中相关细节的全面而详细地评估，并且也缺少对不同多特征融合策略的分析和讨论。

在本章中，我们提出了一种统一的行人再识别算法框架，主要包含常用局部统计特征提取和多特征融合两部分，如图4-2所示。具体来讲，在这个框架下，我们提取了五种不同的基于空间金字塔的统计特征，包括：基于空间金字塔的颜色直方图（spatial pyramid based Color Histogram，spHist）、基于空间金字塔的方向梯度直方图（spatial pyramid based Histogram of Oriented Gradient，spHOG）、基于空间金字塔的局部二值模式（spatial pyramid based Local Binary Pattern，spLBP）、基于空间金字塔的颜色命名特征（spatial pyramid based Color Names，spCN）、和基于空间金字塔的协方差特征（spatial pyramid based Covariance Feature，spCov）；同时，我们利用多核局部费舍尔判别分析算法（multiple kernel Local Fisher Discriminant Analysis，mkLFDA）对五种特征进行了特征融合。更重要的是，我们在四个基准数据集上，对特征提取过程中的关键步骤以及多种特征融合策略，进行了全面而详细地实验评估。

本章的主要贡献如下：

- 提出了一种统一的ReID 算法框架，包括基于空间金字塔的统计特征提取和基于mkLFDA 的多特征融合；
- 在 VIPeR 数据集上进行了大量的对比实验，来评估特征提取每个环节的作用以及不同的特征融合策略的有效性；
- 在 VIPeR、CUHK01、PRID2011、3DPeS 四个基准数据集上，通过实验结果验证了所提ReID 算法的优势。

本章的其余部分组织如下：小节4.2对相关工作进行回顾和总结，小节4.3对基于空间金字塔统计特征及多核学习的行人再识别算法进行详细介绍，小节4.4对相关实验结果进行了详细展示和分析，小节4.5对本章进行了简单总结。

4.2 相关工作

在ReID 任务中，复杂的摄像机配置、摄影环境、行人姿态等都会引起严重的行人外观变化，进而导致严重的类内混淆和类间偏移问题，从而使高精度ReID 难以实现。因此，有大量研究人员投入精力来缓解这些困难。这些工作大致可以分为两类：其一，构造鲁棒的、辨识度高的特征表示，比如[8,10,11,13–16,18–20,30,36,43,66,67,132]；其二，

学习具有区分能力的距离度量，比如[18,23–29,31,33,34,37,42,102,133]。

尽管所有的研究工作都有自己的贡献，但是多数工作都是基于局部统计特征实现的。例如，Zhao 等人[13,30] 提出按照显著性为每个区域内的局部统计特征赋予不同的权重；而 Su[20]、Shi[19] 等人提出从底层局部统计特征中学习更鲁棒的高层属性特征。在度量学习算法中，研究人员通常以行人外观的局部统计特征表示为基础，学习具有辨别能力的距离度量矩阵或者投影矩阵。例如，Zheng 等人[23,31] 以行人的局部统计特征为输入，将ReID 构造成一个概率相对距离比较问题；而 Xiong 等人[33] 同样以行人的局部统计特征为输入，将核技巧引入到度量学习中，从而进一步提升了ReID 的性能。

另外，也有许多研究人员从不同的角度去解决ReID 问题。例如，[32,38,39,103,126,127,134] 中，字典学习和稀疏表示的思想被用来解决跨视域的行人特征匹配以及鲁棒特征学习的问题；在论文[40] 中，基于图像块的局部对应关系学习算法被用来缓解图像匹配中的空间不对齐问题；在论文[21,102] 中，跨视域的特征映射信息被用来辅助行人的再识别；近年来，深度学习也被引入到ReID 任务中，以进一步提升算法准确度，如[66,67]。除此之外，还有许多其他有意思的行人再识别扩展工作被提出，比如基于不完整图片的ReID[126]、基于步态信息的ReID[135]、超分辨率ReID[103,104]、跨域ReID[136,137]、基于摄像机网络的ReID[138]、以及大尺度ReID[3,139]、等等。

尽管这些工作都使ReID 的性能得到了稳步提升，但是我们仍认为，由于缺少对特征提取流程中各环节全面而详细地分析和评估，这些工作并没有将局部统计特征的性能潜力完全发掘出来。同时，我们认为如果能更好的利用这些局部特征表示，就可以进一步提升ReID系统的性能。

在计算机视觉领域近些年的研究里，基于空间金字塔匹配 (Spatial Pyramid Matching，SPM)[140,141] 和积分通道特征 (Integral Channel Features，ChnFtrs)[142] 的特征提取框架已经分别在图像分类和行人检测任务上取得了巨大的成功。然而，由于与图像分类任务相比，行人再识别属于细粒度的实例分类任务，因而需要比SPM 框架所提特征辨识度更高的特征表示；另外，与物体检测相比，行人再识别也需要更加关注具有良好稳定性的高层语义信息，而不单单只是ChnFtrs 框架所提取的原始图像通道信息。显然，这两个特征提取框架都无法完全满足ReID 对特征提取的要求，因此我们需要为ReID 构造专用的特征提取框架。我们在本章提出了一种融合了SPM 和ChnFtrs 的统一特征提取框架，可以针对ReID 任务提取多种基于空间金字塔的统计特征，并且可以进一步利用mkLFDA 对多种特征进行融合。

4.3 我们的算法

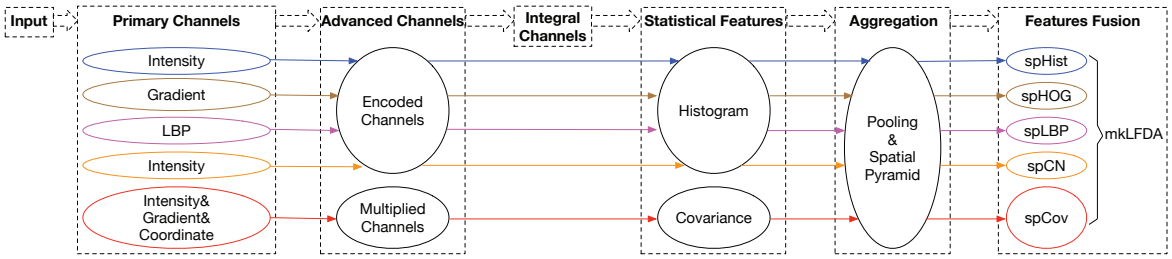


图 4-2　基于空间金字塔统计特征及多核学习的行人再识别算法流程图。

在本小节，我们将对基于空间金字塔统计特征提取和融合的统一框架进行详细介绍，框架流程图如图4-2所示。完整的框架流程由生成初级特征通道（Generating Primary Channels）、构造高级特征通道（Constructing Advanced Channels）、提取局部区域统计特征（Extracting Local Region Statistical Features）、局部特征聚合（Features Aggregation,）、以及多特征融合（Features Combination）五个步骤组成。

4.3.1 基于空间金字塔的统计特征提取框架

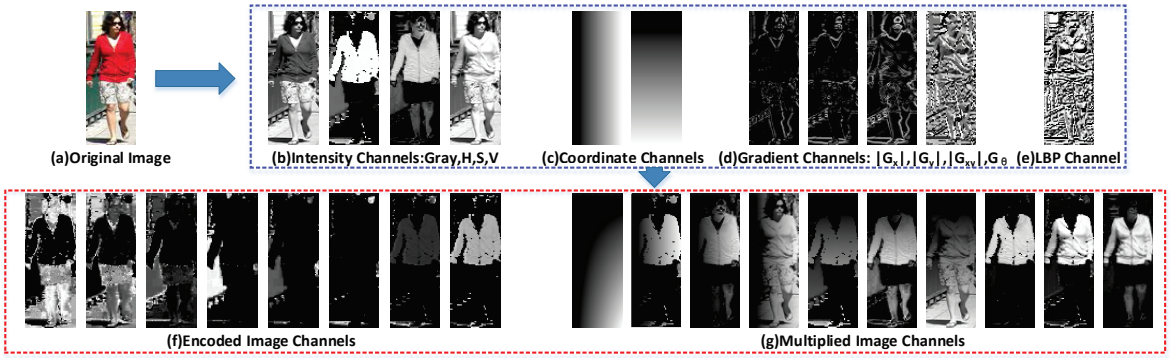


图 4-3　特征通道举例。(a)：原始图像；(b) - (e)：初级特征通道；(f) 和 (g)：高级特征通道。

4.3.1.1 生成初级特征通道

特征通道也可以被称作特征图谱，图谱每个位置的像素值由原始图像相应区域内的像素值映射而来。初级特征通道，往往是将单一通道的原始图像经过某种简单的单一变换而来。具体来讲，假设单一通道的原始图像可表示为 $\mathbf{I} \in \mathbb{R}^{H \times W}$，其中 H 和 W 分别为图像的高度和宽度；那么，初级特征通道可以表示为 $\mathbf{C} = \Omega(\mathbf{I}) \in \mathbb{R}^{H \times W}$，其中 Ω 表示某种简单的变换操作，如论文[142] 所提到的线性滤波、恒等变换、非线性变

换、或者逐像素映射操作等。这些初级特征通道，往往包含了丰富的视觉细节信息。在我们的方法中，主要使用了亮度通道（Intensity Channels）、坐标通道（Coordinate Channels）、梯度通道（Gradient Channels）、LBP 通道（LBP Channels）[107] 四种初级通道信息。下面，我们将对这四种初级通道一一介绍。

（1）亮度通道

等价变换是生成初级特征通道的最简单方式。由等价变换得到的初级特征通道，保留了原始图像中的所有细节信息。图4-3(b) 展示了四种亮度通道，分别是灰度（Gray）通道、色相（Hue）通道、饱和度（Saturation）通道、和明度（Value）通道。

（2）坐标通道

图像中的像素除了包含亮度信息以外，同时还包含了位置信息。因此利用像素的位置坐标，也可以生成相应的初级特征通道，即坐标通道。尽管单独使用坐标通道信息时，并不能区分不同行人图像；但是当将坐标通道与其他通道信息联合使用时，可以大幅度提升辨别行人图像的能力。图4-3(c) 分别展示了 x-坐标通道以及 y-坐标通道。

（3）梯度通道

图像的梯度信息代表了局部区域内图像亮度的变化方向，可以用来很好地刻画物体的形状以及纹理等特征。在我们的方法中，我们利用简单的梯度滤波器 $\mathbf{k}=[-1,0,1]^T$ 提取了图像的四种梯度方向信息，分别是：

$$\begin{aligned}
&\text{水平方向梯度幅值：}\ \mathbf{C}_{|G_x|}=|\mathbf{I}\otimes\mathbf{k}^T|;\\
&\text{垂直方向梯度幅值：}\ \mathbf{C}_{|G_y|}=|\mathbf{I}\otimes\mathbf{k}|;\\
&\text{梯度幅值：}\ \mathbf{C}_{|G_{xy}|}=\sqrt{(\mathbf{I}\otimes\mathbf{k})^2+(\mathbf{I}\otimes\mathbf{k}^T)^2};\\
&\text{梯度方向：}\ \mathbf{C}_{G_\theta}=\text{atan2}(\mathbf{I}\otimes\mathbf{k},\mathbf{I}\otimes\mathbf{k}^T)+\pi);
\end{aligned}\tag{4-1}$$

其中 $\otimes$ 表示离散卷积操作，atan2$(\cdot,\cdot)$ 代表四象限反正切函数。图4-3(d) 分别展示了不同的梯度特征通道。

（4）LBP 通道

局部二值模式是一个由中心像素与领域像素亮度差值决定的二值序列，通常对于图像的单调变换具有一定不变性。在我们的工作中，我们将量化的 LBP 图像当做一种初级特征通道，该特征通道是利用 3×3 的掩模滑动窗口得到的。图4-3(e) 展示了 LBP 特征通道。

4.3.1.2 构造高级特征通道

高级特征通道通常是通过对多个初级特征通道进一步处理得来，在我们的工作中，高级特征通道往往是为后续快速计算局部统计特征做准备。具体来说，给定一系列初级特征通道 $\mathscr{C}=\{\mathbf{C}^{(m)}\in\mathbb{R}^{H\times W},m=1,...,M\}$，则高级特征通道可表示为 $\mathscr{C}_{Adv}=\mathbf{\Omega}(\mathscr{C})\in\mathbb{R}^{H\times W\times M'}$，其中 M' 为高级特征通道的数目，$\mathbf{\Omega}$ 代表某种映射操作，比如编码或者乘积操作。为了更直观的认识高级特征通道，我们在图4-3(f) 和 (g) 中展示了一些样例。

（1）编码特征通道（Encoded Image Channels）

给定一幅图像的 M 个初级特征通道组成的特征图谱以及一个定义好的码书（Codebook）$\mathbf{V}=\{\mathbf{v}_n\in\mathbb{R}^M,n=1,...,N\}$，图谱中每个位置的 M 维特征向量 $\mathbf{f}_{ij}$ 都可以被编码成关于码书 $\mathbf{V}$ 中元素的 N 个组合系数 $\{a_{ij}^n,n=1,...,N\}$。那么，第 n 个编码特征通道的每个像素可以表示为 $\mathbf{C}^{(n)}(i,j)=a_{ij}^n$，此时一共可以构造出 $M'=N$ 个不同的高级特征通道。基于编码特征通道，可以快速地提取类直方图统计特征。

根据不同的编码策略，我们将用到的编码方法分为硬编码（比如直方图编码）和软编码（比如核码书编码、线性插值编码、以及显著颜色编码）。

直方图编码（Histogram Encoding，HE）。在HE 方法中，每一个特征向量 $\mathbf{f}_{ij}$ 按照如下方式编码成 N 维组合系数向量：

$$a_{ij}^n=1\quad\text{if}\quad n=\arg\min_{n'\in\{1,\cdots,N\}}\|\mathbf{f}_{ij}-\mathbf{v}_{n'}\|_2^2, \tag{4-2}$$

否则 $a_{ij}^n=0$。在这种编码方式下，只有距离 $\mathbf{f}_{ij}$ 最近的码字对应的系数被置为 1，其余全为 0。

核码书编码（Kernel Codebook Encoding，KCE）。KCE 方法的基本原理是核密度估计算法，每一个特征向量 $\mathbf{f}_{ij}$ 按照如下方式编码：

$$a_{ij}^n=\frac{k(\mathbf{f}_{ij},\mathbf{v}_n)}{\sum_{l=1}^N k(\mathbf{f}_{ij},\mathbf{v}_l)}, \tag{4-3}$$

其中核函数为 $k(\mathbf{f},\mathbf{v})=\exp(-\frac{\gamma}{2}||\mathbf{f}-\mathbf{v}||_2^2)$。

线性插值编码（Linear Interpolation Encoding，LIE）。当对每个初级特征通道单独进行编码操作时（即 $M=1$），特征 f_{ij} 也可以按照如下方式编码：

$$a_{ij}^n=\max(0,1-|f_{ij}-v_n|/b), \tag{4-4}$$

其中 b 是等分区间的宽度。线性插值策略还可以很容易地扩展到二维或者三维特征

空间，即论文[108] 中涉及到的双线性插值或三线性插值。

显著颜色编码（Salient Color Encoding，SCE）。当对三个颜色通道同时进行编码时，可以采用论文[16] 中提出的SCE 方法，即将颜色空间量化为离散的颜色名称。具体来讲，首先将整个颜色空间（表示为 $\mathbb{F}$）均匀地划分为 $32\times32\times32$ 个小的立方体 $\{\mathbb{F}_c, c=1,...,32768\}$，其中每个立方体都涵盖了 512 个颜色值，即 $\mathbb{F}_c=\{\mathbf{f}^{(l)}, l=1,..,512\}$；然后，计算将任一颜色值 $\mathbf{f}_{ij}\in\mathbb{F}_c$ 命名为颜色名称 $\mathbf{v}_n$ 的概率为：

$$a_{ij}^n = p(\mathbf{v}_n|\mathbb{F}_c),\ \text{for}\ \mathbf{f}_{ij}\in\mathbb{F}_c, \tag{4-5}$$

其中

$$p(\mathbf{v}_n|\mathbb{F}_c)=\sum_{l=1}^{512} p(\mathbf{v}_n|\mathbf{f}^{(l)})p(\mathbf{f}^{(l)}|\mathbb{F}_c), \tag{4-6}$$

其中，如果 $\mathbf{v}_n\in\text{KNN}(\mathbf{f}^{(l)})$，那么

$$p(\mathbf{v}_n|\mathbf{f}^{(l)})=\frac{\exp\left(\frac{-||\mathbf{v}_n-\mathbf{f}^{(l)}||_2^2}{\frac{1}{k-1}\sum_{\mathbf{v}_p\neq\mathbf{v}_n}||\mathbf{v}_p-\mathbf{f}^{(l)}||_2^2}\right)}{\sum_{q=1}^{k}\exp\left(\frac{-||\mathbf{v}_q-\mathbf{f}^{(l)}||_2^2}{\frac{1}{k-1}\sum_{\mathbf{v}_s\neq\mathbf{v}_q}||\mathbf{v}_s-\mathbf{f}^{(l)}||_2^2}\right)}, \tag{4-7}$$

否则 $p(\mathbf{v}_n|\mathbf{f}^{(l)})=0$，并且

$$p(\mathbf{f}^{(l)}|\mathbb{F}_c)=\frac{\exp\left(-\alpha||\mathbf{f}^{(l)}-\boldsymbol{\mu}_c||_2^2\right)}{\sum_{t=1}^{512}\exp\left(-\alpha||\mathbf{f}^{(t)}-\boldsymbol{\mu}_c||_2^2\right)}, \tag{4-8}$$

其中 k 为近邻的数目，$\boldsymbol{\mu}_c$ 为 $\mathbb{F}_c$ 的平均颜色值。

（2）乘积特征通道（Multiplied Image Channels）

乘积征通道通过对两个初级特征通道做逐像素的乘积得到。具体来说，给定一系列初级特征通道 $\mathscr{C}\in\mathbb{R}^{H\times W\times M}$，乘积通道可以按照如下方式构造：

$$\mathbf{C}^{m_1,m_2}=\mathbf{C}^{(m_1)}\odot\mathbf{C}^{(m_2)}, \tag{4-9}$$

其中 $\mathbf{C}^{(m_1)},\mathbf{C}^{(m_2)}\in\mathscr{C}$，$\odot$ 表示逐元素乘积，此时一共可以构造出 $M'=M\times M$ 个不同的高级特征通道。与编码特征通道类似，基于乘积特征通道，可以快速提取局部区域的协方差特征。

4.3.1.3 提取局部区域统计特征

得益于初级和高级特征通道，图像中任意矩形感兴趣区域 (Region of Interest, ROI) 的统计特征（如类直方图特征、均值向量、协方差矩阵）都可以被快速地提取出来。具体来说，给定一系列特征通道，先分别计算每个通道同一位置的 ROI 区域内像素的和，再将这些统计结果组合起来就可以构造出不同的局部统计特征。在我们的工作中，我们以此构造了四种类直方图特征（即spHist、spHOG、spLBP、和spCN）和一种协方差特征（即spCov）。在我们的特征提取框架下，ROI 被定义成对图像进行稠密网格分割得到的小胞（Cell），如图4-4所示。具体来说，当输入图像被缩放到 128×48 大小时，用于提取spHist 和spCN 特征的小胞大小为 4×4，用于提取其他特征的小胞大小为 8×8。

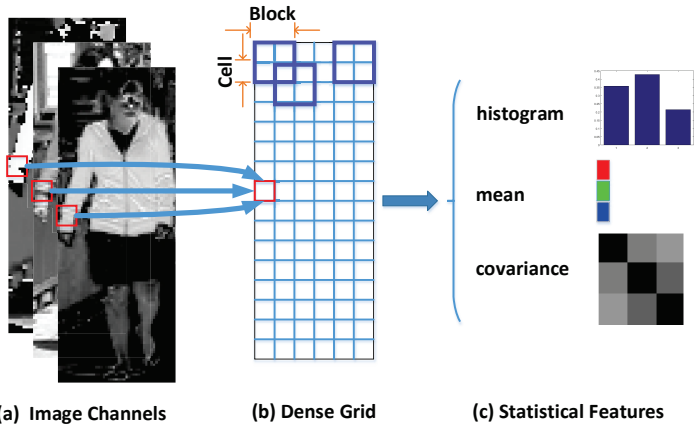


图 4-4 局部统计特征提取过程示意图。通过计算不同特征通道局部区域的像素值之和，可以得到三种不同类型的统计特征，分别是类直方图特征、均值向量、和协方差矩阵。

为了加快局部统计特征的计算过程，我们将可以快速统计局部区域像素值之和的积分图通道作为中间通道引入到我们的特征提取框架中。积分图通道上每个位置的像素值，可以通过对输入特征通道内以该位置为右下角顶点的矩阵区域内部所有像素值的快速求和得到。给定一个输入通道 $\mathbf{C}$，它所对应的积分图通道可以表示为：

$$\mathbf{C}_{Intg}(i', j') = \sum_{i \le i', j \le j'} \mathbf{C}(i, j). \tag{4-10}$$

有了积分图通道的帮助，局部统计特的提取速度得到进一步提升。令 $\mathscr{C}_{Intg} \in \mathbb{R}^{H \times W \times M}$，$\mathscr{C}_{Intg'} \in \mathbb{R}^{H \times W \times N}$，$\mathscr{C}_{Intg''} \in \mathbb{R}^{H \times W \times M \times M}$，分别是初级特征通道、编码特征通道、乘积特征通道对应的积分图通道集合。为了简单起见，我们再次令 $\mathbf{p}_{i,j} = \mathscr{C}_{Intg}(i, j, :)$，$\mathbf{q}_{i,j} = \mathscr{C}_{Intg'}(i, j, :)$，和 $\mathbf{E}_{i,j} = \mathscr{C}_{Intg''}(i, j, :, :)$。那么，基于初级特征通道，某个小胞

$\{(i',j'),(i'',j'')\}$ 内的均值向量可以按照以下方式计算得到：

$$\mathbf{u}_{(i',j';i'',j'')} = \frac{(\mathbf{p}_{i'',j''} + \mathbf{p}_{i'-1,j'-1} - \mathbf{p}_{i'-1,j''} - \mathbf{p}_{i'',j'-1})}{S}; \tag{4-11}$$

基于编码特征通道，小胞内的类直方图统计特征可以按照以下方式计算得到：

$$\mathbf{h}_{(i',j';i'',j'')} = (\mathbf{q}_{i'',j''} + \mathbf{q}_{i'-1,j'-1} - \mathbf{q}_{i'-1,j''} - \mathbf{q}_{i'',j'-1}); \tag{4-12}$$

基于初级特征通道以及乘积特征通道，小胞内的协方差矩阵[143] 可以按照以下方式计算得到：

$$\mathbf{O}_{(i',j';i'',j'')} = \frac{1}{S-1}[\mathbf{E}_{i'',j''} + \mathbf{E}_{i'-1,j'-1} - \mathbf{E}_{i'-1,j''} - \mathbf{E}_{i'',j'-1} - S\mathbf{u}_{(i',j';i'',j'')}\mathbf{u}_{(i',j';i'',j'')}{}^{T}]; \tag{4-13}$$

