密级:_____

中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于无监督局部度量学习的行人再识别研究

作者姓名: 赵恒

指导教师: 韩振军　副教授　中国科学院大学

学位类别: 工程硕士

学科专业: 计算机技术

研究所　: 中国科学院大学电子电气与通信工程学院

二零一七 年 五 月

Research on Person Re-identification via Unsupervised Local Metric Learning

By

Heng Zhao

A Thesis Submitted to

University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Master of Computer Technology

College of Electronics, Electrical and Communication Engineering

University of Chinese Academy of Sciences

May, 2017

中国科学院大学直属院系
研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日　　期：

中国科学院大学直属院系
学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：　　导师签名：

日　　期：　　日　　期：

摘　要

随着社会经济的发展和公共安全意识的加强，越来越多的视频监控设备被部署到人们生活的各个场所，形成一个庞大的监控网络。然而，目前的视频监控分析主要通过相关工作人员对监控视频进行观看和分析来协助完成相关的任务。在大数据时代，人工方法无法适应日益增长的数据规模。行人再识别技术，即通过一个监控摄像头下的人体图像，自动查询其在其他摄像头下的图像，进而得到其在指定监控网络中的行动轨迹，在视频监控和智能化安防领域意义重大。目前的行人再识别技术主要基于监督学习方法，即通过标定数据学习获取适用于特定场景的度量参数。然而，训练数据的获取需要大量人力物力进行数据标定，同时随着监控网络的进一步扩大，用于监督学习的训练数据的标定也会随之变得更加困难。针对这一问题，本文研究使用无标注数据进行训练，提出了基于无监督学习的距离度量算法，即基于样本计算其局部度量参数。进一步，针对初始排序容易受负样本干扰的问题，本文提出了基于近邻的重排序算法来进行排序优化。

本文的主要工作包括：

1. 针对监督方法需要标定训练数据的问题，提出了基于无监督的行人再识别算法。无监督学习方法具有比较好的实用性，适用于目前庞大的视频监控网络。

2. 针对全局的距离度量在处理多样化的数据时存在性能损失的问题，提出了基于样本的局部度量学习算法。针对每个样本学习度量参数，充分考虑样本的特性。

3. 使用重排序方法优化初始排序。重排序考虑了 gallery 样本之间的相似性，进一步优化排序结果，提高行人再识别性能。

关键词：视频监控，行人再识别，局部距离度量，无监督学习

Abstract

With the development of socio-economic and the increasing of public security awareness, more and more video surveillance equipment is deployed to people's living places, forming a huge visual surveillance network. However, the current surveillance video analysis mainly relies on the relevant staff to watch and analyze the surveillance video to help complete the relevant tasks. In the big data era, artificial methods can not adapt to the growing scale of the data. Person re-identification technology, that using target person's body images in a surveillance camera, automatically querying his images in other cameras, and then getting his action track in the designated surveillance network, is very significant in the field of video surveillance and intelligent security. The current pedestrian re-identification technology is mainly based on supervised learning methods, which learn the metric parameters for a specific scene from annotation data. However, a lot of manpower and resources is needed to obtain the annotation data. What's more, it would be more difficult to obtain the annotation data with the expansion of the surveillance network. Aiming at this problem, this paper focuses on the use of unlabeled data for training and proposes a distance metric algorithm based on unsupervised learning, which is based on the target sample when computing metric parameters. Furthermore, for the problem that the initial rank is easy to be influenced by the negative samples, this paper proposes a neighbor based reranking algorithm to optimize the rank.

The main works in this thesis include:

1. Aiming at the problem that the supervised methods need annotating the training data, an unsupervised person re-identification algorithm framework is proposed. Unsupervised learning methods have good practicality and are especially applicable to large video surveillance network.

2. Aiming at the problem that the global distance metric has the performance loss when dealing with the diversified data, a sample-specific local metric learning algorithm is proposed. Learning metric parameters for each sample in order to take the characteristics of each sample into account.

3. Use the reranking method to optimize the initial rank. Reranking considers the similarity between the gallery samples, thus can further optimize the initial rank, improve the person re-identification performance.

Key Words: video surveillance, person re-identification, local distance metric, unsupervised learning

目录

图目录

表目录

第一章 绪论

1.1 研究背景与意义

随着人们安全意识的提高和平安城市等项目的快速推进，我国城市视频监控市场迅速增长。大量的监控摄像头网络被部署到车站、商场、校园等公共场所，为社会稳定和人生安全起着不可替代的保障作用。例如：监控可以提供实时交通路况信息，方便交通部门管理，优化出行；可以协助查找感兴趣目标如走失儿童、犯罪嫌疑人等；可以为偷窃、违法等行为提供事实证据并有效遏制类似事件的发生。随着监控设备相关技术的成熟和成本的降低，数以万计的监控设备无时无刻地在采集数据，形成了海量数据库。如何有效管理和利用监控大数据就成了目前广受关注的问题。

目前，以人工为主的监控方式在大数据背景下遭到了巨大的挑战。人工监控的相关工作人员需要时刻监视视频画面，并分析相关目标和场景，这不仅费时费力，而且错误率会随着数据量的增大而上升。因此，自动化监控技术以其成本低，效率高，扩展性强等优点受到了人们的广泛关注。

近年来，以机器学习、深度学习为基础的计算机视觉技术得到了快速的发展并被广泛运用在一些实际场景中。比如：人脸识别、车牌识别、目标检测等。在视频监控场景中，行人目标检测、跟踪、关联和行为分析等是目前智能化监控主要关注的问题。在一个监控网络中，对某个已经被识别的行人目标进行跨摄像头的关联，即行人再识别，是近 5 年来计算机视觉和视频监控领域的研究热点。

行人再识别技术目的在于获得目标在特定时间，特定监控网络下的整个行动轨迹，如图 1-1 所示。在单个摄像头下，行人检测和跟踪可以很好地发现目标并获得此场景下的行动轨迹。但当目标离开此摄像头覆盖的场景，并在其他摄像头下再次出现时，目标跟踪会由于跨摄像头所形成的目标丢失、摄像头参数和背景不一致等问题导致算法失效。因此，行人再识别应运而生。除了在监控领域，行人再识别技术还可以被运用在多媒体照片归类、图片检索等相关领域。

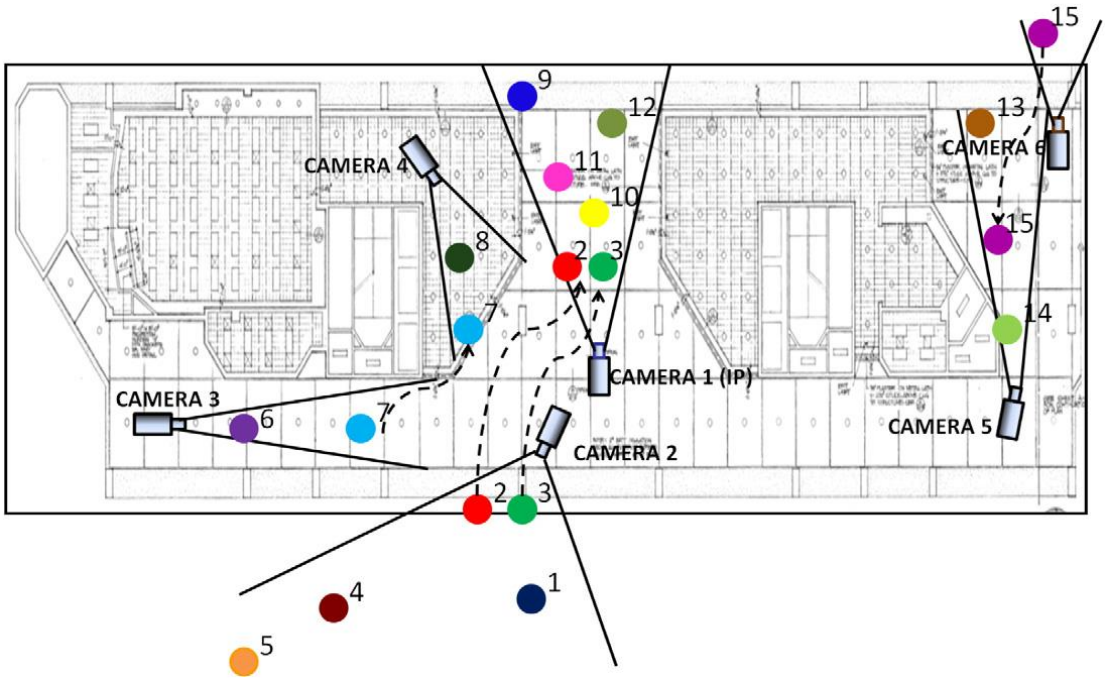


图 1-1 多摄像头监控网络[1]

1.2 国内外研究进展

行人再识别作为计算机视觉中一个热门研究方向，吸引了国内外一大批优秀的研究者投身于此。通过统计近几年发表在计算机视觉领域顶级会议（CVPR、ECCV、ICCV 等）和期刊（PAMI 等）上的论文显示，参与的著名研究机构有：剑桥大学、英国伦敦玛丽女王学院、香港中文大学、北京大学、清华大学、中科院计算所、北京邮电大学等。

监控场景中的行人再识别问题主要面临如下挑战：

1. 光照条件：不同摄像头往往具有不同的光照条件，这使得同一个行人在不同的光照下展现出完全不同的特性，即使是在同一个摄像头下，不同的光照条件也会使得同一个人展现出不同的表观特性。

2. 多视角：由于实际条件限制，摄像头的安装位置往往千差万别，不同的视角如俯视、侧视等获得的人体图片当然也存在着巨大差异。

3. 姿态变化：人体是非刚性的，行走中的人体随着手和腿的摆动会发生巨

大的形变，这就造成了图片对齐的问题。

4. 行人遮挡：遮挡问题在现实场景中普遍存在，尤其是人流密集区域。

5. 相似行人干扰：监控场景中表观相似的行人是影响匹配性能的重要因素，比如穿着类型，颜色等显著性特征。

6. 摄像头参数：摄像头的参数除了焦距、视场角、光圈以外，分辨率是影响匹配性能的主要因素。低分辨率使得人脸、步态等生物特征无法起到有效的识别作用。

为了解决这些问题进而提高行人再识别的性能，研究者们提出了大量行之有效的方法。大部分行人再识别方法主要由两个部分构成：

1. 特征表示[2][3][4][5][6][7]。特征表示的关键在于设计鲁棒性强的行人表示特征，既能够在光照、视角、姿态变化中保持稳定性，又能够在不同的行人之间保持很好的区分性。

2. 距离度量[6][8][9][10]。距离度量的重点在于在样本的特征空间中学习一种映射关系，使得同一个人的不同图片特征之间的距离尽可能小，不同的人的图片的特征之间的距离尽可能大，即最小化类内距，最大化类间距。距离度量通常会以变换矩阵的形式出现。

1.2.1 特征表示

1.2.1.1 底层特征

底层特征表示在各种计算机视觉任务中被广泛运用。如：目标检测，目标识别，图像分类，图像检索，视频分析等。底层特征对目标具有一般性的表示，因此对于不同的视觉任务都非常适用。对于行人再识别问题，常用的底层特征表示有：

1. 颜色直方图：通过统计图像的颜色分布来描述整个图片目标或者其中的某一块图片信息。颜色直方图由于是个全局或者局部的统计描述，因此对于视角、姿态变化比较鲁棒，但是对于光照变化比较敏感。常用的颜色空间有：RGB，HSV 和 YCbCr。