其中 $S = (i''-i'+1)(j''-j'+1)$。

然而，需要注意的是，直接以公式4-10得到的原始积分图通道（Original Integral Channels）为基础提取局部统计特征时，会产生由小胞边界上的像素引起的空间混叠问题。为了克服这个缺陷，我们借鉴了论文[144] 中的空间三线性插值策略，即先对原始特征通道用预定义的卷积核进行预卷积处理。在我们的方法中，我们分别将用于计算 4×4 和 8×8 大小的小胞局部特征的卷积核 $\mathbf{K}$ 定义为：

$$\frac{1}{16}\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix} \text{ or } \frac{1}{256}\begin{bmatrix}1&2&3&4&3&2&1\\2&4&6&8&6&4&2\\3&6&9&12&9&6&3\\4&8&12&16&12&8&4\\3&6&9&12&9&6&3\\2&4&6&8&6&4&2\\1&2&3&4&3&2&1\end{bmatrix},$$

卷积核中的权值按照每个像素位置与其领域像素位置的距离分布。在此基础上，卷积积分图通道（Convoluted Integral Channels）可以按照以下方式计算得到：

$$\mathbf{C}_{Intg}(i',j') = \sum_{i\le i',j\le j'} \mathbf{C}_{Cov}(i,j), \text{ where } \mathbf{C}_{Cov} = \mathbf{C}\otimes\mathbf{K}. \tag{4-14}$$

备注 1: 为了缓解光照变化以及背景干扰对特征性能的影响，我们将基于小胞的局部统计特征做了局部对比归一化处理。与论文[108] 类似，我们将 2×2 个小胞组成有一半重叠的区块（Block），然后对每个区块进行逐一的归一化处理，如图4-4(b) 所

示。对于spHOG 特征来说，特征组合方式是将区块内基于小胞的局部统计特征串联起来；对于其他特征来说，特征组合方式是求区块内基于小胞的局部统计特征的平均值。另外，我们分别测试了五种不同的归一化策略，分别是：

$$\begin{aligned} l_1\text{-norm}: &\quad \mathbf{x} \rightarrow \mathbf{x}/||\mathbf{x}||_1 + \varepsilon; \\ l_1\text{-sqrt}: &\quad \mathbf{x} \rightarrow \sqrt{\mathbf{x}/||\mathbf{x}||_1 + \varepsilon}; \\ l_2\text{-norm}: &\quad \mathbf{x} \rightarrow \mathbf{x}/||\mathbf{x}||_2 + \varepsilon; \\ l_2\text{-clip}: &\quad \text{在}l_2\text{-norm基础上，限制最大值}; \\ l_1^2\text{-norm}: &\quad \mathbf{x} \rightarrow \mathbf{x}/||\mathbf{x}||_1^2 + \varepsilon; \end{aligned} \tag{4-15}$$

其中，$\varepsilon > 0$ 为一个很小的常数。对于协方差特征，我们先按照 $\mathbf{O} \rightarrow \text{diag}(\mathbf{O})^{-\frac{1}{2}}\mathbf{O}\text{diag}(\mathbf{O})^{-\frac{1}{2}}$ 对基于小块的协方差矩阵 $\mathbf{O}$ 进行归一化；考虑到协方差矩阵的对称性，我们接着将归一化后的协方差矩阵的上三角矩阵展开成一个协方差特征向量；最后，协方差特征就可以按照向量的方式进行归一化处理。

4.3.1.4 局部特征聚合

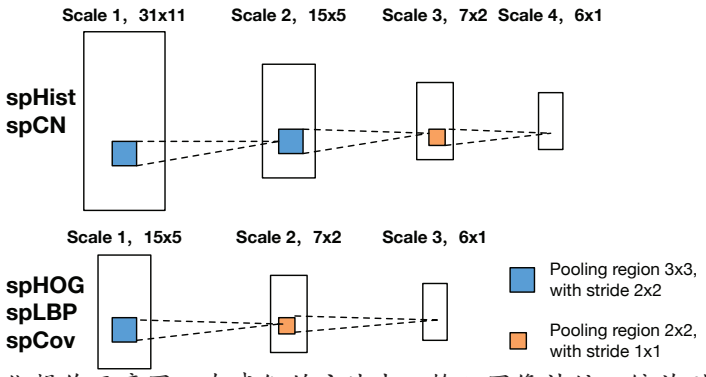


图 4-5 多尺度池化操作示意图。在我们的方法中，输入图像被统一缩放到 128×48 大小，因此局部归一化以后每幅图像会产生 31×11 个基于区块的 spHist 和 spCN 特征，产生 15×5 个基于区块的 spHOG、spLBP、和 spCov 特征。

（1）池化（Pooling）

在我们的特征提取框架中，我们利用多尺度池化操作实现空间金字塔特征的提取，如图4-5所示。具体来说，池化操作以后，每幅图像可以产生 $(31 \times 11 + 15 \times 5 + 7 \times 2 + 6 \times 1 = 436)$ 个基于区块的spHist 和spCN 特征，产生 $(15 \times 5 + 7 \times 2 + 6 \times 1 = 95)$ 个基于区块的spHOG、spLBP、和spCov 特征。另外，我们分别对均值池化（Average Pooling）和最大池化（Max Pooling）两种池化策略进行了测试评估。

（2）空间金字塔

SPM 框架中，图像的最终特征表示通过将多尺度特征以空间金字塔的形式组合而成。然而，为了计算的高效和方便，我们直接将多尺度的特征串联起来组成图像最终特征表示。另外，我们分别在单一尺度特征和多尺度特征上进行了归一化操作。

为了评估空间金字塔特征的性能，我们分别在空间尺度和归一化方法上做了对比实验。具体来说，我们比较了多尺度特征和单一尺度特征，同时我们也比较了两种不同的归一化策略（即 l_1-norm 和 l_2-norm）。

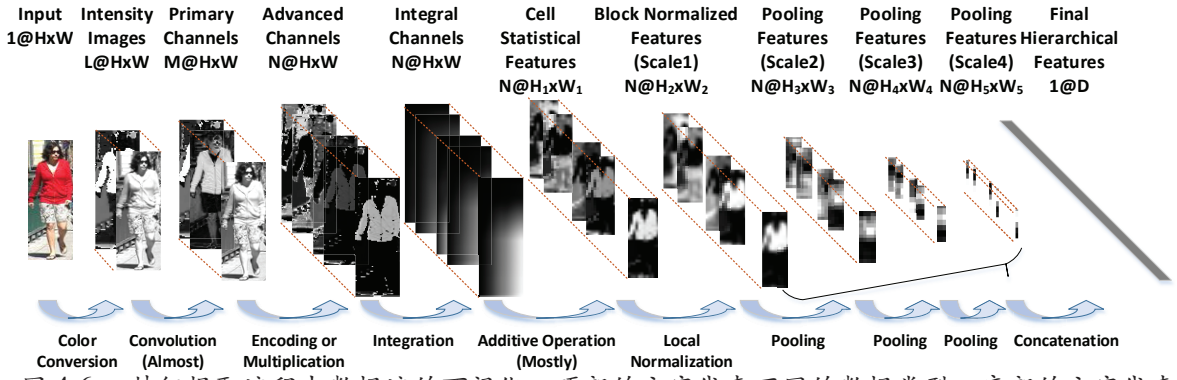


图 4-6 特征提取流程中数据流的可视化。顶部的文字代表不同的数据类型，底部的文字代表对应的操作。

为了更清晰直观地理解基于空间金字塔的统计特征提取流程，我们将特征提取过程中可视化的数据流展示在了图4-6中。同时，在特征提取框架下，提取基于空间金字塔统计特征（即spHist、spHOG、spLBP、spCN、和spCov）时的默认框架参数配置，以及提取对应的原始统计特征（Hist、HOG[108]、LBP[145]、显著颜色命名特征(Salient Color Names，SCN)[16]、和协方差矩阵 (Covariance Martix，Cov)[143]）时的框架参数配置，都被详细地列在了表4-1中。

4.3.2 基于多核局部费舍尔判别分析的特征融合

4.3.2.1 核化局部费舍尔判别分析

论文[146] 提出核化局部费舍尔判别分析 (kernel Local Fisher Discriminant Analysis，kLFDA) 算法，目的是为了在增强样本类间可分性的同时，减小样本类内的差异，同时保持数据的局部邻域结构。它的优化目标函数为：

$$\max_{\mathbf{A}} \frac{\mathrm{tr}(\mathbf{A}^T \widetilde{\mathbf{S}}^{(b)} \mathbf{A})}{\mathrm{tr}(\mathbf{A}^T \widetilde{\mathbf{S}}^{(w)} \mathbf{A})}, \tag{4-16}$$

表 4-1　提取不同特征时，框架的参数配置。

特征	参数配置					
	Input	$\mathscr{C}_{\mathscr{A}}$	$\mathscr{C}_{Intg}$	Contrast Normalize	Pooling	Spatial Scale & Normalize
spHist	HSV	KCE, $\gamma=1/12$	Conv	ℓ_2-norm	Max	Multi, ℓ_1
spHOG	YUV	KCE, $\gamma=\pi/6$	Conv	ℓ_1^2-norm	Avg	Multi, ℓ_2
spLBP	YUV	HE	Conv	ℓ_2-norm	Max	Multi, ℓ_2
spCN	RGB	SCE	Conv	ℓ_2-norm	Max	Multi, ℓ_2
spCov	Gray,HSV, YUV,LAB	MULT	Orig	ℓ_2-norm	Avg	Multi, ℓ_2, +Mean
Hist	HSV	HE	Orig	ℓ_2-norm	–	Single, ℓ_1
HOG	Gray	LIE	Conv	ℓ_2-norm	–	Single, ℓ_2
LBP	Gray	HE	Orig	ℓ_1-sqrt	–	Single, ℓ_1
SCN	RGB	SCE	Orig	ℓ_2-norm	–	Single, ℓ_2
Cov	Gray	MULT	Orig	ℓ_2-norm	–	Single, ℓ_2

其中 tr(·) 代表矩阵的迹，$\mathbf{A}\in\mathbb{R}^{s\times d}$ 代表投影矩阵，s 表示数据集的大小。符号 $\widetilde{\mathbf{S}}^{(b)}$ 和 $\widetilde{\mathbf{S}}^{(w)}$ 分别表示类间和类内的局部散度矩阵，表示为：

$$\begin{aligned}\widetilde{\mathbf{S}}^{(b)}&=\frac{1}{2}\sum_{i,j=1}^{s}\widetilde{\mathbf{W}}_{ij}^{(b)}(\mathbf{k}_i-\mathbf{k}_j)(\mathbf{k}_i-\mathbf{k}_j)^T\in\mathbb{R}^{s\times s},\\ \widetilde{\mathbf{S}}^{(w)}&=\frac{1}{2}\sum_{i,j=1}^{s}\widetilde{\mathbf{W}}_{ij}^{(w)}(\mathbf{k}_i-\mathbf{k}_j)(\mathbf{k}_i-\mathbf{k}_j)^T\in\mathbb{R}^{s\times s},\end{aligned}\tag{4-17}$$

其中 $\mathbf{k}_i=[\kappa(\mathbf{x}_1,\mathbf{x}_i),...,\kappa(\mathbf{x}_s,\mathbf{x}_i)]^T\in\mathbb{R}^s$，$\widetilde{\mathbf{W}}^{(b)}$ 和 $\widetilde{\mathbf{W}}^{(w)}$ 分别表示样本类间和类内的局部邻接图的权重矩阵。

以上优化问题公式4-16可以通过转化为等价的广义特征值问题 $\widetilde{\mathbf{S}}^{(b)}\mathbf{A}=\lambda\widetilde{\mathbf{S}}^{(w)}\mathbf{A}$ 进行求解，其最优解 $\mathbf{A}_*\in\mathbb{R}^{s\times d'}$ 由对应于前 d' 个最大特征值的特征向量组成。

4.3.2.2　多核局部费舍尔判别分析

在利用核技巧提升模型表达能力时，如何选择适合当前问题以及数据分布的核函数是问题的关键。与单纯依赖研究人员经验手工设计核函数不同，多核学习(Multiple Kernel Learning，MKL）算法往往以数据驱动的方式，自动地学习一系列基

础核的最优组合。该过程可以概括如下：

$$\kappa(\mathbf{x},\mathbf{x}') = \sum_{p=1}^{P} \beta_p \kappa^{(p)}(\mathbf{x},\mathbf{x}'), \tag{4-18}$$

其中 $\kappa^{(p)}(\cdot,\cdot)$ 为基础核，而 β_p 为需要优化的、大于零的组合系数。我们可将核理解为一种基于核函数对样本对的相似性度量，因此输入特征不同或者核函数形式、参数不同，都可以引出不同的核。通过将公式4-16中的费舍尔判别比（Fisher Discriminant Ratio，FDR）替换为它的等价二次型，并引入MKL 思想，我们可以构造出如下的mkLFDA 问题：

$$\min_{\mathbf{A},\boldsymbol{\beta}} \operatorname{tr}(\mathbf{A}^T \widetilde{\mathbf{S}}_{\boldsymbol{\beta}}^{(w)} \mathbf{A}),\ \text{s.t. } \operatorname{tr}(\mathbf{A}^T \widetilde{\mathbf{S}}_{\boldsymbol{\beta}}^{(b)} \mathbf{A}) = 1 \text{ and } \boldsymbol{\beta} \geq \mathbf{0}, \tag{4-19}$$

其中，

$$\boldsymbol{\beta} = [\beta_1, ..., \beta_P]^T \in \mathbb{R}^P, \tag{4-20}$$

$$\mathbb{K}^{(i)} = [\mathbf{k}_i^{(1)}, ..., \mathbf{k}_i^{(P)}] \in \mathbb{R}^{s \times P}, \tag{4-21}$$

$$\begin{aligned} \widetilde{\mathbf{S}}_{\boldsymbol{\beta}}^{(w)} &= \sum_{i,j=1}^{s} \frac{1}{2} \widetilde{\mathbf{W}}_{ij}^{(w)} (\mathbb{K}^{(i)} - \mathbb{K}^{(j)}) \boldsymbol{\beta}\boldsymbol{\beta}^T (\mathbb{K}^{(i)} - \mathbb{K}^{(j)})^T, \\ \widetilde{\mathbf{S}}_{\boldsymbol{\beta}}^{(b)} &= \sum_{i,j=1}^{s} \frac{1}{2} \widetilde{\mathbf{W}}_{ij}^{(b)} (\mathbb{K}^{(i)} - \mathbb{K}^{(j)}) \boldsymbol{\beta}\boldsymbol{\beta}^T (\mathbb{K}^{(i)} - \mathbb{K}^{(j)})^T. \end{aligned} \tag{4-22}$$

4.3.2.3 优化算法

对于mkLFDA 问题，直接同时对参数 $\mathbf{A}$ 和 $\boldsymbol{\beta}$ 进行优化不太容易实现。因此，我们采用了论文[147] 中的迭代交替优化的策略，进行模型优化和学习。

具体来说，当参数 $\boldsymbol{\beta}$ 固定，优化参数 $\mathbf{A}$ 时，mkLFDA 问题退化为kLFDA 问题，可以直接利用广义特征值方法求解；当参数 $\mathbf{A}$ 固定，优化参数 $\boldsymbol{\beta}$ 时，可以将原始的mkLFDA 问题重新构造成等价的非凸的带二次约束的二次规划（Quadratically Constrained Quadratic Programming，QCQP）问题，形式如下：

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \boldsymbol{\beta}^T \widetilde{\mathbf{S}}_{\mathbf{A}}^{(w)} \boldsymbol{\beta} \\ &\text{s.t. } \boldsymbol{\beta}^T \widetilde{\mathbf{S}}_{\mathbf{A}}^{(b)} \boldsymbol{\beta} = 1, \text{ and } \boldsymbol{\beta} \geq \mathbf{0}, \end{aligned} \tag{4-23}$$

其中

$$\begin{aligned}\widetilde{\mathbf{S}}_{\mathbf{A}}^{(w)} &= \sum_{i,j=1}^{s} \frac{1}{2}\widetilde{\mathbf{W}}_{ij}^{(w)}(\mathbb{K}^{(i)} - \mathbb{K}^{(j)})^T \mathbf{A}\mathbf{A}^T(\mathbb{K}^{(i)} - \mathbb{K}^{(j)}),\\ \widetilde{\mathbf{S}}_{\mathbf{A}}^{(b)} &= \sum_{i,j=1}^{s} \frac{1}{2}\widetilde{\mathbf{W}}_{ij}^{(b)}(\mathbb{K}^{(i)} - \mathbb{K}^{(j)})^T \mathbf{A}\mathbf{A}^T(\mathbb{K}^{(i)} - \mathbb{K}^{(j)}).\end{aligned} \tag{4-24}$$

由于QCQP 问题属于难解问题，我们可以通过引入一个辅助变量 $\mathbf{B} \in \mathbb{R}^{P\times P}$，将其转化成其半正定规划松弛（Semidefinite Programming Relaxations，SDR）形式，借助SDP算法快速求解[147]。公式4-23的SDR 形式如下：

$$\begin{aligned}&\min_{\boldsymbol{\beta},\mathbf{B}} \operatorname{tr}(\widetilde{\mathbf{S}}_{\mathbf{A}}^{(w)}\mathbf{B})\\ &\text{s.t. } \operatorname{tr}(\widetilde{\mathbf{S}}_{\mathbf{A}}^{(b)}\mathbf{B}) = 1,\ \boldsymbol{\beta} \geq \mathbf{0}, \text{and } \begin{bmatrix} 1 & \boldsymbol{\beta}^T \\ \boldsymbol{\beta} & \mathbf{B} \end{bmatrix} \succeq 0.\end{aligned} \tag{4-25}$$

4.4 实验结果及分析

在本小节中，我们在基于空间金字塔的统计特征提取框架下，按照多种不同的框架参数配置，提取了多种不同类型统计特征，并做了大量对比实验；除此之外，我们将mkLFDA 算法与多种不同的MKL 方法做了性能对比；同时，也将本文提出的基于空间金字塔统计特征和mkLFDA 多特征融合的ReID 算法与大量同期先进算法做了对比，实验结果显示了我们的方法的优越性。

4.4.1 数据集和实验设置

为了更公平可信地展示我们算法的性能，所有的实验主要在四个公开基准数据集上进行实现。我们采用的数据集分别是 VIPeR[4]、CUHK01[5]、PRID2011[100]、和3DPeS[98]。

在实验中，除了数据集 PRID2011 以外，我们将每个数据集都随机等分成两部分，一部分用作训练数据，一部分用作测试数据。而对于数据集 PRID2011，我们随机选取了 100 个同时出现在两个摄像头下的行人用作训练，而将剩下的 100 人联合其余 549 个只出现在一个摄像头下的行人用作测试。

在性能评估时，我们采用了基于单图（Single-Shot）的匹配方法，用CMC 曲线对性能进行量化。每次实验都被随机重复 10 次，平均结果被记录下来作为最终结果。

在实验中，单独评估空间金字塔统计特征提取框架下所提特征性能时，我们以kLFDA 作为默认度量学习方法；另外，kLFDA 也是基于多核学习的特征融合阶段的基础核函数。除了数据集 PRID2011 上，每一种统计特征都利用PCA 算法将维数降低到 300，而在数据集 PRID2011 上，特征维数被降到 100。在kLFDA 算法中，我们采用了高斯核函数，即 $\kappa(\mathbf{x},\mathbf{x}') = \exp(-\frac{||\mathbf{x}-\mathbf{x}'||_2^2}{2\sigma^2})$，默认的带宽参数 $\sigma = 100$；在计算关系矩阵时，对于数据集 CUHK01，我们选择了二近邻约束，对于其他数据集，我们选择了最近邻约束。

4.4.2 空间金字塔统计特征的相关细节

当利用空间金字塔统计特征提取框架，提取三通道彩色图像的spHist、spHOG、和spLBP 特征时，我们采用先逐一提取单一通道特征，再串联成最终特征向量的方式。

当提取spHist 特征时，用于编码初级特征信息的码书，即视觉词典，由每个通道亮度区间的 8 等分点组成。因此，最终的spHist 特征向量维度为 $8 \times 436 \times 3$。

当提取spHOG 特征时，我们将角度 0 到 2π 等分成了 18 个方向间隔。而且，由于在局部对比度归一化阶段，我们采取将区块内基于小胞的特征串联成基于区块的特征的组合方式，因此基于区块的特征维度增大到 $2 \times 2 \times 18 = 72$。因此，最终的spHOG 特征向量维度为 $72 \times 95 \times 3$。

当提取spLBP 特征时，我们使用 3×3 的邻域像素计算LBP 值，因此可以得到 59 个均匀模式。因此，最终的spLBP 特征向量的维度为 $59 \times 95 \times 3$。

当提取spCN 特征时，我们将三维的颜色空间量化到由 N 个颜色名称组成的离散特征空间。实际上，我们选择了论文[16] 中的 16 个颜色名称构成视觉词典。因此，最终的spCN 特征向量维度为 16×436。

当提取spCov 特征时，我们引入了多种初级特征通道来计算乘积特征通道，因此spCov 本身就包含了多种信息的融合。具体来说，我们使用了以下特征信息：

$$[x, y, |\mathrm{G}_x|, |\mathrm{G}_y|, \mathrm{L}, \mathrm{A}, \mathrm{B}, \mathrm{H}, \mathrm{S}, \mathrm{V}_{hsv}, \mathrm{Y}, \mathrm{U}, \mathrm{V}_{yuv}], \tag{4-26}$$

其中 $\mathrm{L}, \mathrm{A}, \mathrm{B}, \mathrm{H}, \mathrm{S}, \mathrm{V}_{hsv}, \mathrm{Y}, \mathrm{U}, \mathrm{V}_{yuv}$ 分别代表不同的初级特征通道。因此，每一小胞内都可以统计出一个 13×13 的协方差矩阵特征。在此基础上，我们又引入了均值向量来弥补协方差特征的不足。最终，完整的spCov 特征向量的维度为 $(91 + 13) \times 95$。

4.4.3 空间金字塔统计特征的详细性能评估

依照空间金字塔统计特征提取框架，我们设计并提取了五种不同种类的局部统计特征，分别是spHist、spHOG、spLBP、spCN、和spCov，涵盖了类直方图、均值向量、和协方差三种统计特性。为了使我们更明了地认识特征提取框架每个步骤对特征性能的影响，我们针对ReID 任务，对整个特征提取过程进行了系统而详细地评估。

在对比实验中，我们采取控制变量法的思想对每个影响因素进行分析，即对比实验实验中只有待评估的因素不同。按照特征提取框架的流程，我们主要对输入图像颜色空间、编码方法、积分图通道类型、局部对比归一化方法、池化方法、特征尺度、以及全局归一化方法这七个因素进行了考量。各因素的默认配置如图4-1所示。具体来说，对于颜色空间，我们对比了 RGB、HSV、YUV、LAB、和 Gray；对于编码方法，我们对比了HE、KCE、LIE、和SCE；对于积分图通道类型，我们对比了原始积分图通道和卷积积分图通道；对于局部对比归一化方法，我们对比了 ℓ_1-norm、ℓ_1-sqrt、ℓ_2-norm、ℓ_2-clip、和 ℓ_1^2-norm；对于池化方法，我们对比了均值池化和最大池化；对于特征尺度，我们对比单尺度和多尺度特征；对于全局归一化方法，我们对比了 ℓ_1-norm 和 ℓ_2-norm。另外，所有的对比实验都在基准数据集 VIPeR 上实现，对比实验结果分别展示在表4-2至表4-6中。