2. 纹理描述：通过描述整张图或者其中一个小区域颜色的结构信息。纹理特征可以很好的弥补单一颜色直方图的缺陷，增强对目标的辨识性。常见的纹理描述子有：LBP（Local Binary Pattern）[11]，Gabor 滤波[12]，共生矩阵（Co-occurrence Matrices）[13]和 HOG（Histogram of Oriented Gradients）[14]等。

3. 局部特征：通过对图片中显著点的表示来描述整个目标。一般选择那些对外界变化比较鲁棒的局部点作为显著点。常见的局部特征描述子有：SIFT（Scale-invariant feature transform）[15]，SURF[16]等。

1.2.1.2 中层语义特征

虽然底层特征以其易于提取，泛化性好被广泛使用在行人再识别问题中，但是在跨摄像头导致的不同的光照，视角及姿态条件下，其性能会受到很大的影响。事实上，人类辨识某一个人时，往往不是通过底层特征，更多的是通过高层的语义属性来主导判断的。比如，发型，衣服的款式类型，人体的高矮胖瘦等一些显著性信息。

与底层特征相比，中层特征[17][18][19][20][21][22]有如下优势：

1. 在跨摄像头的情况下，语义属性对行人的描述比底层特征更加鲁棒，因为其基本不受光照、视角和姿态的干扰。

2. 语义层面的描述更加符合人类的理解和需求，所以中层特征的设计相对底层特征要容易并且符合常理。

3. 中层特征可以不基于图片直接通过语言描述提取。底层特征必须基于图片提取。然而，有时候我们可能得不到需要查询的图片，这个时候，底层特征无能为力，但中层特征可以通过语言描述直接获得，在一定程度上弥补了其他方法的缺陷。

1.2.1.3 深度特征

无论是底层特征、中层语义特征还是它们之间的级联混合，本质上都属于手工设计特征，这样的特征设计方法依靠大量的人工先验和实验验证，而且针对不

同的使用场景和任务，往往需要设计对应的特征。近年来，随着深度学习的发展，越来越多的计算机视觉任务将繁重的特征表示环节交给深度神经网络来自动学习。深度学习方法也逐渐被使用在行人再识别中。然而，深度学习也有它的缺点，比如，在没有足够多数据的特定场景中，无法有效训练深度网络，这个时候，手工设计特征这样的无监督方法具有明显的优势。

目前在特征设计和选择方面已经存在大量的研究工作。2008 年，Gary 等[1]采用了 AdaBoost 算法来装配一组合适的特征表示人体图像，并通过级联的分类器训练学习出特征组合的权值。2010 年，Farenzena 等人[3]提出了对称性驱动的局部特征累加（Symmetry-Driven Accumulation of Local Features，SDALF）方法，先在垂直方向上把人体划分成三部分，再利用行人的对称性进行了水平分块，最终在 5 个区域块上提取 HSV 直方图和纹理特征。2013 年，Weishi Zheng 等人[9]基于行人大多在水平方向上发生视角变化而在垂直方向上基本不变的假设，将人体在垂直方向上分成 6 条带后提取颜色和纹理特征。2013 年，Rui Zhao 等[23]提出一种局部扰动的区域块之间计算距离和的方法，小块的权重由其显著性决定，相比于条带划分，这种方法更为精细，性能更高。2014 年，Giuseppe Lisanti 等[24]通过级联颜色直方图，HOG，LBP 等获得比较完备的图片特征描述，取得了不错的行人再识别效果。同年，Wei Li 等[25]利用当下流行的深度卷积神经网络（Deep Convolutional Neural Network，DCNN）框架自动提取特征来代替手工设计的特征，随后，涌现出许多基于深度网络的特征学习方法。2015 年，Shengcai Liao 等[6]提出一种取局部最大颜色通道和纹理特征（Local Maximal Occurrence，LOMO）来解决光照和视角变化并取得了良好的性能。2016 年，Tetsu Matsukawa 等[7]提出了分层高斯特征描述方法，融合 RGB, Lab, HSV, nRnG 颜色空间，形成最后的特征表示，取得了比较好的行人再识别性能。

1.2.2 距离度量

在计算特征之间的相似度时，距离度量尤为重要，相同的特征在不同的度量准则下将会产生巨大的差异。常规的欧氏距离、余弦距离等由于未考虑样本空间

的特性往往不能取得好的效果。通过训练标注样本得到一个符合样本空间特性的距离度量函数以其显著的效果得到研究者们的青睐。近年来，大量的距离度量方法被提出。

这些方法大部分都可以变换成马氏距离形式的距离度量函数：

$$d_M^2(x_i,x_j)=\|x_i-x_j\|_M^2=(x_i-x_j)^T M(x_i-x_j) \tag{1-1}$$

使用（1-1）可以计算样本之间的距离。其中M是一个半正定矩阵，是通过样本学习得到的参数。

2002 年，Eric Xing 等人[26]首次提出马氏距离形式的距离度量学习。在(x,y)类别数据的基础上：

$$S=\{(x_i,x_j \mid x_i \ and \ x_j \ belong \ to \ the \ same \ class)\} \tag{1-2}$$

$$D=\{(x_i,x_j \mid x_i \ and \ x_j \ belong \ to \ different \ class)\} \tag{1-3}$$

$$\begin{aligned} &\min_{M\in R^{m\times m}} \sum_{(x_i,x_j)\in S} \|x_i-x_j\|_M^2 \\ &\text{s.t } M\geq 0, \sum_{(x_i,x_j)\in D} \|x_i-x_j\|_M^2 \geq 1 \end{aligned} \tag{1-4}$$

基于上述两种样本对，和带约束的凸规划方程，可以学习到一个最优度量矩阵M，最小化相似样本对的距离，同时最大化不相似样本对的距离。2005 年，Weinberger 等人[27]提出了大间隔最近邻分类（Large Margin Nearest Neighbor，LMNN）距离度量算法。2011 年，Weishi Zheng 等人[28]将尺度学习算法的思想引入到行人再识别问题，其采用算法中的三元组形式的样本对，提出基于概率相对距离比较（Probabilistic Relative Distance Comparison，PRDC）的距离度量学习算法。2013 年，Sateesh 等人[29]将局部线性判别分析（Local Fisher Discriminant Analysis，LFDA）用于行人再识别问题。其在特征提取时先分别对不同类型的特征使用 PCA 降维，然后拼接成向量。在尺度学习时，该方法考虑局部样本点，而非赋予所有样本点同样的权重。2014 年，Fei Xiong 等[10]通过核方法将线性的局部线性判别分析方法映射拓展为非线性的基于核的局部线性判别分析（kernel Local Fisher Discriminant Analysis，kLFDA）方法。2015 年，Shengcai Liao 等[6]

提出一种叫跨视域二次判别分析（Cross-view Quadratic Discriminant Analysis，XQDA）的方法，作为贝叶斯人脸识别和 KISSME[30]方法的拓展。2015 年，Sakrapee Paisitkriangkrai 等人[31]提出了度量装配方法，通过融合多个基本度量使得最后结果在单模型基础上有所提升。2016 年，Li Zhang 等[32]提出了学习判别零空间的度量学习方法。2016 年，Peixi Peng 等[52]使用基于字典学习的跨数据集迁移学习的方法无监督地学习具有判别性的特征表示，具有一定实用价值。

这些方法在一定程度上解决了上面提到的光照、姿态、视角变化等问题，使匹配的鲁棒性更强。然而，监督学习方法虽然能取得良好的匹配性能，但是训练过程需要大量人工标注数据。这个条件极大的限制了其在实际场景中的运用。为了解决这个问题，一些学者提出了基于无监督学习的行人再识别方法[3][33][34][35]，这些方法通过使用大量无标签数据学习距离度量而不是通过人工标定的数据，这在一定程度上缓解了监督学习带来的问题。然而，相对于监督学习，通过无标签数据学习得到的距离度量通常不具有很强的判别能力，这导致其性能一般弱于最好的监督学习方法。但其所具有的可扩展性和实用性得到了人们广泛的认可。

1.3 本文研究内容

1. 本文针对监督方法需要标定训练数据的问题，提出了基于无监督的行人再识别算法框架。无监督学习方法具有实用性并适用于大数据情况。

2. 针对全局的距离度量在处理多样化的数据时存在性能损失的问题，提出了基于样本的局部度量学习算法。针对每个样本学习度量，充分考虑样本的特性。

3. 使用重排序方法优化初始排序。重排序考虑了 gallery 样本之间的相似性，从而能够优化初始排序。

1.4 本文的组织结构

第一章，绪论。介绍了行人再识别的研究背景和价值，当前的研究进展以及本文的主要研究内容，并对行人再识别的主要技术进行概述。

第二章，相关工作与技术。分析了行人再识别常用的特征表示以及距离度量方法。介绍了无监督学习方法和重排序方法在行人再识别中的最新进展。

第三章，基于局部度量学习的行人再识别算法。本章提出了基于无监督学习的行人再识别算法框架。不同于监督式学习需要大量成对标注的训练样本，无监督学习只需要无标注的样本数据。

第四章，基于重排序的行人再识别算法。本章提出了一种简单有效的重排序算法。作为一种优化方法，重排序可以显著减少相似样本的干扰从而提高最终排序的正确率。

第五章，总结与展望。总结了本文的研究工作，分析了本文工作的不足、行人再识别目前存在的难点和问题，展望了未来的研究方向。

第二章 相关工作与技术

上一章简要介绍了研究背景和意义、行人再识别领域国内外进展、本文主要研究内容和组织结构。本章将详细介绍本研究涉及的相关理论和技术，作为后续章节展开的背景知识和理论基础。

2.1 特征描述

特征描述普遍存在计算机视觉的各类任务中，良好的特征设计是决定算法最终性能的关键因素。在目标检测、分类等任务中，基于深度学习的特征具有强大的表示能力，这使得后续只需要一个比较简单的分类器就能够取得好的结果，这大大减轻了分类器的负担，简化了整个系统的复杂度。针对行人再识别中任务的特殊性，研究者们提出了一些行之有效的特征描述子。

2.1.1 LOMO 特征

LOMO（Local Maximal Occurrence）是由 Shengcai Liao 等[6]在 2015 年提出的一种对光照和视角变化鲁棒的特征。LOMO 特征主要由三个策略构成：

1. MSR（Multi Scale Retinex）算法[36]。颜色是一种重要的描述人体图片的特征。然而，不同时间，不同摄像头视角下的光照条件差异巨大，这导致如果直接使用原始图片的颜色特征往往无法取得令人满意的效果。针对这一问题，作者使用了 MSR 算法。MSR 可以同时保持图像高保真度和对图像的动态范围进行压缩，它能在一定程度上实现图像的色彩增强、颜色恒常性、局部动态范围压缩、全局动态范围压缩。该算法在处理跨摄像头图片时对光照和颜色有较好的一致性。

2. 颜色直方图和 SILTP（Scale Invariant Local Ternary Pattern）[37]。颜色直方图是图片特征描述时常用的信息，作者采用了 HSV 颜色直方图。对于纹理信息，由于 LBP 特征对图片噪声敏感，作者未直接使用 LBP 特征，而是使用了一种比 LBP 对尺度和图片噪声更鲁棒的描述子 SILTP。

3. 滑动窗口。为了处理视角变化带来的问题，作者使用大小为 10*10，步长为 5 的滑动窗口为基本单元提取两个尺度的 SILTP 直方图和 HSV 直方图，然后对统一水平高度的窗口特征取最大的模式分量。为了进一步考虑多尺度信息，作者又对图片进行了三层金字塔下采样，然后对每一层进行相同的特征提取操作，最后，对所有的模式进行级联，取 log 变换和归一化变化得到最终的特征。