（1）图像颜色空间及初级特征通道

由于图像颜色空间及初级特征通道包含丰富的原始视觉信息，为进一步构造适合ReID 的局部统计特征提供了充足的资源。对于spHist 特征来说（表4-2(a) vs. (c)-(e)），由于 HSV 空间具有更好的光照不变性，因此与 RGB、YUV、和 LAB 相比，基于 HSV 空间的颜色直方图特征可以取得更高的再识别准确度，比如在 Rank-1 上分别高出了 19.53%、10.41%、和 8.89%。对于spHOG（表4-3(a) vs. (c)-(f)）和spLBP（表4-4(a) vs. (c)-(f)）特征来说，由于在ReID 中颜色信息往往比纹理信息更具判别性，因此从颜色空间提取梯度信息或者 LBP 特征，比单单从灰度图像中提取，所包含的信息量更大；这也是在ReID 任务中，原始方向梯度特征和 LBP 特征性能远低于基于颜色空间的方向梯度特征和 LBP 特征的原因所在，比如基于原始 HOG 特征和 LBP 特征的ReID 算法分别只能达到 5.15% 和 4.21% 的 Rank-1 准确度。由于颜色名称通常在 RGB 或者 HSV 空间中都有明确的定义，比如红色在 RGB 中为 $[1,0,0]^T$，在 HSV 空间中为 $[0,1,1]^T$。因此，与其他颜色空间相比，在 RGB 和 HSV 空间中将像素值映射到颜色名称空间会更加合理，也更容易取得更高的准确度（表4-5(a) vs. (c)-(e)）。比如，分别从 RGB 和 HSV 空间中提取的spCN 特征可以分别达到 20.13%

表 4-2 利用不同参数提取的 spHist 特征在 VIPeR 数据集上的 Rank-r 准确率。

参数配置	Rank1	Rank5	Rank10	Rank20	Rank50
(a) **spHist**	35.19	66.30	80.38	**91.23**	98.20
(b) **Hist**	29.49	57.15	71.14	84.15	96.39
(c) RGB	15.66	41.27	57.41	73.70	90.82
(d) YUV	24.78	54.81	69.62	83.58	94.27
(e) LAB	26.30	55.16	70.22	83.83	95.35
(f) KCE, $\gamma = \frac{1}{24}$	34.49	62.66	75.63	87.66	96.84
(g) KCE, $\gamma = \frac{1}{6}$	33.89	65.51	79.05	89.81	98.10
(h) LIE	34.91	65.79	78.58	90.38	97.59
(i) HE	31.93	62.28	75.82	88.04	97.06
(j) Orig	35.70	65.92	79.02	89.59	98.07
(k) No normalization	31.52	63.29	77.15	89.18	98.01
(l) ℓ_1-norm	34.62	67.18	80.51	90.76	**98.23**
(m) ℓ_1-sqrt	**36.74**	**68.20**	80.63	90.66	97.94
(n) ℓ_1^2-norm	35.66	67.85	**80.92**	90.47	98.13
(o) ℓ_2-clip	34.65	67.25	80.38	90.38	98.10
(p) Avg	34.84	67.12	79.97	90.85	98.07
(q) Multi, ℓ_2	35.22	66.27	80.22	90.85	98.13
(r) Single, ℓ_1	35.54	65.98	78.92	89.81	98.10

和 21.17% 的准确度。对于spCov 特征来说（表4-6(a) vs. (c)），通过向初级特征通道中引入更多的颜色信息，可以大幅提升协方差特征的ReID 性能，甚至可以在原始性能基础上提升 30.57%。总之，以上对比实验证明，颜色信息可以极大的丰富初级特征通道的多样性，提升特征在ReID 任务上的辨别能力。

（2）编码方法

生成高级特征通道的方式，对构造高层的ReID 特征至关重要。比如，生成编码特征通道和乘积特征通道分别有助于类直方图特征和协方差特征的提取。然而不幸的是，在直方图特征提取过程中，虽然可以通过将原始的图像信息量化到离散的特征空间得到更鲁棒的特征表示，但却同时带来了不可逆的由量化误差引起的信息损失。与原始的利用硬投票（Hard Voting）方法量化编码的颜色直方图相比，我们通过使用KCE（表4-2(a) vs. (i)）或者LIE（表4-2(h) vs. (i)）等软编码方式，使颜色直方图特征在ReID 中的 Rank-1 准确度分别提升了 3.26% 和 2.98%。尽管原始的HOG 特征

已经通过采用LIE 方法来减少量化误差，我们仍可以通过引入更有效的编码方法进一步提升HOG 特征的准确度，比如通过使用KCE 方法，Rank-1 提升了 3.54%（表4-3(a) vs. (i)）。然而，值得注意的是，在使用软编码算法时，不恰当的参数设置，也可能造成特征性能的轻微下降。例如，当利用KCE 方法时，分别将其参数 γ 设置为 $1/24$ 或者 $1/6$ 时，会使spHist 的 Rank-20 准确度分别下降 3.57% 或 1.42%（表4-2(a) vs. (f) and (g)）；分别将 γ 设置为 $\pi/18$ 或者 $\pi/2$ 时，会使spHOG 的 Rank-1 准确度分别降低 3.04% 和 4.05%(表4-3(a) vs. (g) and (h))。对于spCN 特征来说 (表4-5(a) vs. (f)-(i))，SCE 方法可以取得比HE 和KCE 等方法更好的性能，Rank-1 准确度可以达到 20.13%。总之，以上对比实验结果证明，选择合适的编码算法对所提取特征的性能有着至关重要的作用；一般来说，基于软编码的类直方图特征性能要优于基于硬编码的类直方图特征。

表 4-3　利用不同参数提取的 spHOG 特征在 VIPeR 数据集上的 Rank-r 准确率。

参数配置	Rank1	Rank5	Rank10	Rank20	Rank50
(a) **spHOG**	26.58	**56.96**	**71.39**	**83.67**	**95.66**
(b) **HOG**	5.16	16.65	26.90	39.94	60.76
(c) Gray	5.92	20.32	30.76	46.08	68.39
(d) HSV	23.92	50.92	63.86	78.20	92.34
(e) RGB	9.49	27.75	40.38	56.96	80.82
(f) LAB	24.18	53.07	67.91	81.96	95.06
(g) KCE, $\gamma=\frac{\pi}{18}$	23.54	52.78	66.90	80.66	94.37
(h) KCE, $\gamma=\frac{\pi}{2}$	22.53	50.38	65.70	79.81	93.86
(i) LIE	23.04	52.12	66.55	80.54	94.40
(j) HE	16.52	40.89	55.60	71.87	88.45
(k) Orig	23.83	53.77	68.99	82.59	94.91
(l) No normalization	17.44	41.84	55.35	71.77	89.59
(m) ℓ_1-norm	23.48	52.69	66.30	79.08	92.82
(n) ℓ_1-sqrt	23.45	52.06	65.92	79.59	92.63
(o) ℓ_2-norm	24.24	53.07	67.88	79.75	92.53
(p) ℓ_2-clip	23.07	52.44	66.93	79.49	93.48
(q) Max	24.46	53.67	68.99	82.72	94.62
(r) Multi, ℓ_1-norm	**27.25**	55.89	70.22	83.54	94.91
(s) Single, ℓ_2-norm	21.20	48.45	63.54	78.70	93.23

表 4-4 利用不同参数提取的 spLBP 特征在 VIPeR 数据集上的 Rank-r 准确率。

参数配置	Rank1	Rank5	Rank10	Rank20	Rank50
(a) **spLBP**	20.32	**49.97**	**64.08**	**78.99**	**94.08**
(b) **LBP**	4.21	14.21	22.82	37.78	60.95
(c) Gray	4.49	17.37	29.21	45.13	69.46
(d) HSV	13.58	36.46	51.99	68.13	88.99
(e) RGB	9.72	29.62	43.83	60.98	83.77
(f) LAB	16.61	41.61	56.87	73.20	90.35
(g) Orig	19.18	47.37	61.80	76.93	93.58
(h) No normalization	14.18	38.77	54.78	72.47	92.15
(i) ℓ_1-norm	18.32	45.19	60.32	77.15	92.63
(j) ℓ_1-sqrt	18.01	44.91	59.62	75.44	91.77
(k) ℓ_1^2-norm	19.18	46.74	62.50	78.35	93.54
(l) ℓ_2-clip	18.45	45.22	61.74	76.84	93.04
(m) Avg	19.53	47.78	62.41	78.01	93.77
(n) Multi, ℓ_1-norm	**20.47**	48.29	62.63	78.07	93.39
(o) Single, ℓ_2-norm	19.11	45.47	60.92	76.27	92.47

（3）积分图通道类型

积分图通道被当做一种中间特征通道引入到特征提取框架中，以加快基于矩形块的统计特征的计算过程。通过借助于卷积积分图通道，可以在加快计算的同时，减轻统计特征提取过程中矩形块边缘像素造成的空间混叠问题。与使用原始积分图通道相比，使用卷积积分图通道可以轻微地提升行人匹配的准确度（表4-2(a) vs. (j)、表4-3(a) vs. (k)、表4-4(a) vs. (g)、表4-5(a) vs. (j)）；例如，Rank-5 准确度分别在spHist、spHOG、spLBP、spCN 上提升了 0.38%、3.19%、2.7%、0.45%。另外，我们还发现，使用积分图通道带来的性能提升，在基于较大尺寸局部区域的特征上比基于较小尺寸局部区域的特征上明显；例如在基于 8×8 的小胞特征上比基于 4×4 的小胞特征上提升更明显。

（4）局部对比归一化方法

由于图像上的光照和背景往往会发生剧烈地变化，因此有必要对特征进行局部对比归一化，从而得到更稳定的特征表示。在实验中，我们分别对比了两种特征组合方法和五种不同的局部归一化方法。对于spHOG 特征来说（表4-3(a) vs. (m)-(p)），引入 ℓ_1^2-norm 归一化方法，可以使 Rank-1 的准确度从 23.07% 提升到 26.58%。而对

表 4-5　利用不同参数提取的 spCN 特征在 VIPeR 数据集上的 Rank-r 准确率。

参数配置	Rank1	Rank5	Rank10	Rank20	Rank50
(a) **spSCN**	20.13	48.96	**63.99**	**78.42**	90.92
(b) **SCN**	19.02	46.49	61.11	75.41	88.99
(c) HSV	**21.17**	**49.34**	63.83	77.53	**92.63**
(d) YUV	20.06	44.94	59.43	74.08	89.65
(e) LAB	14.56	37.25	52.88	68.73	86.74
(f) KCE, $\gamma = \frac{1}{10}$	12.25	31.30	43.83	57.85	75.47
(g) KCE, $\gamma = \frac{1}{2}$	15.60	40.70	54.84	71.90	89.59
(h) KCE, $\gamma = 1$	14.97	37.31	51.27	67.69	85.73
(i) HE	16.61	41.52	55.28	70.09	86.68
(j) Orig	20.13	48.51	63.39	77.91	90.79
(k) No normalization	16.27	43.54	58.80	74.94	88.96
(l) ℓ_1-norm	18.89	47.25	61.20	76.74	89.94
(m) ℓ_1-sqrt	20.70	48.64	63.20	77.75	91.46
(n) ℓ_1^2-norm	19.97	47.94	63.07	77.97	90.16
(o) ℓ_2-clip	19.56	47.59	61.96	76.23	89.49
(p) Avg	20.28	48.35	62.72	76.42	90.00
(q) Multi, ℓ_1-norm	20.16	48.48	63.67	78.32	90.98
(r) Single, ℓ_2-norm	19.49	45.82	61.08	76.17	89.75

表 4-6　利用不同参数提取的 spCov 特征在 VIPeR 数据集上的 Rank-r 准确率。

参数配置	Rank1	Rank5	Rank10	Rank20	Rank50
(a) **spCov**	**34.91**	66.80	79.15	**89.87**	98.20
(b) **Cov**	3.01	10.47	17.82	27.41	47.50
(c) Gray	4.34	14.94	23.89	37.34	61.17
(d) No normalization	18.77	47.47	63.23	80.06	94.78
(e) ℓ_1-norm	33.10	66.23	79.18	**89.87**	**98.29**
(f) ℓ_1^2-norm	33.42	64.27	77.91	88.45	97.34
(g) Max	34.46	65.63	79.08	88.99	97.85
(h) Without Mean	30.38	61.46	74.72	86.39	96.39
(i) Multi, ℓ_1-norm	33.73	**67.15**	**79.27**	88.99	98.20
(j) Single, ℓ_2-norm	31.80	63.04	77.56	88.54	97.50

于其他特征来说（表4-2(a) vs. (l)-(o)、表4-4(a) vs. (i)-(l)、表4-5(a) vs. (l)-(o)、表4-6(a) vs. (e) and (f)），不同的归一化方法仅仅带来微弱的 1.5%-2.3% 的性能变化。有趣的是，如果抛弃局部对比归一化步骤，所提取的特征性能会迅速恶化（表4-2(a) vs. (k)、表4-3(a) vs. (l)、表4-4(a) vs. (h)、表4-5(a) vs. (k)、表4-6(a) vs. (d)），例如spHist、spHOG、spLBP、spCN、spCov 的 Rank-1 准确度分别从 35.19% 下降到 31.52%、从 26.58% 下降到 17.44%、从 20.32% 下降到 14.18%、从 20.13% 下降到 16.27%、从 34.91% 下降到 18.77%。

（5）池化、多尺度、以及归一化

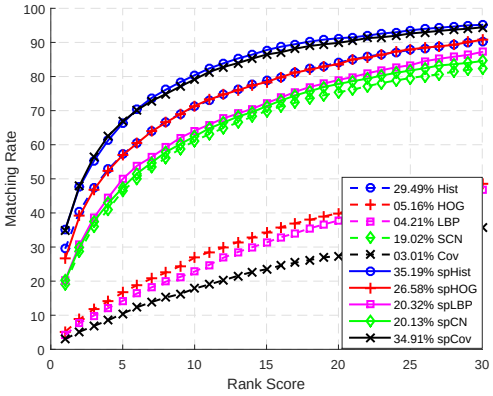
局部特征的池化以及多尺度特征的融合是构造层次化特征的关键，同时也使特征具有了更好的局部干扰不变性，并且缓解了特征的局部不对齐问题。对于某些特征来说，特征中局部显著成分往往直接决定这种特征的性能，因此采用最大池化提取特征的显著成分可以进一步带来特征性能的提升，比如借助于最大池化操作spHOG 的 Rank-1 准确度与均值池化相比可以提升 2.12%（表4-3(a) vs. (q)）。然而这种性能提升大多数情况下并不十分明显（表4-2(a) vs. (p)、表4-4(a) vs. (m)、表4-5(a) vs. (p)、表4-6(a) vs. (g)）。此外，我们也对单尺度的统计特征与多尺度的统计特征进行了性能对比。根据实验结果（表4-2(a) vs. (r)、表4-3(a) vs. (s)、表4-4(a) vs. (o)、表4-5(a) vs. (r)、表4-2(a) vs. (j)），我们发现，多尺度特征通常能带来性能的稳定提升。具体来说，spHist、spHOG、spLBP、spCN、spCov 的 Rank-5 准确度分别提升了 0.32%、8.51%、4.50%、3.14%、3.76%。然而，对不同尺度的特征进行归一化操作，似乎对最终串联特征的影响并不大（表4-2(a) vs. (q)、表4-3(a) vs. (r)、表4-4(a) vs. (n)、表4-5(a) vs. (q)、表4-6(a) vs. (i)）。

除此之外，对于spCov 特征来说，引入均值向量信息（一阶统计信息）作为协方差（二阶统计信息）的补充可以明显提升特征的ReID 性能，例如，Rank-1 准确率从 30.38% 提升到了 34.91%（表4-6(a) vs. (h)）。

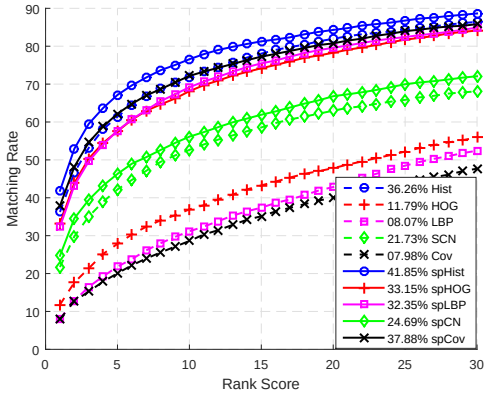
4.4.4 空间金字塔统计特征与原始特征的性能对比

我们将基于空间金字塔的统计特征与原始特征的性能进行了对比，所有的特征都在统一框架下提取，特征提取时对应的参数配置见表4-1。尽管，基于不同金字塔统计特征的ReID 算法性能并不相同，比如在数据集 VIPeR 上（如图4-7(a)），spHist 和spCov 特征分别取得了 35.19% 和 34.19% 的 Rank-1 匹配精度，而spLBP 和spCN 特征分别只能取得 20.32% 和 20.13% 的准确度；并且，每种空间金字塔特征的相对优

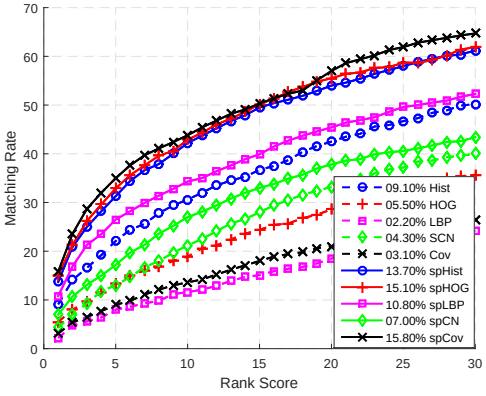
势在不同数据集上的也不尽相同，比如spHist 在数据集 VIPeR 上可以取得最好匹配结果（如图4-7(a)），而在数据集 PRID2011 上与其他特征相比spHist 只能取得中等匹配性能（如图4-7(c)）。然而，图4-7中所有的实验结果都证明，基于空间金字塔的统计特征与其相应的原始统计特征相比，都取得了一致性地大幅度性能提升。具体来说，与原始特征相比，spHist 在数据集 3DPeS 上将 Rank-1 提升了 6.71%；spHOG 在数据集 VIPeR 上将匹配准确度提升了 21.42%；spLBP 在数据集 CUHK01 上将匹配准确度提升了 24.28%；spCN 在数据集 PRID2011 上将准确度提升了 2.7%；spCov 在数据集 VIPeR 上 Rank-1 准确度提升了 31.9%。



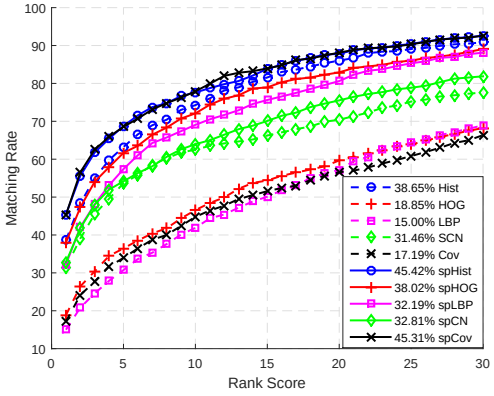
(a) VIPeR



(b) CUHK01



(c) PRID2011



(d) 3DPeS

图 4-7　不同数据集上对比实验的 CMC 曲线结果。(a) - (d)：空间金字塔统计特征与其相应的原始特征在四个基准数据集上的性能对比实验。

经过大量实验和分析，我们可以将空间金字塔统计特征带来的性能提升的原因归结为如下几个方面：

- 丰富的初级特征通道信息可以增强特征的表示能力和可辨识度；
- 合适的编码方法可以减少量化误差带来的信息损失；
- 恰当的局部对比归一化方法可以平滑掉特征的局部噪音变化；
- 空间金字塔的层次化策略可以减轻特征的局部不对齐问题。

4.4.5 mkLFDA 与其他多核学习算法的性能对比

为了进一步提升ReID 算法的准确度，我们利用mkLFDA 算法对空间金塔框架提取的多种特征进行多特征融合。一般来说，最优的集成核，是对若干个基础核的线性组合，如公式4-18所示。为了增加核的多样性，我们一共设计了 9 种不同的基础核，其中包括利用高斯核函数计算的 5 种空间金字塔统计特征的相似度核以及利用径向基函数 (Radial Basis Function，RBF) χ^2 核函数（$\kappa(\mathbf{x},\mathbf{x}') = \exp(\frac{\sum_i (2x_i x_i')/(x_i+x_i')}{2\sigma^2})$）计算的 4 种空间金字塔类直方图特征的相似度核。需要指出的是，当使用RBF χ^2 核函数计算相似核时，我们会去除掉PCA 降维过程；并且依照经验当提取spHist 特征时将 σ 设置为 1，当提取其他特征时将 σ 设置为 10。图4-8展示了基于单一基础核的ReID 算法与基于集成核的ReID 算法的性能对比结果。另外，我们将mkLFDA 算法与多种不同的集成核学习方法进行了对比，分别包括：基于算数平均的多核集成、基于几何平均的多核集成、以及基于核对齐的多核集成。除此之外，我们还将基于mkLFDA 的多特征融合与直接串联的多特征融合方法进行性能对比。为了进一步提升多特征融合的性能，我们又引入了利用预训练的 AlexNet 提取的CNN 特征作为补充。

基于算术平均的多核集成： 在此方法中，集成核可以通过直接求所有基础核的算数平均值得到，即：

$$\kappa(\mathbf{x},\mathbf{x}') = \frac{1}{P}\sum_{p=1}^{P}\kappa^{(p)}(\mathbf{x},\mathbf{x}'). \tag{4-27}$$

基于几何平均的多核集成： 在此方法中，集成核可以通过直接求所有基础核的几何平均值得到，即：

$$\kappa(\mathbf{x},\mathbf{x}') = \sqrt[P]{\prod_{p=1}^{P}\kappa^{(p)}(\mathbf{x},\mathbf{x}')}. \tag{4-28}$$

基于核对齐的多核集成： 在此方法中，首先定义两个核的对齐程度为：

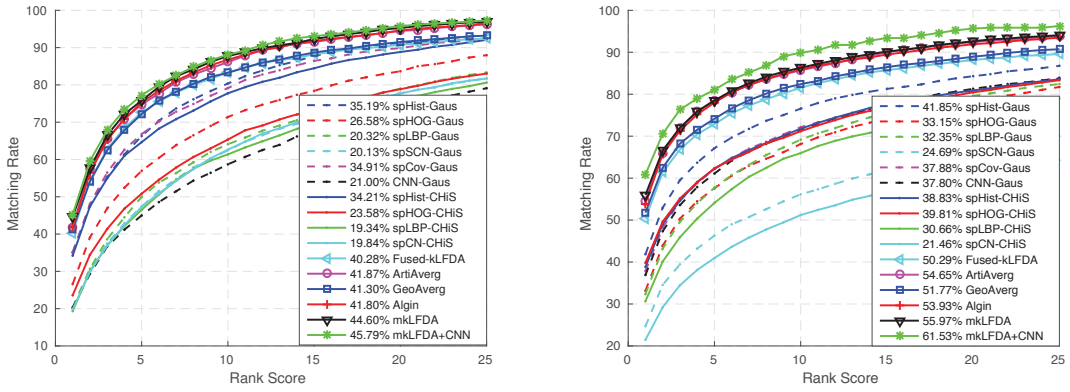
$$A(\kappa_i,\kappa_j) = \frac{\langle \mathbf{G}_i,\mathbf{G}_j\rangle_F}{\sqrt{\langle \mathbf{G}_i,\mathbf{G}_i\rangle_F\langle \mathbf{G}_j,\mathbf{G}_j\rangle_F}}, \tag{4-29}$$

其中 $\mathbf{G}$ 为核所对应的 Gram 矩阵，$\langle\cdot,\cdot\rangle_F$ 定义为矩阵逐元素乘积的和；接着，每个基础核所对应的组合系数可以通过计算它与目标核 $\hat{\boldsymbol{\kappa}}$ 的对齐程度得到，即：

$$\beta_p = \frac{A(\boldsymbol{\kappa}^{(p)}, \hat{\boldsymbol{\kappa}})}{\sum_{p=1}^{P} A(\boldsymbol{\kappa}^{(p)}, \hat{\boldsymbol{\kappa}})}. \tag{4-30}$$

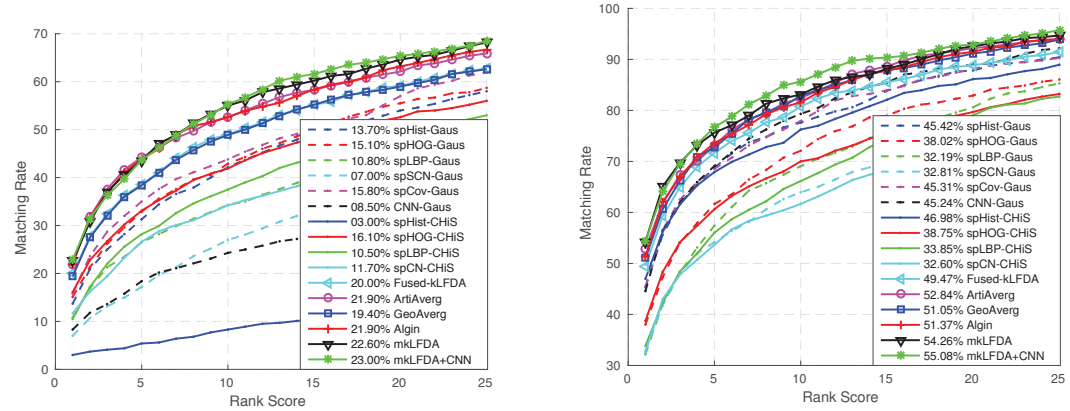
在我们的实验中，如果样本对具有相同标签，我们则令目标核 $\hat{\boldsymbol{\kappa}}$ 对应的 Gram 矩阵中的相应元素为 1；如果样本对具有不同标签，则将相应的 Gram 矩阵中的元素设为 0。

mkLFDA：在模型训练阶段，我们没有直接初始化参数 $\mathbf{A}$，而是将 $\mathbf{AA}^T$ 初始化为单位矩阵。尽管我们没能给出优化算法收敛的理论保证，但是通过实验观察，我们发现迭代交替优化算法可以使mkLFDA 在实验中很快地趋于收敛。



(a) VIPeR

(b) CUHK01



(c) PRID2011

(d) 3DPeS

图 4-8　不同数据集上对比实验的 CMC 曲线结果。(a) - (d)：基于单一基础核的 ReID 算法与基于集成核的 ReID 算法在四个基准数据集上的性能对比实验。

根据在四个基准数据集上的结果，我们发现基于集成核的ReID 算法通常明显地优于基于单一核的ReID 算法，即使仅仅采用最简单的基于算术平均的集成核方法。与其他MKL 方法以及特征融合方法相比，mkLFDA 可以取得更好的行人匹配结果。具体来讲，利用mkLFDA 算法进行多特征融合，分别将数据集 VIPeR、CUHK01、PRID2011、3DPeS 的 Rank-1 准确度提升到了 44.60%、55.97%、22.60%、54.26%。同时，我们还发现CNN 特征可以取得与空间金字塔特征类似的ReID 性能；而且，当我们将其引入到多特征融合框架中时，还可以带来进一步地再识别准确率的提升，比如使 CUHK01 上的 Rank-1 准确率从 56.0% 上升到 61.5%。

4.4.6 与同期其他先进 ReID 算法的性能对比

为了进一步证明我们所提出的基于空间金字塔统计特征和mkLFDA 多特征融合的ReID 算法的有效性，我们将其与大量同期最先进的ReID 算法进行了性能对比。这些算法可以大致被分为以下几类：

- 多特征融合算法（Ensemble[35]、LateFusion[17]、和 ELF[9]），
- 特征提取和特征学习（MidFilter[14]、MTL[20]、Transfer[19]、SCNCD[16]、ExplicitPoly[36]、SalMatch[30]、Salience[13]、ColorInv[15]、ViewInv[43]、eBiCov[148]、SDALF[10]、和 Attribute[149]），
- 度量学习（MLAPG[37]、XQDA[18]、KernelML[33]、RMLLC[42]、LADF[27]、LAFT[29]、MtMCML[34]、RPLM[26]、FuncSpace[102]、LFDA[28]、KISSME[24]、PCCA[25]、RDC[31]、和 PRDC[23]），
- 深度学习（ImprDeep[67] 和 DeepReid[66]），
- 字典学习和稀疏表示（CPDL[39]、ISR[127]、和 SSCDL[32]），
- 其他方法（MirrorRep[21]、CSL[40]、和 CompTemp[132]）。

另外，需要注意的是，每种算法的再识别准确度都是直接引用自原始论文的图表中。

我们将在四个基准数据集上的对比结果分别展示在了表4-7至表4-10中，并且将最优性能标记为红色，将次优性能标记为绿色。通过对比发现，基于空间金字塔统计特征和mkLFDA 多特征融合的ReID 算法的性能堪比同期最好算法，甚至在某些数据集上还超过了其他所有对比算法的结果。

具体来说，由于 VIPeR 数据集的应用最为广泛，大量的ReID 算法都针对该数据集进行了优化，我们的算法在该数据集上可以达到次好水平，而且与最好性能只有很小的差距。在中等规模数据及 CUHK01 上，我们的算法可以达到同期最高的行人

表 4-7　在数据集 VIPeR 上，基于空间金字塔统计特征和 mkLFDA 多特征融合的 ReID 算法与同期其他先进 ReID 算法的 Rank-*r* 准确率比较。

方法	Rank1	Rank5	Rank10	Rank20	Rank50
mkLFDA+CNN	**45.8**	**77.3**	**88.4**	**95.7**	**99.9**
mkLFDA	44.6	75.8	87.7	95.3	**99.5**
Ensemble[35]	**45.9**	**77.5**	**88.9**	**95.8**	**99.5**
LateFusion[17]	30.2	51.6	62.4	73.8	-
ELF[9]	12.0	31.0	41.0	58.0	-
MidFilter[14]	43.4	73.0	85.0	93.7	-
MTL[20]	42.3	72.2	81.6	89.6	-
Transfer[19]	41.6	71.9	86.2	95.1	-
SCNCD[16]	37.8	68.5	81.2	90.4	97.0
ExplicitPoly[36]	36.8	70.4	83.9	91.7	97.8
SalMatch[30]	30.2	52.0	65.5	79.2	-
Salience[13]	26.7	50.7	62.4	76.4	-
ColorInv[15]	24.2	-	57.1	69.7	87.0
ViewInv[43]	21.4	45.9	62.6	79.7	-
eBiCov[148]	20.7	42.0	56.2	68.0	-
SDALF[10]	19.9	38.9	49.4	65.7	92.2
Attribute[149]	17.4	39.0	50.8	-	86.4
MLAPG[37]	40.7	-	82.3	92.4	-
XQDA[18]	40.0	-	80.5	91.1	-
KernelML[33]	36.1	68.7	81.3	91.1	-
RMLLC[42]	31.3	62.1	75.3	86.7	-
LADF[27]	30.0	65.0	79.0	91.0	98.0
LAFT[29]	29.6	-	69.3	-	96.8
MtMCML[34]	28.8	59.3	75.8	88.5	-
RPLM[26]	27.0	-	69.0	83.0	95.0
FuncSpace[102]	25.8	-	69.6	83.7	95.1
LFDA[28]	24.2	-	67.1	-	94.1
KISSME[24]	22.0	-	68.0	-	93.0
PCCA[25]	19.6	48.9	64.9	80.3	-
RDC[31]	18.3	42.7	57.8	72.4	-
ImprDeep[67]	34.8	63.0	76.0	-	-
CPDL[39]	34.0	64.2	77.5	88.6	-
ISR[127]	27.0	-	61.0	72.0	94.1
SSCDL[32]	25.6	53.7	68.1	83.6	-
MirrorRep[21]	43.0	75.8	87.3	94.8	-
CSL[40]	34.8	68.7	82.3	91.8	96.2
CompTemp[132]	24.0	47.0	60.0	75.0	-

再识别水准，并领先于包括其他特征融合算法以及深度学习算法在内所有工作，例如，在 Rank-1 准确率上分别领先 8.1% 和 14.0%。此外，如果按照论文[39] 中的划分探测集和候选集的方式，即在不考虑摄像机标签的情况下进行随机划分，我们的方法甚至可以在 CUHK01 上达到 62.59% 的 Rank-1 准确率，远超论文[39] 中的 59.47%。在更符合实际应用场景的 PRID2011 数据集上，我们的算法依旧可以达到同期最高水平，并进一步提高了行人再识别的准确率，将 Rank-1 提升了 5.1%。然而，由于 3DPeS 数据集所包含的行人图像很不规律，因此无法通过kLFDA 学习单一全局投影矩阵实现正确匹配。因此我们的方法并没有取得同期最好水平，但是尽管如此，我们的方法仍与其他多特征融合、特征提取和学习、深度学习等方法具有一定的可比性。

表 4-8 在数据集 CUHK01 上，基于空间金字塔统计特征和 mkLFDA 多特征融合的 ReID 算法与同期其他先进 ReID 算法的 Rank-r 准确率比较。

方法	Rank1	Rank5	Rank10	Rank20	Rank50
mkLFDA+CNN	**61.5**	**81.7**	**90.0**	**96.3**	**99.1**
mkLFDA	**56.0**	**78.5**	**86.3**	**92.6**	**97.4**
Ensemble[35]	53.4	76.4	84.4	90.7	96.4
MidFilter[14]	34.3	55.1	65.0	74.9	-
Transfer[19]	31.5	52.5	65.8	77.6	-
SalMatch[30]	28.5	46.0	56.0	-	-
ImprDeep[67]	47.5	72.0	80.0	-	-
DeepReid[66]	27.9	64.0	77.0	88.0	-
MirrorRep[21]	40.4	64.6	75.3	84.1	-

4.4.7 算法时间复杂度分析

尽管在我们提出的整个行人再识别算法流程里，包含了多种不同的空间金字塔统计特征提取、以及多特征融合等操作，但是算法整体的执行效率仍是可接受的。表4-11对算法执行效率进行了展示①，以 VIPeR 数据集为例，算法最耗时的是基于mkLFDA 多特征融合的模型训练阶段。而一旦模型完成训练，每幅行人图像的再识别过程都控制在 1 秒之内。在未来的工作中，我们可以通过对特征提取算法进行

① 需要指出，我们使用 Matlab 在 Linux 系统下进行算法开发，系统硬件为普通家用 PC，内存 8GB，主频 3.10GHz。

表 4-9　在数据集 PRID2011 上，基于空间金字塔统计特征和 mkLFDA 多特征融合的 ReID 算法与同期其他先进 ReID 算法的 Rank-*r* 准确率比较。

方法	Rank1	Rank5	Rank10	Rank20	Rank50
mkLFDA+CNN	**23.0**	**43.7**	**55.4**	**65.4**	**78.9**
mkLFDA	**22.6**	**43.5**	**55.0**	64.6	77.5
Ensemble[35]	17.90	39.0	49.0	62.0	-
MidFilter[14]	12.5	23.9	30.7	36.5	51.6
MTL[20]	18.0	37.4	50.1	**66.6**	**82.3**
SalMatch[30]	4.9	17.5	26.1	33.9	47.8
RPLM[26]	15.0	-	42.0	54.0	70.0
PCCA[25]	3.5	10.9	17.9	27.1	45.0
PRDC[23]	4.5	12.6	19.7	29.5	46.0

表 4-10　在数据集 3DPeS 上，基于空间金字塔统计特征和 mkLFDA 多特征融合的 ReID 算法与同期其他先进 ReID 算法的 Rank-*r* 准确率比较。

方法	Rank1	Rank5	Rank10	Rank20	Rank50
mkLFDA+CNN	**55.1**	76.9	**86.2**	**93.0**	**99.7**
mkLFDA	54.3	75.6	83.1	92.6	**99.3**
Ensemble[35]	53.3	77.0	85.0	92.0	-
kernelML[33]	54.0	**77.7**	86.0	92.4	-
LFDA[28]	33.4	-	70.0	-	95.1
PCCA[25]	41.6	70.5	81.3	90.4	-
rPCCA[33]	47.3	75.0	84.5	91.9	-
CSL[40]	**57.9**	**81.1**	**89.5**	**93.7**	-

表 4-11　在数据集 VIPeR 上，算法运行时间复杂度分析。

Feature Extraction Phase					Matching Phase	
spHist	spHOG	spLBP	spCN	spCov	training	testing
~90ms	~80ms	~260ms	~60ms	~150ms	~20min	~260ms

并行优化，进一步提升算法的运行效率。

4.5 本章小结

本章，我们分别从引言、相关工作、基于空间金字塔的统计特征提取框架、基于多核局部费舍尔判别分析的特征融合、以及相关实验结果和分析等方面，对我们提出的基于空间金字塔统计特征及多核学习的行人再识别算法进行详细介绍。我们提出了一种基于多种空间金字塔特征的行人再识别算法，并对其做了全面而详细地实验分析和评估。在这个框架下，我们提取了三种类型的图像局部统计特征，分别是直方图分布、一阶均值向量、和二阶协方差矩阵。具体来说，我们实际上提取了五种不同的空间金字塔统计特征实例，包括：spHist、spHOG、spLBP、spCN、以及spCov；并且我们利用mkLFDA 算法对多种特征进行了融合。在基准数据集上的大量实验证明，丰富的初级特征通道、合适的编码方式、恰当的局部对比归一化、以及多尺度空间金字塔特征提取策略，都大大提升了局部统计特征的性能。与同期其他ReID 算法的比较也证明，我们的算法可以比肩甚至超过同期最高水准。我们希望，我们所提出的特征提取和融合框架，可以更好地引导和帮助计算机视觉领域的其他研究人员设计出更高精度的行人再识别系统，以满足实际应用的需求。

第五章 基于上下文敏感特征序列及双重注意力匹配的行人再识别网络

本章分别从引言、相关工作、上下文敏感的特征序列提取模块、基于双重注意力机制的特征序列匹配模块、以及相关实验结果和分析等方面，对我们提出的基于上下文敏感特征序列及双重注意力匹配的行人再识别网络进行详细介绍。

5.1 引言

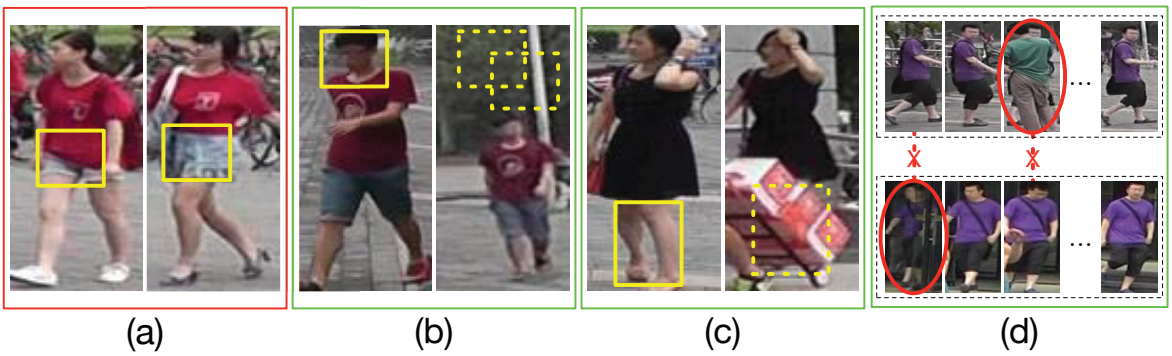


图 5-1 行人再识别中的错误样例。(a)：具有相似衣着的负样本对；(b)：具有较大空间位置不对齐的正样本对；(c)：具有严重身体部分遮挡的正样本对；(d)：含有干扰帧（图中的椭圆标记位置）以及具有时间戳不对齐（图中的“×”符号标记的位置）的正视频样本对。

行人再识别的目的是，对跨摄像头的行人抓拍图像或者行人跟踪序列进行数据关联[6,7]。由于这种跨摄像头行人视觉数据的关联是智能视觉监控、以及基于内容的图像或视频检索等应用的技术基础，因此行人再识别任务开始越来越多的受到广大工业界开发人员以及学术界研究人员的关注。

一种典型的ReID 算法流程是，先将行人图像或者行人跟踪视频表示为单一特征向量，然后再在一个任务专用的度量空间中进行特征向量的距离或者相似度计算。为了取得高精度的再识别结果，我们通常会设计相应的机器学习算法，使具有同一身份的行人特征向量对在度量空间中距离尽量近，同时使具有不同身份的行人特征向量对在度量空间中距离尽量远，如论文[18,27,31,57]。近年来，随着深度学习算法在计算机视觉等领域的成功应用，这种基于单一特征向量的ReID 算法开始取得大幅度的性能提升，如论文[1,56,61,87]。然而，当监控场景中的行人发生严重的外观变化或者

具有不同身份的行人衣着极其相似时，仅仅利用单一特征向量进行准确地行人身份一致性匹配变得困难重重。如图5-1(a) 所示，不同的行人由于衣着风格类似，呈现出极其相似的整体视觉外观，并且只在某些视觉细节上存在细微差异（图中的黄色矩形框位置）；不幸的是，基于单一特征向量的ReID 算法，往往更关注于行人整体的外观特点，而忽略了这些重要细节，因此算法在这种场景中经常会失效。另外，如图5-1(b) 所示，在一些行人跟踪视频中有时也会存在部分干扰帧，这些局部干扰往往会造成视频的全局特征向量的污染，从而导致误匹配的发生。

一种可以缓解以上问题的ReID 算法流程是，先将行人图像或者行人跟踪视频表示成由一系列特征向量组成的特征集合或者特征序列，然后进行基于特征集合或者特征序列的距离或者相似度匹配①。例如，在论文[40,66,67,82,83] 中，研究人员利用基于空间图像块的特征序列或者利用基于人体部分的特征集合表示图像中的行人外观，然后借助于启发式的局部对应关系结构指导特征序列或者特征集合的匹配；又如，在论文[54,60,62] 中，研究人员利用基于子视频段或者基于视频帧图像的特征序列表示视频中的行人外观和动作，然后利用特征序列的距离或者相似性度量实现基于跟踪视频的行人匹配。以上这些方法，要么是依靠一个通用的局部对应关系结构指导局部空间或者局部时空特征序列的距离度量或相似度匹配，要么是依靠人体部分结构指导局部语义特征集合的距离度量或者相似度匹配。然而，对于前一类算法来说，如图5-1(b) 所示的空间局部不对齐或者如图5-1(d) 所示的局部干扰帧都会导致特征序列严重的局部不对齐，从而使ReID 算法失效；而对于后一类算法来说，如图5-1(d) 所示的严重的身体遮挡会导致身体局部结构的丢失，从而使ReID 算法发生误匹配。

为了解决以上所列的ReID 中可能面临的挑战，我们提出了一种新颖的端到端可训练的ReID 模型，称之为双重注意力匹配网络 (Dual Attention Matching Network，DuATM) 。利用该模型，我们可以同时学习上下文敏感的行人特征序列，并且实现基于注意力机制的特征序列的自动匹配。图5-2对模型的基本框架以及基本原理进行了图解展示。具体来说，DuATM 分别包含特征序列提取模块（Feature Sequence Extraction Module）和特征序列匹配模块（Sequence Matching Module）两部分。对于行人图像来说，特征序列提取模块以CNN 为基础，借助于CNN 结构可以提取出图像的空间上下文敏感信息；对于行人视频来说，特征序列提取模块以双向循环卷积神经网络 (Bi-Recurrent Convolutional Neural Network，Bi-Recurrent CNN) 为基础，借助

① 需要指出，在这里我们将一组包含了空间或者时间相邻关系的特征称为特征序列，而将一组不含有空间或者时间相邻关系的特征称为特征集合。

于Bi-Recurrent CNN 结构可以提取视频的时空上下文敏感信息。而无论是基于图像的ReID，还是基于视频的 ReID，特征序列匹配模块都以双重注意机制为基础。其中序列内的注意力机制（Intra-Sequence Attention）可以利用被污染的局部特征向量的上下文信息，实现特征序列内部的局部去噪和精炼（Refinement）；而序列间的注意力机制（Inter-Sequence Attention）可以从被匹配特征序列中自动地为每个局部特征向量选择最佳匹配项，从而实现序列间的局部语义对齐（Alignment）。经过去噪精炼和对齐之后，特征序列对的距离或者相似性度量可以通过直接地计算和融合相应的局部特征向量对之间的距离或相似性度量得到。另外，在模型训练时，我们将DuATM构造成含有三个分支的孪生网络（Siamese Network）结构，利用参数共享的特征序列提取模块提取样本三元组的特征表示；此外，我们使用 Triplet Loss 来促使网络预测的序列距离适用于ReID 任务，并且辅助使用 Cross Entropy Loss 来促使模型学习到辨识度高的中间特征，同时再辅助使用新提出的 De-Correlation Loss 来抑制所学特征序列内特征向量间的相关性。

本章的主要贡献是：

- 提出了一种新颖的端到端可训练的ReID 模型DuATM，该模型可以同时学习上下文敏感的行人特征序列，并实现基于双重注意机制的特征序列匹配；
- 提出利用双重注意力机制进行特征序列内的特征去噪精炼和特征序列间的特征对齐；
- 将DuATM 构造成含有三分支的孪生网络进行训练，并联合使用了 Triplet Loss、Cross Entropy Loss、和 De-Correlation Loss 进行模型参数优化；
- 在多个行人图像和行人视频基准数据集上对算法的有效性进行了验证，证明利用DuATM 可以达到同期最高的再识别准确度。

5.2 相关工作

一个完整的ReID 算法通常由两部分组成，即特征提取和度量学习。目前大量的研究工作也主要集中在这两个方面，或者是构造信息量丰富的特征表示，或者是学习判别能力高的距离度量。按照输入到特征匹配阶段的特征表示的构成形式，我们将现有的算法分为以下两类：其一，基于特征向量的ReID 算法，如论文[68,69,80,87,88,112,122,150]；其二，基于特征集合或者特征序列的ReID 算法，如论文[62,66,67,72–74,82,83]。