LOMO 特征在设计的过程中充分考虑了光照和视角变化这两个行人再识别中普遍存在并且影响巨大的因素，有针对性地对常用的颜色特征和纹理特征进行预处理或者修改提升，然后进行滑窗提取。实验表明此特征在大多数行人再识别任务中均能取得比较好的效果。

2.1.2 WHOS 特征

WHOS（Weighted Histograms of Overlapping Stripes）是由 Giuseppe Lisanti 等[24]在 2014 年提出的一种特征描述子。该特征充分使用了比较著名的各种底层特征：颜色直方图（HS、RGB、Lab），HOG（Histogram of Oriented Gradient，方向梯度直方图），LBP（Local Binary Pattern，局部二值模式）。其中，HOG 特征是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子。它通过计算和统计图像局部区域的梯度方向直方图来构成特征，能比较好地描述物体的边缘和纹理信息。HOG 特征结合 SVM 分类器已经被广泛应用于图像识别中，尤其在行人检测中获得了极大的成功。LBP 是一种用来描述图像局部纹理特征的算子，它具有旋转不变性和灰度不变性等显著的优点，在纹理分类、人脸识别领域被广泛运用，取得了不错的效果。通过级联颜色直方图（HS、RGB、Lab）、HOG 和 LBP 所形成的特征能够较为全面地描述人体目标所带有的信息。在视角变化的处理上，此方法也通过将图片划分成水平条带然后提取相应模式的特征向量。在背景压制处理上，通过使用各向异性的高斯核权重来减小人体周围背景的影响。WHOS 特征设计的细节如图 2-1 所示。

原图　Epanechnikov 核权重掩膜　分块　HOG单元

图 2-1 WHOS 特征部分细节

2.1.3 深度特征

Hinton 等人[38]在 2006 年提出了深度学习的概念和改进的神经网络模型训练方法，这打破了早期 BP 神经网络发展的瓶颈。Hinton 在论文中提出了两个观点：（1）多层人工神经网络模型有很强的特征学习能力，深度学习模型学习得到的特征数据对原始数据有更本质的代表性，这将大大便于分类和可视化问题；（2）对于深度神经网络很难训练达到最优的问题，可以采用逐层训练方法解决。将上层训练好的结果作为下层训练过程中的初始化参数。在这一文献中深度模型的训练过程中逐层初始化采用无监督学习方式。

2012 年，Hinton 课题组为了证明深度学习的潜力，首次参加 ImageNet 图像识别比赛，其通过卷积神经网络 AlexNet[39]（如图 2-2 所示）一举夺得冠军，且大幅度超过第二名（SVM 方法）的分类性能。经过该比赛，CNN 吸引到了众多研究者的注意。随后各种不同结构和深度的神经网络如雨后春笋般涌现，不断刷新着各种计算机视觉任务的性能。其中比较有代表性的如 VGG16[40]、GoogleNet[41]、ResNet[42]等。从比较有代表性的卷积神经网络 AlexNet[39]开始，新的网络结构隐层数量不断加深，以取得更好的非线性效果，然而网络层数的增加也带来了一系列问题比如梯度弥撒，计算复杂度增加等。也有一些学者试图通过改善网络结构的方法，在不增加网络参数或者计算复杂度的情况下提高网络的性能，如以 GoogleNet[41]为代表的一系列网络结构。

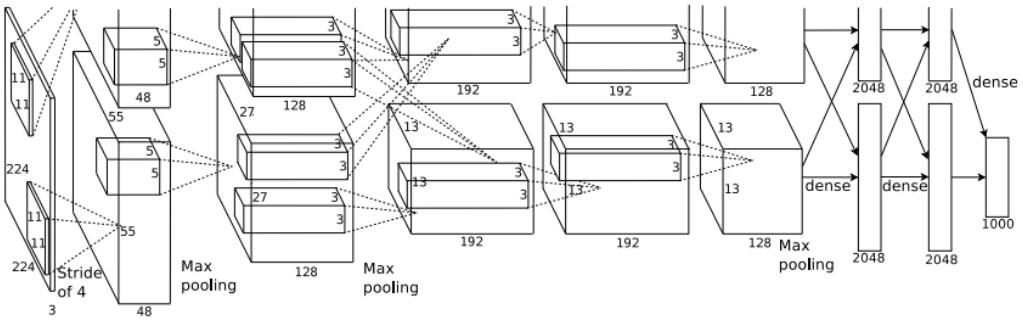


图 2-2 Alex-Net 模型结构

典型的卷积神经网络主要由卷积层、池化层和全连接层组成。为了改进全连接造成网络参数随着输入维度增加呈指数次方增长的问题，卷积层采用了局部连接和权值共享的方法，即输出的特征图中每一个点只与输入特征图的局部有关，输出特征图中每个点都是输入特征图和同一个卷积核卷积计算得到的。为了进一步减少参数，降低计算复杂度并且使特征具有一定的平移不变性，在一些卷积层后面接入了池化层，这样即使图像经过了一些平移和缩放，图像的输出特征仍然可以保持不变，增强了特征的鲁棒性。为了增强网络的非线性变换能力，控制输出特征的维度，在网络的最后面一般会接入几层全连接层。当然，最新的研究表明这不是必须的，而且全连接层一般包含大量的参数，一定程度上造成了网络的复杂度。

相较于手工设计的特征，深度网络学习得到的特征表示不需要太多的人工先验，而且一般来讲网络的深度越大，参数越多，所需的训练数据越多，这样得到的模型所对应的特征空间就越完备。所以在目标检测和分类任务中，深度学习取得了绝对的性能优势。然而，在行人再识别任务中，无法直接使用这些性能优秀的网络模型，原因在于：（1）目前的监控网络中的摄像头数量巨大，对应着海量的视频数据，加上人的行为具有随机性，这给标注带来了困难，导致大规模的标注数据集的获得比较困难。深度神经网络的训练需要大量的标注数据。（2）监控摄像头中采集到的低分辨图片在经过多层卷积、池化后特征维度太小，导致最终性能的损失。因此，必须设计适合于行人再识别任务的特殊网络才能获得较好