在基于特征向量的ReID 算法中，一张行人图像或者一段行人视频通常先被表示

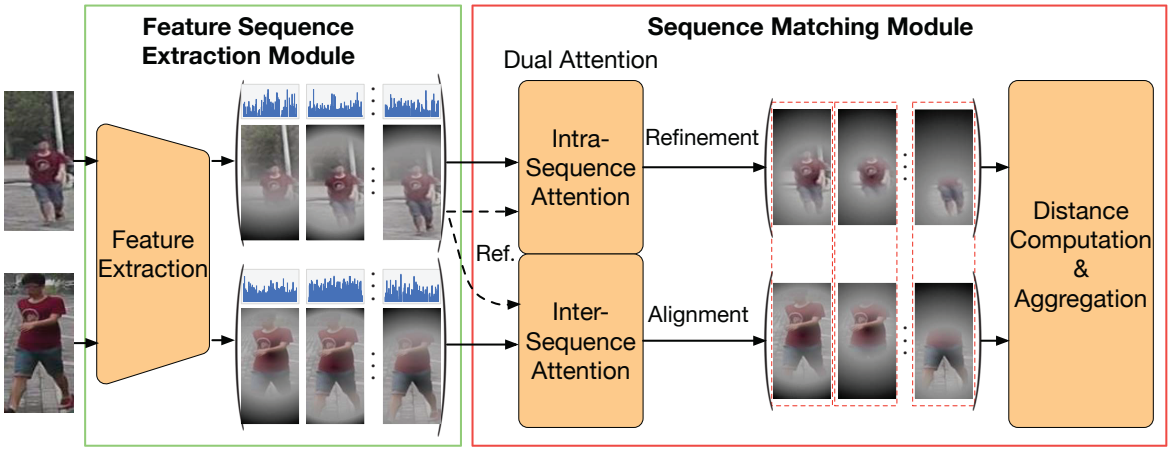


图 5-2　DuATM 框架示意图。

成单一特征向量，然后在此基础上进行度量学习。比如，在论文[3,18,25,27,31,33,44,47,57,59]中，研究人员先提取手工设计的行人局部特征，接着将其集成为单一全局特征向量；然后在此基础上，依照最小化类内差异、最大化类间距离的原则，进行距离度量学习。另外，近年来研究人员也开始尝试利用深度神经网络，直接从原始数据中学习高辨识度的行人特征嵌入。比如，论文[61,64]中，借助巧妙的网络结构设计，行人图像或者视频中的关键细节被自动发掘出来，并集成为行人的全局特征表示；在论文[49,56,58]中，借助于RNN 网络，循环复现的外观特征被刻画出来，并且通过时域池化的方式集成为行人的全局特征表示；在论文[1,79]中，为了提升神经网络所学得的特征嵌入的可辨识度，研究人员将基于样本对的对比损失函数，分别扩展为基于三元组或者四元组的距离度量损失函数。然而，尽管以上算法可以针对特定任务，自动地从数据中学习专用的特征向量表示，但是一旦最终特征中忽略掉了某些关键视觉细节，也会导致误匹配的发生。

与基于特征向量的ReID 算法不同，基于特征集合或者特征序列的ReID 算法，通过利用多种互补的特征向量或者时空信息，可以保留更多的视觉细节；而如何构造合理的局部对应关系结构，指导特征集合或者特征序列间的距离计算，是在这类算法下实现准确行人再识别的关键。比如，在论文[40,48]中，局部空间约束被用来指导基于图像块的特征集合的相似度计算；在论文[54]中，稠密的逐元素对的局部特征对间相似度被用来集成为时序特征序列对的整体相似度；在论文[66,67,72,74]中，通过引入图像块比较层到端到端的神经网络模型中，局部空间位置的对应结构可以直接从数据中学习出来，并且被应用到特征匹配中；在论文[82,83]中，人体的结构信息被用来指导基于语义的特征集合的对齐和匹配。尽管以上方法，可以利用启发式的局部

对应关系指导特征集合或者特征序列的距离度量，但是输入样本中剧烈的局部不对齐以及局部干扰问题，仍会使算法性能下降。

近些年，注意力机制被引入到包括序列匹配和特征表示学习在内的多种机器学习和模式识别任务中，例如论文[62,76,151–154]。在论文[151,154] 中，注意力机制被用来解决文本匹配中的字符对齐问题；在论文[76,152] 中，研究人员利用注意力机制对图像对进行反复比对，每次只关注图像的部分细节信息；在论文[62,153] 中，研究人员利用注意力机制从不定长视频数据中发掘关键细节，并分别针对人脸识别和行人再识别任务学习固定长度的特征表示。这些算法，要么是只单独利用了序列内注意力机制进行特征选择，要么是只利用了序列间注意力机制进行跨序列对齐。然而，只利用序列间注意力机制的话，尽管可以很好地解决局部不对齐问题，但是无法克服序列内部的干扰或者污染；另一方面，只利用序列内注意力机制的话，尽管可以处理局部污染问题，却没有办法很好地消除局部不对齐现象。

在我们的工作中，我们将双重注意力机制引入到了基于特征序列的行人再识别任务中。与以往工作不同，大部分工作借助于启发式的局部对应关系进行特征序列的比较，而我们尝试借助于双重注意力机制进行特征序列的匹配。具体来说，我们利用序列间注意力机制实现特征序列的语义对齐，同时利用序列内注意力机制实现序列内部的特征去噪精炼。

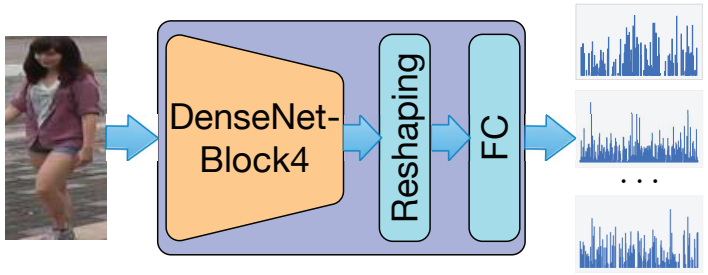
5.3 我们的方法

在本小节中，我们将对提出的端到端可训练的 ReID 模型DuATM 进行详细的介绍。如图5-2所示，模型主要由特征序列提取模块和特征序列匹配模块两个部分组成。

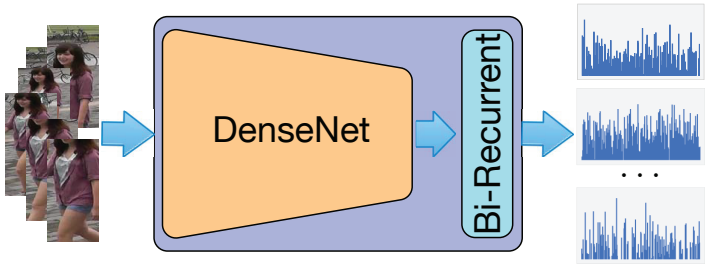
5.3.1 上下文敏感的特征序列提取模块

在 DuATM 中，我们采用 DenseNet-121[155] 作为特征序列提取模块的基础架构。由于在 DenseNet 中每一层的输入都是同一区块（Block）内前面所有层的输出的堆叠，因此底层的细节信息可以更容易的传递到网络末尾，从而丰富网络所提取的最终特征的信息量；而且在这种结构下，也更容易得到融合了底层细节与高层语义的层次化特征表示。以 DenseNet 为基础，我们分别为图像特征序列提取和视频特征序列提取设计了在结构上有轻微差异的子网络，如图5-3所示。

如图5-3(a) 所示，给定一张行人图像 $\mathscr{X} \in \mathbb{R}^{H\times W\times 3}$，我们首先利用 DenseNet-Block4（将 DenseNet 最后的最大池化层和全连接层移除掉后的网络）提取图像的



(a) 图像特征序列提取模块



(b) 视频特征序列提取模块

图 5-3　特征序列提取模块。

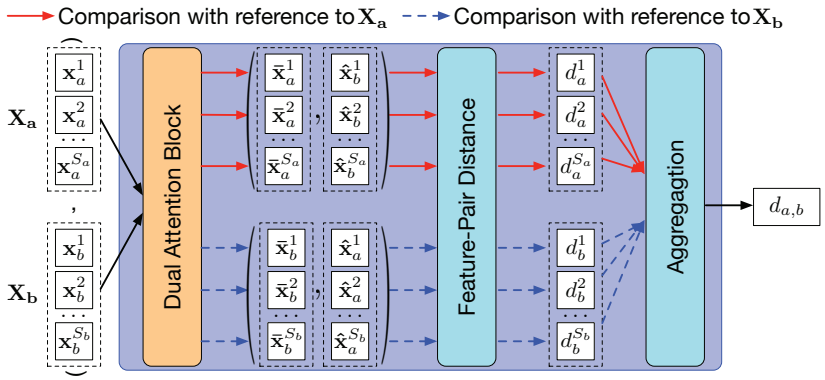
层次化特征，得到相应的特征图谱。假设特征图谱的数目为 D，则图谱上每个位置都可以看作一个 D 维的特征向量。由于特征图谱保留了空间位置信息，而且由于 DenseNet 足够深以至于图谱上每个位置对应的感受野都几乎覆盖了整个图像，因此每个 D 维的特征向量既包含了局部空间位置信息、也包含了全图的层次化语义上下文信息。我们将这些特征向按照在特征图谱中的位置，从上到下从左到右排列成一个特征序列，并让每个特征向量通过全连接层实现降维。最终我们可以得到行人图像所对应的紧致的特征序列表示。

如图5-3(b) 所示，给定一个长度为 T 的行人跟踪序列 $\mathscr{X} \in \mathbb{R}^{H\times W\times 3\times T}$，我们首先利用完整的 DenseNet 提取视频中每帧行人图像对应的特征向量，得到各向量之间相互独立的特征序列。由于与静态图像数据相比，视频数据不仅包含了更丰富的静态外观特征，同时还包含了丰富的动态信息，比如运动、光流等等。为此，我们将得到的特征序列传入到双向循环网络中，借助循环网络中信息可以在不同时间戳之间流动的特点，挖掘特征序列中蕴含的时间关联性。双向循环网络在每个时间戳上输出的特征向量既包含了局部时空中的行人外观信息、又包含了整个视频段中的时空上下文信息。最终我们将所有时间戳输出的特征向量组合在一起，构成行人视频所对应的特征序列表示。

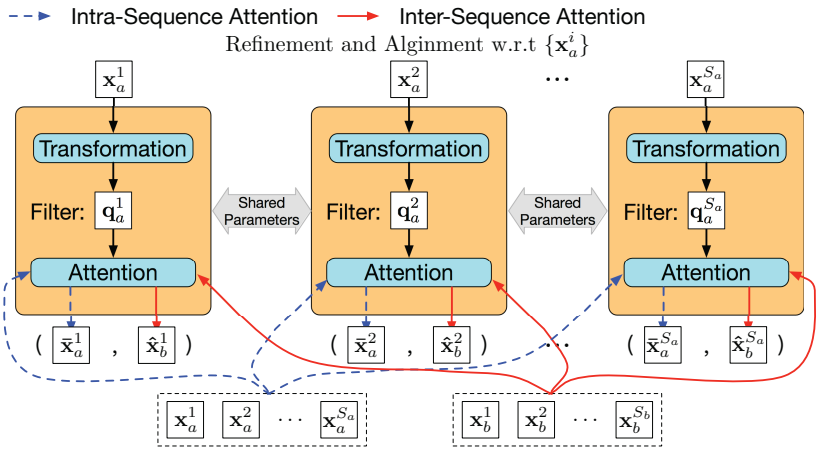
为了方便起见，我们将特征序列提取过程表示为 $\mathbf{X} = \mathscr{F}(\mathscr{X};\Theta_{\mathscr{F}})$，其中 $\mathbf{X}$ 为提

取得到的特征序列，它刻画了输入数据的空间细节信息或者时空细节信息；$\mathscr{F}(\cdot;\Theta_{\mathscr{F}})$ 代表特征序列提取子网络对应的参数化模型，网络所包含的参数为 $\Theta_{\mathscr{F}}$。更具体地，我们将长度为 S 的特征序列表示为 $\mathbf{X}=\{\mathbf{x}^i\in\mathbb{R}^D\}_{i=1}^{S}$，而且每个特征向量 $\mathbf{x}^i$ 在传入到匹配子网络之前都会先被进一步归一化成 ℓ_2 范数为 1。

5.3.2　基于双重注意力机制的特征序列匹配模块



(a) 特征序列匹配模块



(b) 双重注意力区块

图 5-4　特征序列匹配模块以及双重注意区块的结构和原理示意图。

特征序列匹配模块是 DuATM 模型最重要的组成部分。需要指出的是，由于不存在直接的监督信号来指导特征序列提取模块提取语义对齐的特征序列，因此序列匹配模块面对的首要挑战就是如何合理地计算未对齐特征序列对 $(\mathbf{X}_a,\mathbf{X}_b)$ 之间的距离或者相似度，其中 $\mathbf{X}_a=\{\mathbf{x}_a^i\}_{i=1}^{S_a}$，$\mathbf{X}_b=\{\mathbf{x}_b^j\}_{j=1}^{S_b}$。由于，受输入噪音或者不完美的特征提取器的影响，每个特征序列都可能一定程度上包含了已经被污染了的局部特征

向量。因此，利用直接对特征序列求均值将其转换成单一特征向量的方式，尽管可以避免（或者忽略）局部不对齐问题，但是却造成了局部特征信息的损失，而且还向最终特征中糅合了噪音。与此不同，我们提出先对特征序列进行去噪精炼和语义对齐，再计算已对齐的特征向量对的距离，最终将其汇总成特征序列对之间的距离。

由于前面设计的特征序列提取模块提取的中间特征，除了隐含特定的局部空间相关或者局部时空相关信息以外，还包含了丰富的语义相关上下文信息，因此我们可以借助这些上下文去实现局部去噪以及序列对齐。具体来说，我们将借助注意力机制，去发掘和利用序列元素的上下文信息。对于待匹配的特征序列对，如果将其中一个当做参考序列（Reference）②，将另一个当做注意力机制中的记忆序列（Menmory）③，那么对于参考序列中的任一元素，都可以利用注意力机制并结合其上下文信息，从记忆序列中选择出与其语义对齐的元素；而如果将参考序列自身同时也当做记忆序列，那么对于参考序列中的任一元素，也可以利用注意力机制并结合其上下文信息，从自身序列中选择出不含噪音的替代元素；从而实现了相对于参考序列的序列间的语义对齐，以及参考序列内的去噪精炼。此外，为了使去噪更彻底，同时为了保证后续计算的距离度量的对称性，我们将参考序列和记忆序列交换位置，并再次重复以上操作。

为了更清楚地阐述双重注意力机制及其实现细节，我们给出了示意图5-4进行说明。图5-4(a) 展示了基于双重注意力机制的特征序列匹配模块的实现框架，给定待匹配特征序列，在模块内部可以实现自动地序列内特征去噪精炼和序列间语义对齐（图5-4(b)）。经过去噪精炼和语义对齐的特征序列中的特征向量对，可以直接用来计算特征对间的局部距离度量，并最终汇总成序列对间的整体距离度量。

5.3.2.1 双重注意力区块（Dual Attention Block）

双重注意力区块由一个转换层（Transformation Layer）和一个注意力层（Attention Layer）组成。其中，转换层将每一个输入特征向量投影到滤波器空间，生成一个与输入特征相关的注意力滤波器；而注意力层将会利用滤波器生成一系列与记忆序列元素一一对应的注意力权重。为了不失一般地阐述整个过程，下面我们将以去噪精炼和对齐了的特征向量对 $(\bar{\mathbf{x}}_a^i, \hat{\mathbf{x}}_b^i)$ 的生成过程为例，参考图5-4(b)，对序列去噪精炼和对齐原理进行详细解释。要指出的是，该过程是以序列 $\mathbf{X}_a$ 中的向量元素 $\mathbf{x}_a^i$ 为参

② 参考序列内的元素将会被逐一去噪精炼，同时也将会被当做另一个序列的对齐标准。
③ 记忆序列中的元素将会被按照参考序列元素位置进行重新组合和排序。

考，进行的特征去噪精炼和特征对对齐过程。

- 首先，对于输入的参考特征向量，利用转换层生成对应的注意力滤波器，过程如下：

$$\mathbf{q}_a^i = ReLU(BN(\mathbf{W}\mathbf{x}_a^i + \mathbf{b})), \tag{5-1}$$

其中，$\mathbf{W}$ 和 $\mathbf{b}$ 分别是线性全连接层的权重矩阵和偏置向量，BN 和 ReLU 分别代表批归一化（Batch Normalization，BN）操作[156] 和线性整流单元（Rectified Linear Unit，ReLU）函数。

- 然后，利用注意力层，分别生成用于特征序列内部去噪精炼的注意力权重、以及用于特征序列间语义对齐的注意力权重，过程分别如下所示：

$$\begin{aligned} \bar{e}_a^{i,m} &= \langle \mathbf{q}_a^i, \mathbf{x}_a^m \rangle, \\ \hat{e}_b^{i,n} &= \langle \mathbf{q}_a^i, \mathbf{x}_b^n \rangle, \end{aligned} \tag{5-2}$$

其中，$\langle \cdot, \cdot \rangle$ 表示求内积的操作。

- 最终，已经被去噪精炼以及语义对齐了的特征对 $(\bar{\mathbf{x}}_a^i, \hat{\mathbf{x}}_b^i)$，可以通过分别对相应原始特征序列中特征向量元素进行线性组合得到，而组合系数即为归一化后的注意力权重。过程如下：

$$\begin{aligned} \bar{\mathbf{x}}_a^i &= \sum_{m=1}^{S_a} \sigma(\bar{e}_a^{i,m}) \mathbf{x}_a^m, \\ \hat{\mathbf{x}}_b^i &= \sum_{n=1}^{S_b} \sigma(\hat{e}_b^{i,n}) \mathbf{x}_b^n, \end{aligned} \tag{5-3}$$

其中，$\sigma(\cdot)$ 是用来实现权重归一化的 Softmax 函数，定义为 $\sigma(t_j) = \frac{\exp(t_j)}{\sum_{j=1}^{S} \exp(t_j)}$，其中 $\mathbf{t} \in \mathbb{R}^S$。

按照公式5-1至公式5-3所描述的过程，以输入序列中任一特征向量为参考，都可以生成相应的去噪精炼的、语义上可比较的特征向量对。

5.3.2.2 距离计算和汇总

得益于双重注意力区块输出的已经完成了去噪精炼和语义对齐了的特征序列对，我们可以直接在局部上计算两两特征向量之间的距离，并最终将其汇总成特征序列对之间的整体距离。

实际上，在 DuATM 模型中，我们采用了双向的双重注意力机制进行序列匹配，即双重注意力过程分别以序列 $\{\mathbf{x}_a^i\}$ 和 $\{\mathbf{x}_b^j\}$ 为参考序列执行两次，并且我们结合了两次的距离度量结果作为最终的特征序列距离。具体来说，我们先利用欧氏距离计算每一对具有可比性的特征向量的距离，过程如下：

$$\begin{aligned} d_a^i &= ||\bar{\mathbf{x}}_a^i - \hat{\mathbf{x}}_b^i||_2, \quad i = 1, \cdots, S_a, \\ d_b^j &= ||\bar{\mathbf{x}}_b^j - \hat{\mathbf{x}}_a^j||_2, \quad j = 1, \cdots, S_b. \end{aligned} \tag{5-4}$$

接着，我们利用计算均值的方式将这些局部距离汇总成序列对 $(\mathbf{X}_a, \mathbf{X}_b)$ 的整体距离，过程如下：

$$||\mathbf{X}_a - \mathbf{X}_b||_{\mathscr{M}} = \frac{1}{2S_a}\sum_{i=1}^{S_a} d_a^i + \frac{1}{2S_b}\sum_{j=1}^{S_b} d_b^j, \tag{5-5}$$

其中 $||\mathbf{X}_a - \mathbf{X}_b||_{\mathscr{M}}$ 表示由特征序列匹配模块所定义的序列距离。为了简便起见，我们把包括 $\mathbf{W}$、$\mathbf{b}$、以及BN 层中参数在内所有序列匹配模块中的参数表示为 $\Theta_{\mathscr{M}}$。

5.3.3 损失函数

为了学习到能够解决ReID 任务，并且在未见数据集上具有良好泛化能力的模型，我们将 DuATM 构造成含有三分支的孪生网络，并借助于 Triplet Loss 进行训练优化。训练阶段，网络模型结构如图5-5所示。同时，为了使模型学习到的中间特征更加紧致、更加鲁棒、也更加有辨识度，我们引入并联合使用了另外两个辅助损失函数，分别是 De-Correlation Loss 和 Cross Entropy Loss。因此，网络训练的最终损失函数定义为：

$$\ell = \ell^{(0)}(\mathscr{X}, \Theta_{\mathscr{F}}, \Theta_{\mathscr{M}}) + \lambda_1 \ell^{(1)}(\mathscr{X}, \Theta_{\mathscr{F}}) + \lambda_2 \ell^{(2)}(\mathscr{X}, \Theta_{\mathscr{F}}, \boldsymbol{\theta}), \tag{5-6}$$

其中，$\lambda_1 > 0$ 和 $\lambda_2 > 0$ 分别是两个平衡系数。

5.3.3.1 Triplet Loss

我们在模型训练中使用 Triplet Loss 的目的是，使网络学习到的正样本对的距离尽量小于负样本对的距离。给定由行人图像或者行人跟踪视频组成样本三元组，特征序列提取模块会利用含有三个分支的孪生子网络分别提取三元组的空间或者时空上下文相关的特征序列；而序列匹配模块会利用含有两个分支的孪生子网络分别计算正样本对和负样本对内部的距离度量。

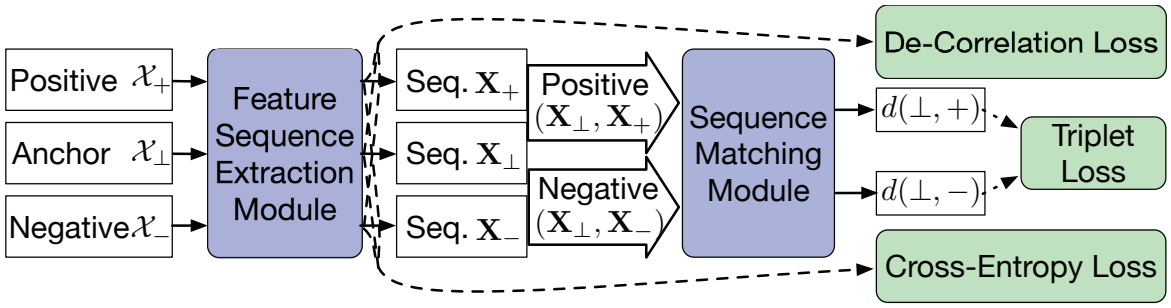


图 5-5 训练阶段 DuATM 模型结构示意图。

如果将输入的三元组样本表示为 $\mathscr{X} = (\mathscr{X}_\perp, \mathscr{X}_+, \mathscr{X}_-)$。为了促使网络，对正样本对的预测距离比负样本对的预测距离小间隔 γ，我们可以将 Triplet Loss 定义为：

$$\ell^{(0)}(\mathscr{X}, \Theta_{\mathscr{F}}, \Theta_{\mathscr{M}}) = \max\{0, \gamma + ||\mathscr{F}(\mathscr{X}_\perp) - \mathscr{F}(\mathscr{X}_+)||_{\mathscr{M}} - ||\mathscr{F}(\mathscr{X}_\perp), \mathscr{F}(\mathscr{X}_-)||_{\mathscr{M}}\}, \tag{5-7}$$

其中，往往会要求 $\gamma > 0$，比如依据经验在我们的实验中将其设置为 0.2。

5.3.3.2 De-Correlation Loss

在论文[157] 中，研究人员为了减小深度神经网络的过拟合问题，提取了一种去相关表示（De-Correlating Representation）。受此启发，在我们的工作中，为了降低特征序列内部特征向量之间的相关性，使特征表示更加的紧致，我们提出了一种相似却又不同的去相关损失函数（De-Correlation Loss）。具体来说，我们为特征提取模块添加了以下关于特征序列内部相关矩阵的强制约束条件：

$$\ell^{(1)}(\mathscr{X}, \Theta_{\mathscr{F}}) = \frac{1}{N^2}||\mathbf{I} - \mathscr{F}(\mathscr{X})^T \mathscr{F}(\mathscr{X})||_F^2, \tag{5-8}$$

其中，$\mathbf{I}$ 表示单位矩阵，N 为特征序列中特征向量元素的个数。

5.3.3.3 Cross Entropy Loss

为了使模型学习到更加信息丰富以及更加鲁棒的特征序列，我们在模型训练时同时引入了 Cross Entropy Loss，并且在使用过程中添加了数据扩充策略（Data Augmentation）。具体来说，在使用原始数据进行分类训练的同时，我们利用插值的方式为具有同一标签的特征创造更多的同类虚拟数据。

假设令 $\mathbf{X} = \mathscr{F}(\mathscr{X})$ 为网络中间所学习到的特征序列，我们首先可以将特征序列

内元素组合成单一特征向量，过程如下：

$$\mathbf{z} = \sum_{i=1}^{S} \omega_i \mathbf{x}_i, \tag{5-9}$$

其中，$\sum_i \omega_i = 1$，并且 $\omega_i \geq 0$。然后将融合得到的特征向量传入到一个全连接层，实现分类打分，并利用 Cross Entropy Loss 对结果进行约束，过程如下：

$$\ell^{(2)}(\mathscr{X}, \Theta_{\mathscr{F}}, \boldsymbol{\theta}) = -\ln \sigma(\mathbf{w}_c \mathbf{z} + b_c), \tag{5-10}$$

其中，c 表示输入样本 $\mathscr{X}$ 所对应的类别标签，$\{\mathbf{w}_c, b_c\}$ 分别表示全连接层的权重矩阵和偏置向量的第 c 行元素，$\boldsymbol{\theta}$ 包含了新增的全连接层中的所有参数。值得注意的是，与以往直接简单地对特征序列求均值，得到单一特征向量 $\mathbf{z}$ 的方式不同；我们随机生成组合系数 $\omega_i \in [0,1]$，并以概率 $p > 0$ 随机地将其中一部分置成 0，同时保持保证 $\sum_i \omega_i = 1$，这样就可以通过凸组合策略将特征序列元素合成单一特征向量，并通过引入一定程度的噪音来进一步降低特征向量间的依赖性。这种组合方式可以看成论文[158] 中提出的基于插值方法生成新数据的简化版本。

5.4 实验结果及分析

为了对 DuATM 的模型性能进行全面而详细地评估，我们在 Market1501、DukeMTMC-reID、以及 MARS 等三个大规模行人再识别数据集上做了大量对比实验。

5.4.1 数据集和实验设置

5.4.1.1 数据集和评价指标

对于数据集 Market1501，我们按照论文[3] 提出的方式，将整个数据集分成没有类别重叠的训练数据和测试数据两部分，这两部分分别包含了 12,936 和 19,732 张行人图像，同时我们也采用了单图匹配模式（Single-Quey）；对于数据集 DukeMTMC-reID，我们将出现在多个摄像头下的 1,404 个行人平均分成训练数据和测试数据两部分，具体来说，数据集中包含了 2,228 张行人图像组成的PSet、17,661 张行人图像组成的GSet、以及 16,522 张行人图像组成的训练集；对于数据集 MARS，我们将 625 个行人的 8,298 个跟踪视频用作训练，将 636 个行人的 12,180 个跟踪视频用于测试。

在本次实验中，我们采用了ReID 任务中常用的性能评价指标CMC 和mAP。为了计算指标幅值，我们利用 Python 语言复现了论文[101] 提供的性能评估程序。

5.4.1.2 算法实现细节

实验中，我们利用在 ImageNet 上预训练好的 DenseNet-121 模型的参数值，去初始化DuATM 模型中的 DenseNet 部分，并使用随机梯度下降法 (Stochastic Gradient Descent，SGD) 进行模型的训练。具体来说，为了防止在训练之初，含有噪音的反向传播误差对 DenseNet 部分初始化参数的破坏，我们先将 DenseNet 部分参数固定之后对模型训练 100 个 Epoch，接着再对整个网络训练 200 个 Epoch。在训练过程中，模型的学习率初始化为 0.01，并在最后 50 个训练 Epoch 内变为 0.001。

在利用 Triplet Loss 进行 DuATM 模型训练时，我们所面临的一个最为严峻的问题是，训练数据中的正样本对远远低于负样本对，正负样本不平衡。为了缓解这个问题，我们采用了论文[1,159] 中所提出的难分类三元组挖掘 (Hard Triplet Mining，HTM) 策略来生成由三元组构成的小批量集合（Mini-Batch）。具体来讲，每一个小批量集合训练数据，包含了 P 个行人，每个行人拥有 V 张图像或者 V 段跟踪序列；所有这些数据都会被当做参考样本（Anchor Point），同时对于每个参考样本会从小数据集合中选择最难正确分类的正样本和负样本，即距离参考样本最远的正样本和距离参考样本最近的负样本；由此，构成一个三元组，最终生成一个三元组构成的小批量集合训练数据。在我们的实验中，对于图片尺寸为 256×128 的行人图像数据集，我们设置 $(P=10,V=4)$；而对于图片尺寸为 128×64 的行人跟踪序列数据集，我们设置 $(P=7,V=3)$。另外，我们借鉴了常用的深度模型训练经验，利用随机旋转、随机裁剪等方式实现训练数据的扩充，并且利用随机截取长度为 16 的连续视频子片段的方式实现视频样本的扩充，如论文[56] 所述。

表 5-1 在数据集 Market1501 上，DuATM 与基准模型及不同损失函数下模型性能对比结果。* 在这组实验中，我们将损失函数中的超参按照参数分析中的结果调节到最优配置。** 在这组实验中，我们在性能评估阶段采取了数据扩充的策略。

方法 & 损失函数	R1	R5	R20	mAP
AvePool+$\ell^{(0)}$	74.20	89.67	95.58	56.88
DuATM+$\ell^{(0)}$	79.66	91.15	96.73	63.46
DuATM+$\ell^{(0)}$+$\ell^{(1)}$	81.83	92.46	97.33	65.21
DuATM+$\ell^{(0)}$+$\ell^{(2)}$	87.50	95.37	98.01	70.02
DuATM+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	88.75	95.78	98.46	70.46
DuATM*+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	89.96	96.53	98.72	75.22
DuATM**+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	**91.42**	**97.09**	**98.96**	**76.62**

实验中，从行人图像或者行人跟踪视频中提取的特征向量的维度 D，被默认设置为 256。此外，损失函数中的超参（如 λ_1、λ_2、以及污染度 p），分别被默认设置为 $\lambda_1 = 0.1$、$\lambda_2 = 0.5$、以及 $p = 0.2$。在进行参数分析时，这些超参会被调整到最优值。在性能评估阶段，当与同期最优算法进行性能对比时，我们省去了数据扩充的策略，并且在行人视频数据集上，我们将视频子片段长度设为 64④。本章中，所有的实验都利用 PyTorch 实现，硬件为两块 Nvidia 公司的 Titan-X 型号显卡。

5.4.2 DuATM 模型的性能评估

5.4.2.1 DuATM 中的损失函数评估

表 5-2 在数据集 DukeMTMC-reID 上，DuATM 与基准模型及不同损失函数下模型性能对比结果。* 在这组实验中，我们将损失函数中的超参按照参数分析中的结果调节到最优配置。** 在这组实验中，我们在性能评估阶段采取了数据扩充的策略。

方法 & 损失函数	R1	R5	R20	mAP
AvePool+$\ell^{(0)}$	64.05	79.44	87.52	43.79
DuATM+$\ell^{(0)}$	68.40	81.73	89.77	48.65
DuATM+$\ell^{(0)}$+$\ell^{(1)}$	69.17	82.23	89.36	49.48
DuATM+$\ell^{(0)}$+$\ell^{(2)}$	79.40	90.04	94.25	61.55
DuATM+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	81.06	**91.11**	95.02	62.27
DuATM*+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	81.46	90.75	95.11	63.14
DuATM**+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	**81.82**	90.17	**95.38**	**64.58**

为了更好地评估不同损失函数以及双重注意力区块在 DuATM 模型中的作用，我们按照不同设置进行了大量对比实验，具体包括：a) DuATM+$\ell^{(0)}$、b) DuATM+$\ell^{(0)}$+$\ell^{(1)}$、c) DuATM+$\ell^{(0)}$+$\ell^{(2)}$,、和 d) DuATM+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$。需要注意的是，由于我们的 DuATM 模型以 DenseNet 为基础，因此为了公平地反映出模型的优越性，我们按以下方式构造基准模型：首先，利用 DenseNet 提取输入图像的特征序列，然后通过求序列平均值的方式将其融合成单一整体特征向量，最后在欧式空间里计算向量的距离作为行人的距离度量。基准模型同样用 Triplet Loss 训练，并且采用同样的训练策略，我们将基准模型的实验结果表示为 AvePool+$\ell^{(0)}$。在不同基准数据集上的实验结果都分别被展示在表格5-1至表格5-3中。由实验结果我们发现，在三个基准数据集上，模型

④ 如果原始行人跟踪序列长度不足 64，我们利用循环采样的方式进行视频长度扩充。

表 5-3 在数据集 MARS 上，DuATM 与基准模型及不同损失函数下模型性能对比结果。* 在这组实验中，我们将损失函数中的超参按照参数分析中的结果调节到最优配置。** 在这组实验中，我们在性能评估阶段采取了数据扩充的策略。

方法 & 损失函数	R1	R5	R20	mAP
AvePool+$\ell^{(0)}$	65.45	81.92	90.10	47.26
DuATM+$\ell^{(0)}$	66.36	83.13	90.40	48.44
DuATM+$\ell^{(0)}$+$\ell^{(1)}$	66.52	83.78	91.21	49.07
DuATM+$\ell^{(0)}$+$\ell^{(2)}$	73.74	87.73	93.84	56.36
DuATM+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	74.43	89.08	94.13	58.19
DuATM*+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	76.36	90.10	95.30	58.96
DuATM**+$\ell^{(0)}$+$\ell^{(1)}$+$\ell^{(2)}$	**78.74**	**90.86**	**95.76**	**62.26**

DuATM+$\ell^{(0)}$ 的性能都要优于模型 AvePool+$\ell^{(0)}$ 的性能。实验结果证明了在 DuATM 模型中引入双重注意力区块的有效性，即利用双重注意力机制进行上下文敏感特征序列的匹配要比基于单一整体特征向量的匹配效果好。另外，当我们分别联合 De-Correlation Loss 或者 Cross Entropy Loss 进行模型训练优化时，DuATM 的模型性能仍会有不同程度的提升。然而，由于 De-Correlation Loss 并没有为模型的训练引入额外的监督信息，因此相比模型 DuATM+$\ell^{(0)}$，模型 DuATM+$\ell^{(0)}$+$\ell^{(1)}$ 的性能提升并不十分明显。另外，有趣的是，当在模型训练中引入 Cross Entropy Loss 时，模型性能的提升十分显著，这可能要归功于行人身份标签所引入的额外的模型训练监督信号。最终，当联合了所有的损失函数进行训练时，DuATM 在ReID 任务上的准确度会进一步提升。

5.4.2.2 DuATM 的模型简化测试（Ablation Study）

表 5-4 在数据集 Market1501 上，DuATM 的模型简化测试结果。

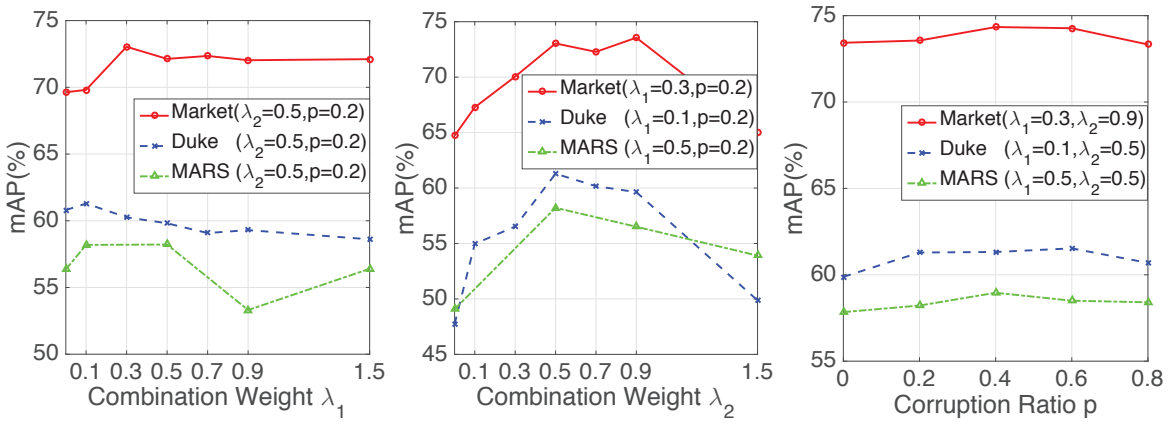
方法 & 损失函数	R1	R5	R20	mAP
AvePool+$\ell^{(0)}$	74.20	89.67	95.58	56.88
Intra+$\ell^{(0)}$	78.78	90.69	96.73	61.76
Inter+$\ell^{(0)}$	72.36	87.74	95.19	53.91
DuATM+$\ell^{(0)}$	79.66	91.15	96.73	63.46

为了更好地验证不同注意力机制在模型 DuATM 中的有效性，我们在基准数据

集 Market1501 上，分别对序列内注意力机制和序列间注意力机制进行了对比性能评估，实验结果分别表示为 Intra+$\ell^{(0)}$ 和 Inter+$\ell^{(0)}$，并被展示在表格5-4中。需要指出的是，表格5-4中的 DuATM+$\ell^{(0)}$ 其实就是 Intra+Inter+$\ell^{(0)}$。根据实验我们发现，同时采用两种注意力（即双重注意力）比单纯使用一种注意力（如序列内或者序列间注意力）的模型再识别准确度高。这也再次证明了双重注意力在 DuATM 中的重要性。

5.4.2.3 DuATM 中的参数评估

我们发现，在 DuATM 的整体损失函数中存在着两个平衡系数参数 λ_1 和 λ_2；同时在 Cross Entropy Loss 中也存在着一个污染度参数 p，控制着被合成新数据中引入的噪音程度。为了评估不同参数对模型训练的影响，我们在三个基准数据集上做了大量对比实验，实验结果被展示在图5-6中。我们采用了控制变量法来研究每个参数，即每组实验只改变被考量的参数的值，而使其他参数保持一致。



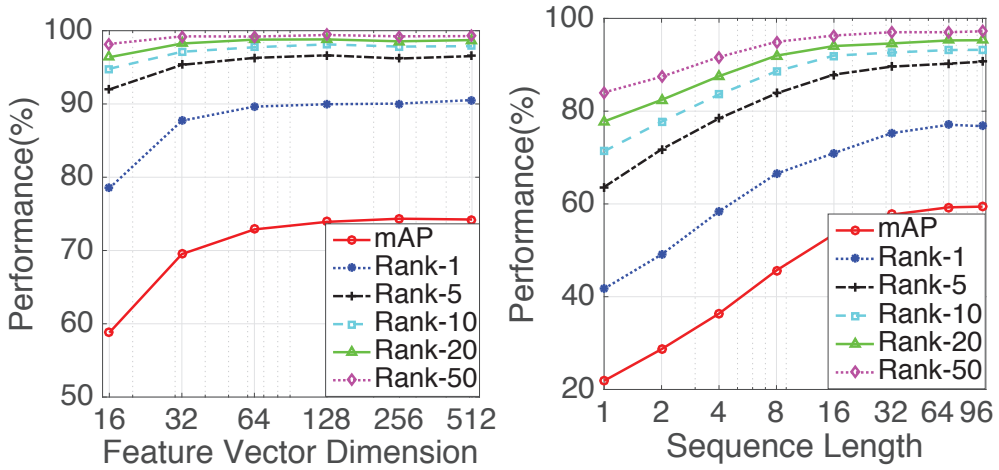
(a) (b) (c)

图 5-6 损失函数中参数设置对模型性能影响的评估实验。

通过观察实验结果，我们可以得到三个结论：其一，当设置适当的 λ_1 值时，可以使学习到的特征序列更紧致，从而带来匹配性能的提升；然而，当 λ_1 被设置的过大时，会因此破坏了特征序列内部的上下文关系，从而导致性能的轻微下降。其二，当设置适当的 λ_1 值时，会因引入了额外的监督信息，而使所学特征可辨识度更高；然而，当 λ_1 被设置的过大时，会容易导致模型的过拟合，使模型在测试数据上性能有所下降。其三，模型的性能对参数 p 不敏感。同时，我们也发现了三个数据集 Market1501、DukeMTMC-reID、和 MARS 上对应的较好的参数配置 $(\lambda_1, \lambda_2, p)$ 分别为 $(0.3, 0.9, 0.4)$、$(0.1, 0.5, 0.6)$、和 $(0.5, 0.5, 0.4)$，我们将对应的ReID 结果展示在了

表格5-1至表格5-3末尾两行。

除此之外，我们分别在行人图像数据集 Market1501 和行人跟踪视频数据集 MARS 上，对特征向量维度 D 和视频序列长度 T 的取值做了简要对比分析。实验结果被展示在图5-7中。我们发现，对于基于图像的行人再识别，当我们将每个行人表示成长度为 $N \times T$ 的特征序列时，即使每个特征向量的维数很低，整个序列表示也会包含足够的行人身份判别信息。比如，即使特征向量维度降到了 16 或者 32，模型在ReID 上 Rank-1 的准确度仍可以保持在 78.50% 或者 87.71%。对于基于视频的行人再识别，由于特征序列的长度直接由视频的长度所决定，因此长的视频段很可能包含更丰富的不同时间点上行人的视觉信息，因此会带来更高的再识别准确度。比如，当视频段长度从 1 增加到 96 时，ReID 的 mAP 得分从 21.87% 提升到了 59.42%。



(a) (b)

图 5-7 特征向量维数和视频段长度对模型性能影响的评估实验。

5.4.3 DuATM 与其他 ReID 模型的性能对比

5.4.3.1 DuATM 与其他注意力模型的性能对比

为了验证我们所设计的双重注意机制的优越性，我们将 DuATM 模型与包括 CAN[76]、HP-Net[84]、ST-RNN[62]、和 QAN[61] 在内的多种注意力模型进行了对比，并将结果汇总在表格5-5中。在被对比的模型中，研究人员尝试借助注意力机制进行图像或者视频中关键特征的筛选，比如发掘显著性模式、提取关键局部特征、或者自适应地融合局部特征集合。与此不同的是，我们的算法会在特征提取阶段保留所有的细节信息，并利用双重注意力机制在匹配阶段实现局部模式的去噪精炼以及模

表 5-5　DuATM 与其他注意力模型的性能对比结果。

数据集	方法	R1	mAP	文献
Market1501	CAN	48.24	24.43	2017 TIP[76]
	HP-Net	76.90	-	2017 ICCV[84]
	DuATM	**91.42**	**76.62**	This paper
MARS	ST-RNN	70.60	50.70	2017 CVPR[62]
	QAN	73.74	51.70	2017 CVPR[61]
	DuATM	**78.74**	**62.26**	This paper

式对的语义对齐。因此，我们的模型可以进行更合理地图像或者视频对比，并取得优越的性能。

5.4.3.2　DuATM 与其他特征集合或特征序列匹配模型的性能对比

表 5-6　DuATM 与其他特征集合或特征序列匹配模型的性能对比结果。

数据集	方法	R1	mAP	文献
Market1501	SCSP	51.90	26.35	2016 CVPR[48]
	SpindleNet	76.90	-	2017 CVPR[82]
	DLPAR	81.00	63.40	2017 ICCV[83]
	DRL-PL	88.20	69.30	2017 Arxiv[160]
	DuATM	**91.42**	**76.62**	This paper

在表格5-6中，我们将 DuATM 模型与论文[48,82,83,160] 中多种不同的特征集合或特征序列匹配模型进行了对比。由于 DuATM 模型借助于双重注意力机制，不仅可以有效地推理出序列间的局部对应关系结构，还可以自动去除序列内所包含的局部干扰，因此我们的方法在数据集 Market1501 上，比基于身体部分的模型（如，SpindleNet[82]、DRL-PL[160]）和基于稠密匹配的方法（如 SCSP[48]）取得的再识别准确率都要高。

5.4.3.3　DuATM 与同期最先进 ReID 算法的性能对比

在表格5-7至表格5-9中，我们分别将 DuATM 模型与同期最好的ReID 算法在 Market1501、DukeMTMC-reID、和 MARS 三个数据集上做了对比。实验证明 DuATM

表 5-7 在数据集 Market1501 上，DuATM 与同期其他先进 ReID 算法的性能对比。

方法	R1	R5	mAP	文献
BOW	44.42	63.90	20.76	2015 ICCV[3]
LDNS	61.02	-	35.68	2016 CVPR[46]
Re-Rank	77.11	-	63.63	2017 CVPR[113]
SSM	82.21	-	68.80	2017 CVPR[114]
S-LSTM	61.60	-	35.30	2016 ECCV[49]
G-CNN	65.88	-	39.55	2016 ECCV[72]
CRAFT	68.70	-	42.30	2017 TPAMI[121]
P2S	70.72	-	44.27	2017 CVPR[81]
CADL	73.84	-	47.11	2017 CVPR[161]
USG-GAN	78.06	-	56.23	2017 ICCV[99]
LDCAF	80.31	-	57.53	2017 CVPR[80]
SVDNet	82.30	92.30	62.10	2017 ICCV[87]
TriNet	84.92	**94.21**	69.14	2017 Arxiv[1]
JLML	85.10	-	65.50	2017 IJCAI[92]
DML	87.73	-	68.83	2017 Arxiv[162]
REDA	87.08	-	71.31	2017 Arxiv[163]
DarkRank	**89.80**	-	**74.30**	2017 Arxiv[164]
DuATM	**91.42**	**97.09**	**76.62**	This paper