的特征表示。2014 年 Wei Li 等[25]首先提出了基于深度卷积神经网络的行人再识别方法 DeepReID。随后，基于深度学习的方法受到了研究者们的广泛关注，各种新的网络结构被提出，大大推动了深度学习在行人再识别领域的发展。

2.2 距离度量

行人再识别本质上是要计算样本之间的相似度或者距离，然后根据相似度对样本进行排序，进而找到与查询样本属于同一个人的样本图像。在将图片转化成特征之后，如何计算特征之间的距离是我们需要考虑的问题。仅仅使用简单的欧式距离或者余弦距离往往无法兼顾到不同摄像头场景之间存在的光照、视角、分辨率等的差异性和多变性。通过先验人工设计特征权重加入这些存在的因素也不太现实，目前研究者们采用的普遍方法是通过采集样本数据集然后通过数据来学习特定场景的距离度量。学习距离度量本质上是学习一种映射，使得原始特征经过映射后在新的子空间里可以有效地和不同类别区分开来，即最大化类间距离，最小化类内距离，度量学习在图像检索和分类、人脸识别等领域有广泛的运用。如图 2-3 所示，为了更加直观，图中特征向量使用对应图片表示。接下来介绍两个比较典型的度量学习算法。

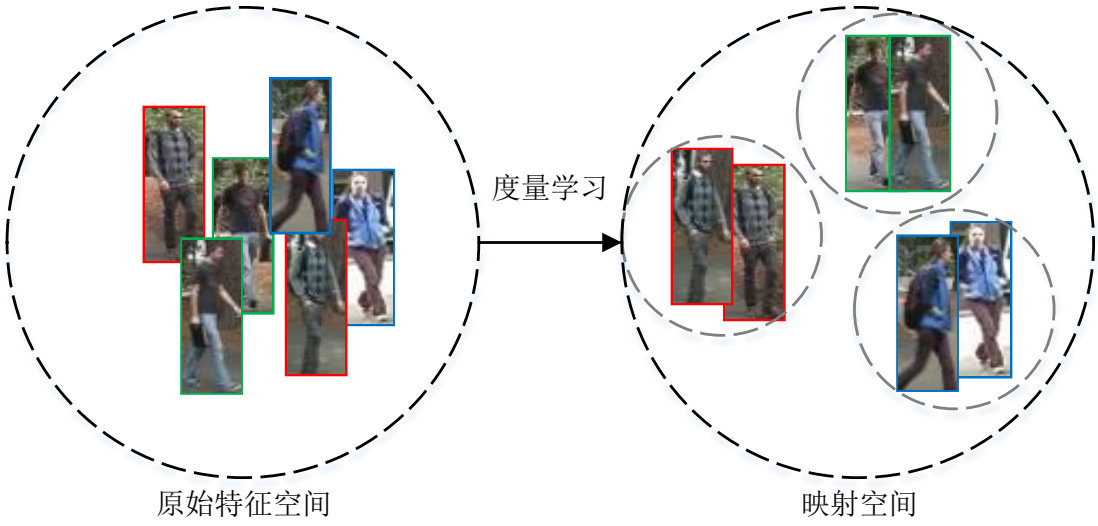


图 2-3 不同空间特征分布

2.2.1 XQDA 算法

XQDA 度量算法是在 KISSME 算法的基础上加入了具有辨别性的降维算法。所以我们先介绍一下 KISSME 算法。该算法从统计推断的角度，使用 log 似然比检验的方法判断给定的样本对是否是同一个行人。首先使用特征向量x_i和x_j分别代表两个样本。假设H_0代表特征向量对不相似，即x_i和x_j代表两个不同的人，H_1代表特征向量对相似，即x_i和x_j代表同一个人的两个样本，并定义$x_{ij} = x_i - x_j$为样本对的差分向量，则H_0和H_1的后验概率比的对数形式为：

$$\delta\left(x_{ij}\right)=\log\left(\frac{p\left(x_{ij}\mid H_0\right)}{p\left(x_{ij}\mid H_1\right)}\right)=\log\left(\frac{f\left(x_{ij}\mid \theta_0\right)}{f\left(x_{ij}\mid \theta_1\right)}\right) \tag{2-1}$$

其中，$\delta(x_{ij})$为正值时，表示x_i和x_j是同一个行人的样本；$\delta(x_{ij})$为负值时，表示x_i和x_j不是同一个人的样本。在零均值高斯分布的假设下，差分向量x_{ij}在假设H_0和H_1下的后验概率可以表示为：

$$p\left(x_{ij}\mid H_E\right)=\frac{1}{(2\pi)^{d/2}\left|\sum_E\right|^{1/2}}e^{-\frac{1}{2}x_{ij}^T\sum_E^{-1}x_{ij}} \tag{2-2}$$

$$p\left(x_{ij}\mid H_I\right)=\frac{1}{(2\pi)^{d/2}\left|\sum_I\right|^{1/2}}e^{-\frac{1}{2}x_{ij}^T\sum_I^{-1}x_{ij}} \tag{2-3}$$

其中，Σ_E和Σ_I分别是假设H_0和H_1下差分向量x_{ij}对应的协方差矩阵，d 是特征向量的维数。令N_E表示不相似特征向量对的个数，N_I表示相似特征向量对的个数，则差分向量x_{ij}在假设H_0和H_1下的协方差矩阵Σ_E和Σ_I的估计为：

$$\sum_E=\frac{1}{N_E}\sum_E\left(x_i-x_j\right)\left(x_i-x_j\right)^T \tag{2-4}$$

$$\sum_I=\frac{1}{N_I}\sum_I\left(x_i-x_j\right)\left(x_i-x_j\right)^T \tag{2-5}$$