表 5-8 在数据集 DukeMTMC-reID 上，DuATM 与同期其他先进 ReID 算法的性能对比。

方法	R1	R5	mAP	文献
BOW	25.13	-	12.17	2015 ICCV[3]
LOMO	30.75	-	17.04	2015 CVPR[18]
USG-GAN	67.68	-	47.13	2017 ICCV[99]
OIM	68.10	-	-	2017 CVPR[90]
APR	70.69	-	51.88	2017 Arxiv[165]
SVDNet	76.70	**86.40**	56.80	2017 ICCV[87]
DPFL	79.20	-	60.60	2017 ICCVW[166]
REDA	**79.31**	-	**62.44**	2017 Arxiv[163]
DuATM	**81.82**	**90.17**	**64.58**	This paper

模型在三个数据集都取得了同期最高的ReID 准确度，从而再次验证了基于上下文敏感特征序列和双重注意力机制匹配算法的有效性。具体来说，在行人图像数据集 Market1501 和 DukeMTMC-reID 上，DuATM 模型超过所有同期的分步优化模型以及端到端优化模型，并分别取得了 91.24% 和 81.37% 的 Rank-1 准确度；在行人跟踪序列数据集 MARS 上，DuATM 模型的性能仍然要比列表中的大多数算法好，另外，如果我们采用与论文[1] 中一样的 256×128 的输入图像尺寸，DuATM 模型依旧可以超过所有同期最优模型。

表 5-9　在数据集 MARS 上，DuATM 与同期其他先进 ReID 算法的性能对比。DuATM*: 采用与论文[1] 一样的图像尺寸 256×128 重新训练模型。

方法	R1	R5	mAP	文献
SMP	23.59	35.81	10.54	2017 ICCV[105]
BOW	30.60	46.20	15.50	2015 ICCV[3]
DGM	36.80	54.00	21.30	2017 ICCV[116]
Re-Rank	73.93	-	68.45	2017 CVPR[113]
IDE	65.10	81.10	45.60	2016 ECCV[101]
LDCAF	71.77	86.57	56.50	2017 CVPR[80]
TriNet	**79.80**	**91.36**	**67.70**	2017 Arxiv[1]
DuATM	78.74	90.86	62.26	This paper
DuATM*	**81.16**	**92.47**	**67.73**	This paper

5.4.4　双重注意力机制的可视化

为了更好地理解 DuATM 模型中双重注意力机制的作用，我们在图5-8中展示了一些模型输出的中间结果的可视化图像。由于特征提取模块所提取的序列中的所有特征向量都包含丰富的上下文信息，因此以每个特征向量为参照，序列内部的注意力机制可以将关注点放在图像或者视频内部与其语义相关的身体部分或者步态信息上，从而借助于这些上下文来消除自身的噪音；同时序列间的注意力机制可以将注意力放在图像或者视频之间与其语义一致的相关的身体部分或者步态信息上，从而借助于这些上下文来进行局部对齐。即使如图5-8(c) 所示，参照的特征向量采集于干扰区域，双重注意力机制也可以进行合理地特征去噪精炼和对齐。

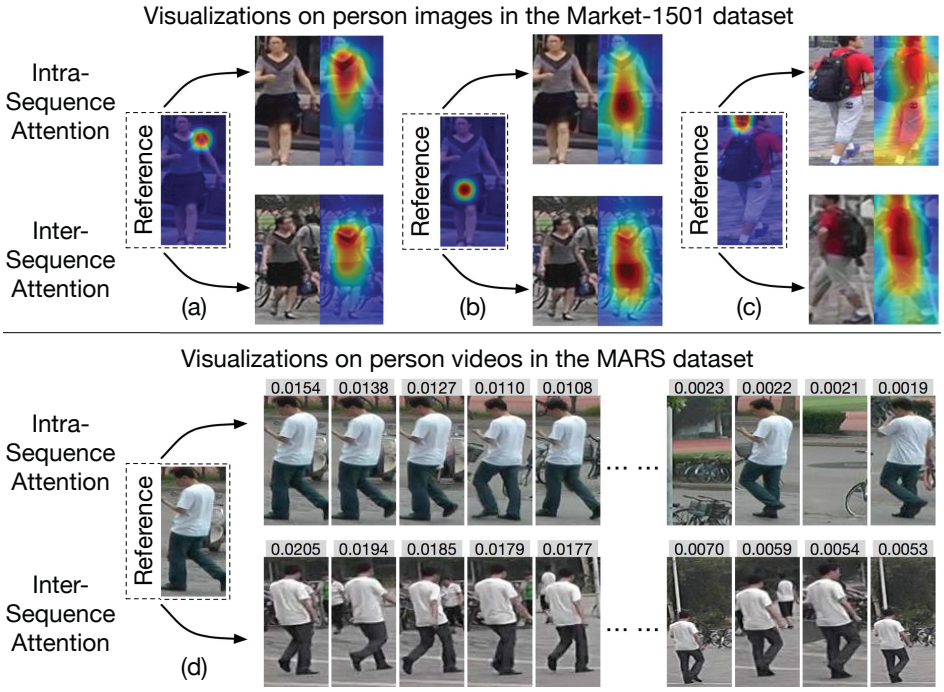


图 5-8 可视化地展示序列内和序列间注意力权重。

5.5 本章小结

本章分别从引言、相关工作、上下文敏感的特征序列提取模块、基于双重注意力机制的特征序列匹配模块、以及相关实验结果和分析等方面，对我们提出的基于上下文敏感特征序列及双重注意力匹配的行人再识别网络进行详细介绍。我们为解决ReID 问题提出了一种称为DuATM 的端到端的深度模型，该模型可以自动地从样本中学习上下文敏感的特征序列表示，并且同时利用双重注意力机制自适应地进行序列匹配。DuATM 的核心部件是双重注意力区块，该部分实现了特征序列对的自动去噪精炼和语义对齐。另外，我们联合使用 Triplet Loss、De-Correlation Loss、和 Cross Entropy Loss 三种损失函数进行模型训练。在大规模行人图像和行人跟踪视频数据集上的实验结果，证明了 DuATM 模型的有效性和优越性。

第六章 总结与展望

6.1 工作总结

本文首先从ReID 的研究背景、研究进展、存在的主要挑战、以及学术界常用的算法性能指标和常用的公开数据集等方面，对本课题的研究内容做了简单介绍。接着我们回顾了现阶段行人再识别中的相关算法，从特征表示方法、特征匹配算法、深度神经网络算法等方面对相关算法进行总结和分析，并进一步阐明本课题的研究动机。最终，本文在国内外大量研究成果的基础上，结合ReID 所面临的主要挑战以及我们对ReID 问题的认识，分别从小数据集上的度量学习模型泛化能力、基于手工特征设计和多特征融合的分步处理模型、基于特征序列提取和序列匹配的端到端处理模型这三个角度，分别提出了不同的改进的ReID 算法。本文的主要工作可以总结如下：

（1）提出利用正则化的度量学习方法来增强小数据集上ReID 模型的泛化能力

众所周知，度量学习的研究在 ReID 技术发展过程中扮演着重要的角色。然而，受某些应用场景的限制，研究人员往往无法获取充足的标记样本进行模型训练和学习，从而导致 ReID 算法的泛化能力较弱。为此，我们从限制模型复杂度的角度，提出利用正则化的度量学习算法实现 ReID 中的特征距离度量，从而提升小数据集上模型的泛化能力。具体来讲，我们分别从马氏距离学习、对称投影学习、以及非对称投影学习三个不同的角度理解度量学习模型，并构造了四种不同的正则化度量学习算法实例来实现 ReID。在数据集 VIPeR 和 CUHK01 上的实验验证了，正则化的模型约束往往可以带来整体算法性能的提升。

（2）提出了一种统一的局部统计特征提取框架，并结合多核学习实现ReID 中的多特征融合

在监控场景中实现跨摄像头视域的行人匹配是一项极富挑战的任务，因为不同拍摄角度、不同行人姿态、不同光照条件、以及局部遮挡等都会引起行人外观的剧烈变化，从而增大匹配难度。目前，大量的研究工作主要集中在，构造优秀的特征表示或者学习合理的特征匹配模型这两个方面。然而，由于影响行人外观的因素众多，很难构造单一特征来全面地刻画行人外观；而且，不同特征的提取过程相对独立、缺乏系统而详细地评估分析，很难启发研究人员充分发掘特征的性能或设计其

他更有效的特征表示。为此，我们提出了一种空间金字塔统计特征提取框架，在此框架基础上去实现多种 ReID 中常用统计特征的提取以及改进；同时，我们还利用基于多核学习的费舍尔判别分析方法实现 ReID 中的度量学习和多特征融合。实验结果证明，在 ReID 任务中，此框架下提取的改进局部统计特征性能要优于原始特征，并且结合多特征融合算法后可以进一步提升再识别的准确率。

（3）提出了一个端到端的上下文信息敏感的特征序列提取以及基于双重注意力机制的序列匹配ReID 模型

传统的 ReID 算法在匹配行人图像或者行人跟踪序列之前，往往先将其表示成为一个单独的特征向量，然后在向量空间进行度量学习。然而在复杂的拍摄环境下，单一特征向量并不足以消除行人外观上的模糊性。为此我们提出，将每个行人表示成一系列包含细节信息的特征集合或者序列，并利用双重注意力机制进行序列匹配，从而实现高精度的行人再识别。模型中采用的双重注意力机制是整个算法的核心，其中序列内部的注意力机制用来进行特征序列的去噪精炼，而序列间的注意力机制用来实现序列对的语义对齐。借助这两种注意力机制，包含在特征序列对中的细节信息可以被自动地挖掘出来，并被合理地比较，从而得到恰当的行人距离或相似性度量。实验结果证明我们的模型在多个大规模数据集上都取得了同期最优的性能。

6.2 研究展望

本文主要针对目前ReID 中存在困难以及现有算法在特定应用场景下的不足，分别从小数据集上的度量学习模型泛化能力、基于手工特征设计和多特征融合的分步处理模型、基于特征序列提取和序列匹配的端到端处理模型三个方面，提出高精的ReID 算法或者改进算法，使再识别的准确度进一步提升，甚至达到同期最高水平。但是由于监控场景的复杂性，目前仍有以下问题值得我们继续研究和探讨：

- ReID 中面临的一个实际问题是标记训练样本的缺乏。然而随着监控视频网络的广泛应用，甚至互联网摄像头的普及，每天都有大量原始监控视频数据产生。如何有效利用这些丰富的无标签或者弱标签数据，进一步提升ReID 算法的泛化能力，是值得研究的技术热点之一。
- 在以往的研究中，大家通常更关注ReID 算法的再识别准确率，投入了大量工作使其逐渐能满足实际应用对算法精度的限制。然而，在智能监控场景下，受系统实时性以及监控数据规模庞大的要求，算法的运行效率，也是必须要考量的指标之一。如何设计高效以及可以在超大规模数据集上执行的ReID 算法是今

后研究的关键点之一。

- ReID 是智能视觉监控系统的关键，连接着单摄像头行人检测跟踪技术和多摄像头协同监控技术。目前主流的研究方法是将行人检测跟踪、ReID、以及多摄像头监控网络优化等任务单独建模优化，因此往往无法保证整个系统的性能最优。如何联合不同智能视觉监控阶段，实现整体监控系统的性能最优，是更能满足实际应用的待研究问题之一。

附录 A 缩略语表

AdaBoost	Adaptive Boosting，自适应增强
AMC	Ambiguity-Sensitive Matching Classifier，模糊性敏感的匹配分类器
AP	Average Precision，平均精度
Bi-Recurrent CNN	Bi-Recurrent Convolutional Neural Network，双向循环卷积神经网络
BN	Batch Normalization，批归一化
BoW	Bag-of-Words，词袋模型
CCA	Canonical Correlation Analysis，典型关联分析
CDL	Coupled Dictionary Learning，对偶字典学习
ChnFtrs	Integral Channel Features，积分通道特征
CMC	Cumulative Matching Characteristic，累计匹配性能
CNN	Convolutional Neural Network，卷积神经网络
Cov	Covariance Martix，协方差矩阵
DFL	Decision Function Learning，决策函数学习
DL	Deep Learning，深度学习
DL-SR	Dictionary Learning and Sparse Representation，字典学习与稀疏表示
DPM	Deformable Parts Model，可变形组件模型
DSIFT	Dense Scale-Invariant Feature Transform，稠密的尺度不变特征变换
DuATM	Dual Attention Matching Network，双重注意力匹配网络
DVR	Discriminative Video Ranking，可判别视频排序
ELF	Ensemble of Localized Features，局部特征集成
FDA	Fisher Discriminant Analysis，费舍尔判别分析
FDR	Fisher Discriminant Ratio，费舍尔判别比
FEP	Flow Energy Profile，光流能量分布图
FV	Fisher Vector，费舍尔向量
GMMCP Tracker	Generalized Maximum Multi Clique Problem，广义最大多团问题跟踪器
GOG	Gaussian Of Gaussian，层次化的高斯描述子
GRQ	Generalized Rayleigh Quotient，广义瑞利商
GSet	Gallery Set，候选集

HE	Histogram Encoding，直方图编码
Hist	Color Histogram，颜色直方图
HOG	Histogram of Oriented Gradient，梯度方向直方图
HOG3D	Spatio-Temporal Descriptor based on 3D Gradients，基于三维梯度的时空特征描述子
HTM	Hard Triplet Mining，难分类三元组挖掘
IDE	ID-discriminative Embedding，具有身份区分度的嵌入特征
ISR	Iterative Re-weighted Sparse Ranking，迭代稀疏系数重分配排序
IVS	Intelligent Visual Surveillance，智能视觉监控
KCE	Kernel Codebook Encoding，核码书编码
KISSME	Keep It Simple and Straightforward Metric，简单直观的距离度量
kLFDA	kernel Local Fisher Discriminant Analysis，核化局部费舍尔判别分析
LADF	Locally-Adaptive Decision Functions，局部自适应决策函数
LAFT	Locally Aligned Feature Transforms，局部对齐特征变换
LBP	Local Binary Pattern，局部二值模式
ldLMNN	LogDet regularized LMNN，LogDet 散度正则化的 LMNN
LFDA	Local Fisher Discriminant Analysis，局部费舍尔判别分析
LIE	Linear Interpolation Encoding，线性插值编码
LMNN	Large Margin Nearest Neighbors，最大间隔最近邻
LOMO	Local Maximal Occurrence Representation，局部最大共生表示
mAP	Mean Average Precision，均值平均精度
MFA	Marginal Fisher Analysis，边缘费舍尔分析
MKL	Multiple Kernel Learning，多核学习
mkLFDA	multiple kernel Local Fisher Discriminant Analysis，多核局部费舍尔判别分析算法
MTMCT	Multi-Target Multi-Camera Tracking，多目标多摄像头跟踪
NLML	Nonlinear Local Metric Learning，非线性局部度量学习
nuLMNN	nuclear norm Regularized LMNN，核范数正则化的 LMNN
OP-ReID	Open-Set Person Re-Identification，开集的行人再识别
PCA	Principal Component Analysis，主成分分析
PCCA	Pairwise Constrained Component Analysis，成对约束成分分析
PeRe	Person Retrieval，行人检索

PeSe	Person Search，行人搜寻
PRC	Precision-Recall Curve，精确率-召回率曲线
PRDC	Probabilistic Relative Distance Comparison，概率相对距离比较
PSD	Positive Semi-Definite，半正定
PSet	Probe Set，探测集
QCQP	Quadratically Constrained Quadratic Programming，带二次约束的二次规划
Rank-1	Rank-1 Accuracy，首位准确度
RBF	Radial Basis Function，径向基函数
rDFL	regularized Decision Function Learning，正则化的决策函学习
ReID	Person Re-Identification，行人再识别
ReLU	Rectified Linear Unit，线性整流单元
RNN	Recurrent Neural Network，循环神经网络
ROI	Region of Interest，感兴趣区域
SCE	Salient Color Encoding，显著颜色编码
SCN	Salient Color Names，显著颜色命名特征
SDP	Semi-Definite Programming，半正定规划
SDR	Semidefinite Programming Relaxations，半正定规划松弛
sFDA	stable Fisher Discriminant Analysis，稳定费舍尔判别分析
SGD	Stochastic Gradient Descent，随机梯度下降法
SILTP	Scale Invariant Local Ternary Pattern，尺度不变的局部三元模式
spCN	spatial pyramid based Color Names，基于空间金字塔的颜色命名特征
spCov	spatial pyramid based Covariance Feature，基于空间金字塔的协方差特征
spHist	spatial pyramid based Color Histogram，基于空间金字塔的颜色直方图
spHOG	spatial pyramid based Histogram of Oriented Gradient，基于空间金字塔的方向梯度直方图
spLBP	spatial pyramid based Local Binary Pattern，基于空间金字塔的局部二值模式
SPM	Spatial Pyramid Matching，空间金字塔匹配
SRC	Sparse Representation Classification，稀疏表示分类
trLMNN	trace norm Regularized LMNN，迹范数正则化的 LMNN
X-ReID	Cross-Modality Person Re-Identification，跨模态行人检索

参考文献

[1] Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-Identification[J]. CoRR, 2017, abs/1703.07737.

[2] Zheng L, Zhang H, Sun S, et al. Person Re-identification in the Wild[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 3346–3355.

[3] Zheng L, Shen L, Tian L, et al. Scalable Person Re-identification: A Benchmark[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 1116–1124.

[4] Gray D, Brennan S, Tao H. Evaluating Appearance Models for Recognition, Reacquisition, and Tracking[A]. // Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)[C]. 2007.

[5] Li W, Zhao R, Wang X. Human Reidentification with Transferred Metric Learning[A]. // Proc. Asian Conference on Computer Vision (ACCV)[C]. 2012.

[6] Gong S, Cristani M, Yan S, et al. Person Re-Identification[M]. Springer, 2014.

[7] Zheng L, Yang Y, Hauptmann A G. Person Re-identification: Past, Present and Future[J]. CoRR, 2016, abs/1610.02984.

[8] Gheissari N, Sebastian T B, Hartley R. Person Reidentification Using Spatiotemporal Appearance[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2006: 1528–1535.

[9] Gray D, Tao H. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2008: 262–275.

[10] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2010: 2360–2367.

[11] Bąk S, Charpiat G, Corvée E, et al. Learning to Match Appearances by Correlations in a Covariance Metric Space[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2012: 806–820.

[12] Bazzani L, Cristani M, Murino V. Symmetry-driven accumulation of local features for human characterization and re-identification[J]. Computer Vision and Image Understanding (CVIU), 2013, 117(2): 130 – 144.

[13] Zhao R, Ouyang W, Wang X. Unsupervised Salience Learning for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2013: 3586–3593.

[14] Zhao R, Ouyang W, Wang X. Learning Mid-level Filters for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2014: 144–151.

[15] Kviatkovsky I, Adam A, Rivlin E. Color Invariants for Person Reidentification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2013, 35(7): 1622–1634.

[16] Yang Y, Yang J, Yan J, et al. Salient Color Names for Person Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2014: 536–551.

[17] Zheng L, Wang S, Tian L, et al. Query-adaptive late fusion for image search and person re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 1741–1750.

[18] Liao S, Hu Y, Zhu X, et al. Person re-identification by Local Maximal Occurrence representation and metric learning[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 2197–2206.

[19] Shi Z, Hospedales T M, Xiang T. Transferring a semantic representation for person re-identification and search[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 4184–4193.

[20] Su C, Yang F, Zhang S, et al. Multi-Task Learning with Low Rank Attribute Embedding for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 3739–3747.

[21] Chen Y C, Zheng W S, Lai J. Mirror Representation for Modeling View-specific Transform in Person Re-identification[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2015: 3402–3408.

[22] Matsukawa T, Okabe T, Suzuki E, et al. Hierarchical Gaussian Descriptor for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1363–1372.

[23] Zheng W S, Gong S, Xiang T. Person re-identification by probabilistic relative distance comparison[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2011: 649–656.

[24] Kostinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2012: 2288–2295.

[25] Mignon A, Jurie F. PCCA: A new approach for distance learning from sparse pairwise constraints[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2012: 2666–2672.

[26] Hirzer M, Roth P M, Köstinger M, et al. Relaxed Pairwise Learned Metric for Person Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2012: 780–793.

[27] Li Z, Chang S, Liang F, et al. Learning Locally-Adaptive Decision Functions for Person Verification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2013: 3610–3617.

[28] Pedagadi S, Orwell J, Velastin S, et al. Local Fisher Discriminant Analysis for Pedestrian Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2013: 3318–3325.

[29] Li W, Wang X. Locally Aligned Feature Transforms across Views[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2013: 3594–3601.

[30] Zhao R, Ouyang W, Wang X. Person Re-identification by Salience Matching[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2013: 2528–2535.

[31] Zheng W S, Gong S, Xiang T. Reidentification by Relative Distance Comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2013, 35(3): 653–668.

[32] Liu X, Song M, Tao D, et al. Semi-supervised Coupled Dictionary Learning for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2014: 3550–3557.

[33] Xiong F, Gou M, Camps O, et al. Person Re-Identification Using Kernel-Based Metric Learning Methods[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2014: 1–16.

[34] Ma L, Yang X, Tao D. Person Re-Identification Over Camera Networks Using Multi-Task Distance Metric Learning[J]. IEEE Transactions on Image Processing (TIP), 2014, 23(8): 3656–3670.

[35] Paisitkriangkrai S, Shen C, van den Hengel A. Learning to rank in person re-identification with metric ensembles[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 1846–1855.

[36] Chen D, Yuan Z, Hua G, et al. Similarity learning on an explicit polynomial kernel feature map for person re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 1565–1573.

[37] Liao S, Li S Z. Efficient PSD Constrained Asymmetric Metric Learning for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 3685–3693.

[38] Karanam S, Li Y, Radke R J. Person Re-Identification with Discriminatively Trained Viewpoint Invariant Dictionaries[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 4516–4524.

[39] Li S, Shao M, Fu Y. Cross-view Projective Dictionary Learning for Person Re-identification[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2015: 2155–2161.

[40] Shen Y, Lin W, Yan J, et al. Person Re-Identification with Correspondence Structure Learning[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 3200–3208.

[41] Martinel N, Micheloni C, Foresti G L. Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning[J]. IEEE Transactions on Image Processing (TIP), 2015, 24(12): 5645–5658.

[42] Chen J, Zhang Z, Wang Y. Relevance Metric Learning for Person Re-Identification by Exploiting Listwise Similarities[J]. IEEE Transactions on Image Processing (TIP), 2015, 24(12): 4741–4755.

[43] Wu Z, Li Y, Radke R J. Viewpoint Invariant Human Re-Identification in Camera Networks Using Pose Priors and Subject-Discriminative Features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015, 37(5): 1095–1108.

[44] Yang Y, Liao S, Lei Z, et al. Large Scale Similarity Learning Using Similar Pairs for Person Verification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2016: 3655–3661.

[45] Wang G, Lin L, Ding S, et al. DARI: Distance Metric and Representation Integration for Person Verification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2016: 3611–3617.

[46] Zhang L, Xiang T, Gong S. Learning a discriminative null space for person re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016.

[47] Zhang Y, Li B, Lu H, et al. Sample-Specific SVM Learning for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1278–1287.

[48] Chen D, Yuan Z, Chen B, et al. Similarity Learning with Spatial Constraints for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1268–1277.

[49] Varior R R, Shuai B, Lu J, et al. A Siamese Long Short-Term Memory Architecture for Human Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016.

[50] Tao D, Guo Y, Song M, et al. Person Re-Identification by Dual-Regularized KISS Metric Learning[J]. IEEE Transactions on Image Processing (TIP), 2016, 25(6): 2726–2738.

[51] Sun C, Wang D, Lu H. Person Re-Identification via Distance Metric Learning With Latent Variables[J]. IEEE Transactions on Image Processing (TIP), 2017, 26(1): 23–34.

[52] Zhou J, Yu P, Tang W, et al. Efficient Online Local Metric Adaptation via Negative Samples for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[53] Lin W, Shen Y, Yan J, et al. Learning Correspondence Structures for Person Re-Identification[J]. IEEE Transactions on Image Processing (TIP), 2017, 26(5): 2438–2453.

[54] Wang T, Gong S, Zhu X, et al. Person Re-identification by Video Ranking[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2014: 688–703.

[55] Liu K, Ma B, Zhang W, et al. A Spatio-Temporal Appearance Representation for Video-Based Pedestrian Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 3810–3818.

[56] McLaughlin N, d Rincon J M, Miller P. Recurrent Convolutional Network for Video-Based Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1325–1334.

[57] You J, Wu A, Li X, et al. Top-Push Video-Based Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1345–1353.

[58] Yan Y, Ni B, Song Z, et al. Person Re-identification via Recurrent Feature Aggregation[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016.

[59] Zhu X, Jing X Y, Wu F, et al. Video-based Person Re-identification by Simultaneously Learning Intra-video and Inter-video Distance Metrics[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2016: 3552–3558.

[60] Wang T, Gong S, Zhu X, et al. Person Re-Identification by Discriminative Selection in Video Ranking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016, 38(12): 2501–2514.

[61] Liu Y, Yan J, Ouyang W. Quality Aware Network for Set to Set Recognition[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 4694–4703.

[62] Zhou Z, Huang Y, Wang W, et al. See the Forest for the Trees: Joint Spatial and Temporal Recurrent Neural Networks for Video-Based Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 6776–6785.

[63] Chung D, Tahboub K, Delp E J. A Two Stream Siamese Convolutional Neural Network for Person Re-identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 1992–2000.

[64] Xu S, Cheng Y, Gu K, et al. Jointly Attentive Spatial-Temporal Pooling Networks for Video-Based Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[65] Zheng K, Fan X, Lin Y, et al. Learning View-Invariant Features for Person Identification in Temporally Synchronized Videos Taken by Wearable Cameras[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[66] Li W, Zhao R, Xiao T, et al. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2014: 152–159.

[67] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 3908–3916.

[68] Wang F, Zuo W, Lin L, et al. Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1288–1296.

[69] Xiao T, Li H, Ouyang W, et al. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1249–1258.

[70] Cheng D, Gong Y, Zhou S, et al. Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 1335–1344.

[71] Su C, Zhang S, Xing J, et al. Deep Attributes Driven Multi-camera Person Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016: 475–491.

[72] Varior R R, Haloi M, Wang G. Gated Siamese Convolutional Neural Network Architecture for Human Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016.

[73] Zhang Y, Li X, Zhao L, et al. Semantics-Aware Deep Correspondence Structure Learning for Robust Person Re-Identification[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2016.

[74] Subramaniam A, Chatterjee M, Mittal A. Deep Neural Networks with Inexact Matching for Person Re-Identification[A]. // Proc. Advances in Neural Information Processing Systems (NIPS)[C]. 2016: 2667–2675.

[75] Chen S Z, Guo C C, Lai J H. Deep Ranking for Person Re-Identification via Joint Representation Learning[J]. IEEE Transactions on Image Processing (TIP), 2016, 25(5): 2353–2367.

[76] Liu H, Feng J, Qi M, et al. End-to-End Comparative Attention Networks for Person Re-Identification[J]. IEEE Transactions on Image Processing (TIP), 2017, 26(7): 3492–3506.

[77] Chen W, Chen X, Zhang J, et al. A Multi-Task Deep Network for Person Re-Identification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2017.

[78] Zhang X, Luo H, Fan X, et al. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification[J]. CoRR, 2017, abs/1711.08184.

[79] Chen W, Chen X, Zhang J, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017.

[80] Li D, Chen X, Zhang Z, et al. Learning Deep Context-Aware Features over Body and Latent Parts for Person Re-identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 7398–7407.

[81] Zhou S, Wang J, Wang J, et al. Point to Set Similarity Based Deep Feature Learning for Person Re-Identification[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 5028–5037.

[82] Zhao H, Tian M, Sun S, et al. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 907–915.

[83] Zhao L, Li X, Zhuang Y, et al. Deeply-Learned Part-Aligned Representations for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 3219–3228.

[84] Liu X, Zhao H, Tian M, et al. HydraPlus-Net: Attentive Deep Features for Pedestrian Analysis[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 350–359.

[85] Qian X, Fu Y, Jiang Y G, et al. Multi-Scale Deep Learning Architectures for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[86] Su C, Li J, Zhang S, et al. Pose-Driven Deep Convolutional Model for Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[87] Sun Y, Zheng L, Deng W, et al. SVDNet for Pedestrian Retrieval[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[88] Li W, Zhu X, Gong S. Person Re-Identification by Deep Joint Learning of Multi-Loss Classification[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2017.

[89] Xiao J, Xie Y, Tillo T, et al. IAN: The Individual Aggregation Network for Person Search[J]. CoRR, 2017, abs/1705.05552.

[90] Xiao T, Li S, Wang B, et al. Joint Detection and Identification Feature Learning for Person Search[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 3376–3385.

[91] Liu H, Feng J, Jie Z, et al. Neural Person Search Machines[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 493–501.

[92] Li S, Xiao T, Li H, et al. Person Search with Natural Language Description[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 5187–5196.

[93] Yamaguchi M, Saito K, Ushiku Y, et al. Spatio-Temporal Person Retrieval via Natural Language Queries[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 1462–1471.

[94] Wu A, Zheng W S, Yu H X, et al. RGB-Infrared Cross-Modality Person Re-identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 5390–5399.

[95] Ye M, Lan X, Li J, et al. Hierarchical Discriminative Learning for Visible Thermal Person Re-Identification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2018.

[96] Zheng W S, Gong S, Xiang T. Towards Open-World Person Re-Identification by One-Shot Group-Based Verification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016, 38(3): 591–606.

[97] Liao S, Mo Z, Hu Y, et al. Open-set Person Re-identification[J]. CoRR, 2014, abs/1408.0872.

[98] Baltieri D, Vezzani R, Cucchiara R. 3DPes: 3D People Dataset for Surveillance and Forensics[A]. // Proc. International ACM Workshop on Multimedia access to 3D Human Objects[C]. 2011: 59–64.

[99] Zheng Z, Zheng L, Yang Y. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[100] Hirzer M, Beleznai C, Roth P M, et al. Person Re-Identification by Descriptive and Discriminative Classification[A]. // Proc. Scandinavian Conference on Image Analysis (SCIA)[C]. 2011.

[101] Hirzer M, Beleznai C, Roth P M, et al. Proc. European Conference on Computer Vision (ECCV)[C]. 2016.

[102] Martinel N, Das A, Micheloni C, et al. Re-Identification in the Function Space of Feature Warps[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015, 37(8): 1656–1669.

[103] Jing X Y, Zhu X, Wu F, et al. Super-resolution Person re-identification with semi-coupled low-rank discriminant dictionary learning[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 695–704.

[104] Li X, Zheng W S, Wang X, et al. Multi-Scale Learning for Low-Resolution Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 3765–3773.

[105] Liu Z, Wang D, Lu H. Stepwise Metric Promotion for Unsupervised Video Person Re-identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 2448–2457.

[106] Yang Y, Lei Z, Zhang S, et al. Metric Embedded Discriminative Vocabulary Learning for High-level Person Representation[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2016.

[107] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2002, 24(7): 971–987.

[108] Dalal N, Triggs B. Histograms of oriented gradients for human detection[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2005: 886–893 vol. 1.

[109] Jobson D J, Rahman Z, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. IEEE Transactions on Image Processing (TIP), 1997, 6(7): 965–976.

[110] Liao S, Zhao G, Kellokumpu V, et al. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2010: 1301–1306.

[111] Martinel N, Das A, Micheloni C, et al. Temporal Model Adaptation for Person Re-identification[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016: 858–877.

[112] Chen J, Wang Y, Qin J, et al. Fast Person Re-identification via Cross-Camera Semantic Binary Transformation[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 5330–5339.

[113] Zhong Z, Zheng L, Cao D, et al. Re-ranking Person Re-identification with k-reciprocal Encoding[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017.

[114] Bai S, Bai X, Tian Q. Scalable Person Re-identification on Supervised Smoothed Manifold[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 3356–3365.

[115] Panda R, Bhuiyan A, Murino V, et al. Unsupervised Adaptive Re-identification in Open World Dynamic Camera Networks[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 1377–1386.

[116] Ye M, Ma A J, Zheng L, et al. Dynamic Label Graph Matching for Unsupervised Video Re-identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 5152–5160.

[117] Klaser A, Marszalek M, Schmid C. A Spatio-Temporal Descriptor Based on 3D-Gradients[A]. // Proc. British Machine Vision Conference (BMVC)[C]. Leeds, United Kingdom: British Machine Vision Association, 2008: 275:1–10.

[118] Li D, Zhang Z, Chen X, et al. A Richly Annotated Dataset for Pedestrian Attribute Recognition[J]. CoRR, 2016.

[119] Lin Y, Zheng L, Zheng Z, et al. Improving Person Re-identification by Attribute and Identity Learning[J]. CoRR, 2017, abs/1703.07220.

[120] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[A]. // Proc. Advances in Neural Information Processing Systems (NIPS)[C]. USA: Curran Associates Inc., 2012: 1097–1105.

[121] Chen Y C, Zhu X, Zheng W S, et al. Person Re-Identification by Camera Correlation Aware Feature Augmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018, 40(2): 392–408.

[122] De Cheng L L A G H Y G N Z Xiaojun Chang. Discriminative Dictionary Learning With Ranking Metric Embedded for Person Re-Identification[A]. // Proc. International Joint Conference on Artificial Intelligence (IJCAI)[C]. 2017: 964–970.

[123] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2016: 770–778.

[124] Huang S, Lu J, Zhou J, et al. Nonlinear Local Metric Learning for Person Re-identification[J]. CoRR, 2015, abs/1511.05169.

[125] Yu H X, Wu A, Zheng W S. Cross-View Asymmetric Metric Learning for Unsupervised Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2017: 994–1002.

[126] Zheng W S, Li X, Xiang T, et al. Partial Person Re-Identification[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2015: 4678–4686.

[127] Lisanti G, Masi I, Bagdanov A D, et al. Person Re-Identification by Iterative Re-Weighted Sparse Ranking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015, 37(8): 1629–1642.

[128] Zhou Q, Fan H, Zheng S, et al. Graph Correspondence Transfer for Person Re-identification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2018.

[129] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 1–9.

[130] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research (JMLR), 2009, 10: 207–244.

[131] Davis J V, Kulis B, Jain P, et al. Information-theoretic metric learning[A]. // Proc. International Conference on Machine Learning (ICML)[C]. 2007.

[132] Xu Y, Lin L, Zheng W S, et al. Human Re-identification by Matching Compositional Template with Cluster Sampling[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2013: 3152–3159.

[133] Tao D, Jin L, Wang Y, et al. Person Reidentification by Minimum Classification Error-Based KISS Metric Learning[J]. IEEE Transactions on Cybernetics (ToC), 2015, 45(2): 242–252.

[134] Kodirov E, Xiang T, Fu Z, et al. Person Re-identification by Unsupervised L1 Graph Learning[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2016: 178–195.

[135] Wei L, Tian Y, Wang Y, et al. Swiss-system Based Cascade Ranking for Gait-based Person Re-identification[A]. // Proc. AAAI Conference on Artificial Intelligence (AAAI)[C]. 2015: 1882–1888.

[136] Ma A J, Yuen P C, Li J. Domain Transfer Support Vector Ranking for Person Re-identification without Target Camera Label Information[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2013: 3567–3574.

[137] Ma A J, Li J, Yuen P C, et al. Cross-Domain Person Reidentification Using Domain Adaptation Ranking SVMs[J]. IEEE Transactions on Image Processing (TIP), 2015, 24(5): 1599–1613.

[138] Das A, Chakraborty A, Roy-Chowdhury A K. Consistent Re-identification in a Camera Network[A]. // Proc. European Conference on Computer Vision (ECCV)[C]. 2014: 330–345.

[139] Zhang R, Lin L, Zhang R, et al. Bit-Scalable Deep Hashing With Regularized Similarity Learning for Image Retrieval and Person Re-Identification[J]. IEEE Transactions on Image Processing (TIP), 2015, 24(12): 4766–4779.

[140] Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2006: 2169–2178.

[141] Ken Chatfield A V Victor Lempitsky, Zisserman A. The devil is in the details: an evaluation of recent feature encoding methods[A]. // Proc. British Machine Vision Conference (BMVC)[C]. 2011: 76.1–76.12.

[142] Dollar P, Tu Z, Perona P, et al. Integral Channel Features[A]. // Proc. British Machine Vision Conference (BMVC)[C]. 2009: 91.1–91.11.

[143] Tuzel O, Porikli F, Meer P. Pedestrian Detection via Classification on Riemannian Manifolds[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2008, 30(10): 1713–1727.

[144] Wang X, Han T X, Yan S. An HOG-LBP human detector with partial occlusion handling[A]. // Proc. IEEE International Conference on Computer Vision (ICCV)[C]. 2009: 32–39.

[145] Vedaldi A, Fulkerson B. Vlfeat: An Open and Portable Library of Computer Vision Algorithms[A]. // Proc. ACM International Conference on Multimedia (ACMM)[C]. 2010: 1469–1472.

[146] Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis[J]. Journal of Machine Learning Research (JMLR), 2007, 8: 1027–1061.

[147] Lin Y Y, Liu T L, Fuh C S. Multiple Kernel Learning for Dimensionality Reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2011, 33(6): 1147–1160.

[148] Ma B, Su Y, Jurie F. BiCov: a novel image representation for person re-identification and face verification[A]. // Proc. British Machine Vision Conference (BMVC)[C]. Guildford, United Kingdom: 2012: 11 pages.

[149] Layne R, Hospedales T M, Gong S. Person Re-identification by Attributes[A]. // Proc. British Machine Vision Conference (BMVC)[C]. 2012: 24.1–24.11.

[150] Geng M, Wang Y, Xiang T, et al. Deep Transfer Learning for Person Re-identification[J]. CoRR, 2016, abs/1611.05244.

[151] Vinyals O, Bengio S, Kudlur M. Order matters: Sequence to sequence for sets[A]. // Proc. International Conference on Learning Representations (ICLR)[C]. 2016.

[152] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention[A]. // Proc. International Conference on Machine Learning (ICML)[C]. 2015: 2048–2057.

[153] Yang J, Ren P, Zhang D, et al. Neural Aggregation Network for Video Face Recognition[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017.

[154] Wang S, Jiang J. A Compare-Aggregate Model for Matching Text Sequences[A]. // Proc. International Conference on Learning Representations (ICLR)[C]. 2017.

[155] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017.

[156] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[A]. // Proc. International Conference on Machine Learning (ICML)[C]. 2015: 448–456.

[157] Cogswell M, Ahmed F, Girshick R, et al. Reducing overfitting in deep networks by decorrelating representations[A]. // Proc. International Conference on Learning Representations (ICLR)[C]. 2016.

[158] DeVries T, Taylor G W. Dataset Augmentation in Feature Space[A]. // Proc. International Conference on Learning Representations Workshop (ICLRW)[C]. 2017.

[159] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015: 815–823.

[160] Yao H, Zhang S, Zhang Y, et al. Deep Representation Learning with Part Loss for Person Re-Identification[J]. CoRR, 2017, abs/1707.00798.

[161] Lin J, Ren L, Lu J, et al. Consistent-Aware Deep Learning for Person Re-identification in a Camera Network[A]. // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2017: 5771–5780.

[162] Zhang Y, Xiang T, Hospedales T M, et al. Deep Mutual Learning[J]. CoRR, 2017, abs/1706.00384.

[163] Zhong Z, Zheng L, Kang G, et al. Random Erasing Data Augmentation[J]. CoRR, 2017.

[164] Chen Y, Wang N, Zhang Z. DarkRank: Accelerating Deep Metric Learning via Cross Sample Similarities Transfer[J]. CoRR, 2017.

[165] Lin Y, Zheng L, Zheng Z, et al. Improving person re-identification by attribute and identity learning[J]. CoRR, 2017.

[166] Chen Y, Zhu X, Gong S. Person Re-Identification by Deep Learning Multi-Scale Representations[A]. // Proc. IEEE International Conference on Computer Vision Workshop (ICCVW)[C]. 2017.

致 谢

在求学生涯即将结束之际，不禁百感交集，思绪万千。

从考入大学到博士毕业整整十年。十年倥偬，十年转瞬，在求知若渴的岁月里，风雨兼程。硕博连读的六年纵然充满挑战，但梦想和执着也常伴左右。项目、科研、论文、访问交流，求学生活充实而多彩。曾因，没有好的科研思路而抓耳挠腮甚至郁郁寡欢过；曾因，一直无法得到理想的实验结果而暗自神伤甚至彻夜难眠过；曾因，生活的苟且与诗和远方而纠结怀疑过。但也曾在论文投稿截止期前通宵达旦；也曾在论文被录用的那一刻激动得双拳紧握，对自己说，“YES”！回忆起这段艰辛而快乐的日子，需要感谢的人很多很多，感谢你们无私的帮助与付出！

衷心感谢我的导师张洪刚老师。是您给了我勇气，引领我步入科研的大门，鞭策我勇攀学术的高峰；是您一次次为我创造学习、锻炼的机会，培养我勤奋、自强的求学态度。张老师开阔的研究视野、创新求是的科研精神，激励我奋勇向前。虽然有时科研任务沉重，但您总能鼓励我保持积极向上的态度，以健康的兴趣爱好调剂生活。每当看到您球场上飒爽的英姿、爆发的激情，我都会深受感染。您身上满满的正能量，让我十分钦佩。

衷心感谢李春光老师，感谢您在我读博期间给予我的细致和周到的帮助，感谢您在我科研迷茫之际给予我的指导和鼓励。从理论基础的学习积累、到论文公式的解释推导，从学术想法的讨论规划、到论文撰写的修改润色，李老师让我切切实实领略到了深厚扎实的学术功底，也教会了我严谨踏实的治学态度。您将是我一生的良师益友。

衷心感谢郭军老师，您严谨治学、勤奋做事、为人谦和、待人友善，不仅是我科研路上的灯塔，还是我人生路上的榜样。您对科研的热忱和渊博的专业知识，让我十分钦佩，虽然身为校长，但是仍几十年如一日地坚持在科研道路上，身体力行；每次博士例会上，您都能一针见血地指出学生的不足，提出高屋建瓴的建议，为我们的学术研究带来建设性的启迪。很庆幸在自己的学生时代最末步入社会之前能得到郭老师的指导，藉此论文完成之际，谨向您献上我崇高的敬意和衷心的感谢。

感谢北邮模式识别实验室的马占宇老师、徐蔚然老师、高升老师、邓伟洪老师、陈光老师、肖波老师、李思老师、刘瑞芳老师、胡佳妮老师、刘刚老师、徐雅静老师、徐前方老师，感谢各位老师在论文和学术上给予的指导。还要感谢在南洋理工

访问交流期间给予我很多帮助的王刚老师和 Alex C. Kot 教授，您们国际化的视野带给我许多科研上的启发。同时，感谢南洋理工大学 ROSE 实验室为我提供了先进的实验设备、舒适的科研办公环境以及经济上的照顾，使我在异国求学路上倍感温暖与力量。

感谢实验室同门齐勇刚、张佳玥、赵凯莉、张春云、秦臻、邱琳、戚园园、于泓、赵宇、狄帅等师兄师姐，感谢张军建、李珂、林毅等同学，感谢蒲石、徐鹏、郭若沛、刘旭卿等师弟，在生活和学习上给予的帮助和关心。感谢硕士研究生班小伙伴张斌、徐彬、陈代武、王明明、石宽、李思远、谷元庆、郭杰、黄一、张囡囡、侯成文、梁冬、张文静，是你们让原本枯燥的研究生活变得丰富多彩。感谢在南洋理工大学交流期间，一起奋斗玩耍的朋友梅剑寒、胡平、刘俊、顾久祥、陈坚达、陈增海、孔翔飞、Henghui Ding、Feixiang He、Xingxing Wang、Jason Kuen，是你们让原本孤独的异国求学变得温暖。

我还要感谢我的家人们，你们是我拼搏奋斗的源动力。感谢相识十余年的妻子，未来的人生岁月里，我会给你最长情的告白—陪伴。感谢我的父母和岳父母，感谢你们含辛茹苦地养育，也感谢你们始终无条件的支持和信任，以后我会为你们撑起这个家。

最后，真诚感谢参与本论文开题、中期及毕业答辩工作的各位专家和老师们，感谢你们对我提出的宝贵意见和建议。同时，向所有关注我、关心我支持我成长的人士致以真诚的感谢和最诚挚的祝福。感谢命运和生活，为我打开了求学之门，学生时代的成长与历练为以后的工作与生活打下了坚实的基础，我将以坚持不懈的努力去续写更为美丽精彩的人生篇章！

攻读学位期间发表的学术论文目录

期刊论文

[1] **Si, Jianlou**, Zhang H, Li C G, et al. Spatial Pyramid-Based Statistical Features for Person Re-Identification: A Comprehensive Evaluation[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems (TSMC), 2017, PP(99): 1–15. (SCI 收录，10.1109/TSMC.2016.2645660).

会议论文

[2] **Si, Jianlou**, Zhang H, Li C G, et al. Dual Attention Matching Network for Context-Aware Feature Sequence based Person Re-Identification[A]. // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2018. (CCF A 类会议，已发表，2018.06.19 开会).

[3] **Si, Jianlou**, Zhang H, Li C G. Regularization in metric learning for person re-identification[A]. // 2015 IEEE International Conference on Image Processing (ICIP)[C]. 2015: 2309–2313. (EI 收录，检索号：20160601895710).

[4] **Si, Jianlou**, Zhang H, Li C G. Person re-identification via region-of-interest based features[A]. // 2014 IEEE Visual Communications and Image Processing Conference (VCIP)[C]. 2014: 249–252. (EI 收录，检索号：20151300686977).

[5] Zhang H, Zhang H, **Si, Jianlou**. Fusing multiple statistical features via explicit feature mapping for person re-identification[A]. // 2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)[C]. 2016: 371–375. (EI 收录，检索号：20173404060007).

专利

[6] **四建楼** , 张亨洋, 王光宇. 一种行驶证识别方法及装置 [P]. 中国: CN106874901A, 2017–01–17.