由式（2-1）到式（2-5）可得：

$$\delta\left(x_{ij}\right)=\frac{1}{2}x_{ij}^T\left(\sum_I^{-1}-\sum_E^{-1}\right)x_{ij}+\frac{1}{2}\log\left(\left|\sum_I\right|\right)-\frac{1}{2}\log\left(\left|\sum_E\right|\right) \tag{2-6}$$

这里的常数项不影响之后距离的相对值，因此可以舍去，并简化为：

$$\delta\left(x_{ij}\right)=x_{ij}^{T}\left(\sum_{I}^{-1}-\sum_{E}^{-1}\right)x_{ij} \tag{2-7}$$

最后，令$\widehat{M}=\Sigma_{I}^{-1}-\Sigma_{E}^{-1}$，将$\widehat{M}$重投影到半正定矩阵锥面，最终得到反映对数似然比测试性能的马氏距离度量矩阵：

$$d_{M}^{2}\left(x_{i},x_{j}\right)=\left(x_{i}-x_{j}\right)^{T}M\left(x_{i}-x_{j}\right) \tag{2-8}$$

因此，可以通过对$\widehat{M}$进行特征值分析来估计 M。

然而，通常情况下原始的特征向量维度是巨大的，需要对其进行降维处理。KISSME 算法在对特征数据进行降维时直接使用了主成分分析（PCA）方法，主成分分析在降维的过程中没有考虑数据的类别信息，这在一定程度上会降低特征的分类能力。因此，为了在对高维特征数据降维的过程中考虑类别信息，Shengcai Liao 等[6]提出了 XQDA 算法。

由于同类和异类样本差x_{ij}都符合零均值分布，所以传统的线性判别分析不在适用。然而，它们的协方差可以被用来区分这两个类，所以在降维的过程中可以最大化异类样本差的协方差和同类样本差的协方差比值：

$$J\left(w\right)=\frac{w^{T}\sum_{E}w}{w^{T}\sum_{I}w} \tag{2-9}$$

$$\max_{w}=w^{T}\sum_{E}w,\ s.t.\ w^{T}\sum_{I}w=1 \tag{2-10}$$

上式可以通过类似线性判别分析的特征值分解方法取最大的几个特征值对应的特征向量组成最终的降维矩阵，实现特征的降维。最终的距离表达式如下所示：

$$d_{W}\left(x,z\right)=\left(x-z\right)W\left(\sum_{I}^{\prime-1}-\sum_{E}^{\prime-1}\right)W^{T}\left(x-z\right) \tag{2-11}$$

XQDA 算法在特征降维的过程中考虑了距离度量信息，所以能够在匹配的过程中取得比 KISSME 算法好的性能。

2.2.2 判别性零空间学习算法

判别性零空间（Discriminative Null Space）学习算法是由 Li Zhang 等[32]在 2016 年提出的一种度量学习算法。众所周知，大多数的度量学习算法意在学习

马氏距离形式的度量矩阵。如果有一个特征的线性映射 W 使得原始特征x_i被映射为$y_i = W^T x_i$，那么y_i和y_j之间的欧式距离可以写为$\|y_i - y_j\|^2 = (x_i - x_j)^T A(x_i - x_j)$，其中$\mathrm{A} = W^T W$是一个半正定矩阵。这意味着，学习一个判别性的子空间然后计算欧氏距离等效于在原特征空间直接计算判别性的马氏距离。

在介绍具体算法之前，先说明一下零空间的定义：已知 A 为一个$m \times n$矩阵。A 的零空间（null space），又称核（kernel），是一组由下列公式定义的 n 维向量：$\ker(A) = \{x \in R^n : Ax = 0\}$，即线性方程组$Ax = 0$ 的所有解 x 的集合。

一般地，我们可以仿照类似线性判别分析的方法，通过最大化：

$$J(w) = \frac{w^T S_b w}{w^T S_w w} \tag{2-12}$$

来得到这个具有判别性的零空间 W。其中，S_b是类间散度矩阵，S_w是类内散度矩阵。显然，如果S_w是非奇异的，$C-1$个特征向量$w_1, \ldots, w_{c-1}$可以通过$S_w^{-1} S_b$最大的$C-1$个特征值求得，将它们作为列向量就能构成一个$C-1$维的 W 映射矩阵。然而，在数据集规模比较小时，S_w是奇异矩阵无法求逆，通常的做法是用 PCA 降维或者加小的正则项，但这样做会受到性能的损失。所以作者使用了 NFST[43]的方法来求映射矩阵 W。具体公式如下：

$$w^T S_w w = 0 \tag{2-13}$$

$$w^T S_b w > 0 \tag{2-14}$$

定义全散度矩阵$S_t = S_b + S_w$，记S_t和S_w的零空间为：

$$Z_t = \left\{ z \in R^d \mid S_t z = 0 \right\} \tag{2-15}$$

$$Z_w = \left\{ z \in R^d \mid S_w z = 0 \right\} \tag{2-16}$$

它们的正交补为$Z_t^{\perp}$和$Z_w^{\perp}$。因为S_b是非负定的，为了同时满足公式（2-13）和（2-14），特征向量必须满足：

$$\mathrm{w} \in \left(Z_t^{\perp} \cap Z_w \right) \tag{2-17}$$

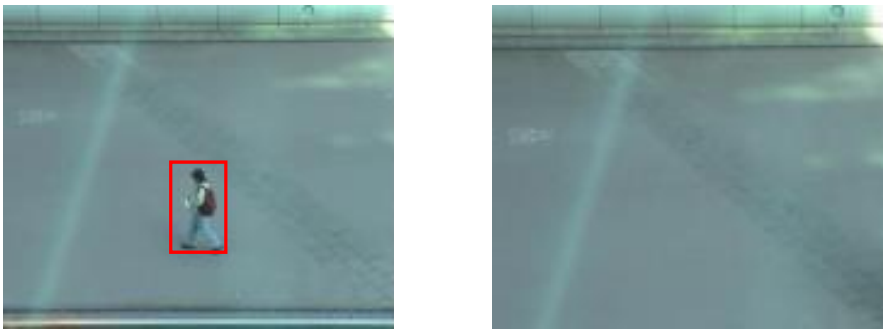
具体的求解过程可以参见相关论文。

作者通过改进线性判别分析使之适用于数据集规模较小的场景，很好地解决了行人再识别的问题。

2.3 基于无监督学习的方法

本节将介绍无监督学习在行人再识别领域的运用。首先介绍一下无监督学习的背景和概念，然后列举近年来提出的基于无监督学习的行人再识别算法。

依据训练数据是否需要标注，以及标注的类型，我们一般可以将学习方法分为监督学习，弱监督学习和无监督学习。监督学习广泛使用于各类机器学习任务中，它通过使用标注样本学习模型的参数使其达到要求的性能。然而，这种方法只适用于数据量不大的情况，随着大数据时代的来临，大部分数据都是无标记样本，标定数据成本巨大。因此，研究者们开始探索如何使用无标记样本来学习模型，于是就有了弱监督学习和无监督学习。弱监督学习只需要一些弱标注信息即可进行训练，具体的，比如我们不需要知道一张图片中目标的具体位置，具体个数，只需要知道这张图片中包含有目标就把它当做正例样本，如图 2-4 所示（$Y=1$表示正例，$Y=0$表示反例；L 表示位置信息，x 和 y 表示目标中心点位置，w 和 h 表示目标宽和高），我们知道图片中有人但不知道具体位置。还有一种被称为半监督的弱监督学习，它使用少量的已标注样本和大量的无标注样本训练模型，在一定程度上解决了监督方法存在的问题。



强标注：Y=1；L=(x,y,w,h)　　强标注：Y=0

弱标注：Y=1　　弱标注：Y=0

图 2-4 正负样本标注

针对标注问题，最终极的解决方法就是无监督学习方法。直观的讲就是我们不告诉计算机怎么做，完全让它自己去学习如何做事情。一个典型的应用场景就是聚类，因为事先没有给出数据的任何类别信息，完全通过数据的分布将数据聚成几类。然而，在其他任务中比如检测、分类等还没有非常有效的无监督学习的解决方案，有待于进一步探索。

在行人再识别中，字典学习和稀疏编码被许多研究者用来进行无监督的学习特征子空间。其中比较具有代表性的是 Elyor Kodirov[35]等提出的基于迭代拉普拉斯正则的字典学习方法。此算法想要学习构建一个字典$D \in R^{k\times m}$，将原本 n 维的特征 x 映射到 k 维子空间，并使映射后的特征 y 具有稀疏特性，然后求解子空间特征向量之间的余弦距离。原始的字典学习公式：

$$\left(D^*,Y^*\right)=\arg\min_{D,Y}\left\|X-DY\right\|_F^2+\alpha\left\|Y\right\|_1 \qquad (2\text{-}18)$$

只考虑到了重构误差和一范数正则最小，没有考虑到重构的稀疏特征在跨摄像头匹配的判别信息。为了解决这个问题，作者引入了图拉普拉斯正则项，公式（2-18）可以重写为：

$$\left(D^*,Y^*\right)=\arg\min_{D,Y}\left\|X-DY\right\|_F^2+\alpha\left\|Y\right\|_1+\beta\sum_{i,j}^{m}\left\|y_i^a-y_j^b\right\|_2^2W_{ij} \qquad (2\text{-}19)$$

其中，β 是新的正则项的权重，$W\in R^{m\times m}$代表跨摄像头特征之间的关系矩阵。由于没有标注信息，真正的特征之间的关系信息无法知道。因此这里的 W 代表一种软的特征关系。具体的，如果原始特征x_i^a在x_j^b的 k 近邻中，则$W_{i,j}=\left((x_i^a)^Tx_j^b\right)/\left(\left\|x_i^a\right\|\left\|x_j^b\right\|\right)$，否则，$W_{i,j}=0$。在给定的拉普拉斯正则项中，作者假设了视觉上相似的样本之间很有可能是同一个人，进而它们子空间的特征向量之间的距离也应该很小。直观地讲，就是在原始空间中距离相近的特征在子空间中我们需要保证它们之间的这种关系。

通过学习得到的字典就可以对原始特征进行稀疏重构，然后计算重构特征之间的余弦距离作为样本之间的距离。该方法通过无标注样本学习字典，一定程度上解决了标注问题。但该方法存在两个问题：（1）在字典学习的过程中原始特

征空间中相似的样本中大部分不是同一类别，这会对下一次迭代和最终性能造成影响。（2）计算复杂度较大，和数据量规模不是线性关系，无法适用于真实场景。

2016 年 Peixi Peng 等[52]在图拉普拉斯正则迭代更新学习的基础上引入了跨数据集的迁移学习。该方法同时使用标注的源数据集和无标注的目标数据集，使用非对称多任务字典学习模型从训练数据中学习得到具有场景不变性和判别性的信息。该方法进一步提升了无监督行人再识别的性能，但其缺点也显而易见：（1）基于字典学习，在数据扩展、训练复杂度等方面具有一定局限性。（2）需要大量的外部标定数据来协同训练，代价高。

无监督行人再识别方法虽然具有比较高的实际使用价值，吸引了大批研究者投身于此，也取得了上述这样比较优秀的成果，但是目前的进展离实际运用还有很长的距离，需要继续探索更好的解决方案。

2.4 基于重排序的方法

重排序即对初始排序进行再次排序，使得最终结果变得更好。重排序在搜索、个性化推荐等领域被广泛运用。比如在个性化推荐中，当我们通过协同过滤等算法得到了初始的推荐排序后，一般直接将这个结果呈现给用户也没什么问题，但为了提高用户的满意度，通过考虑商品的新颖性、多样性、用户反馈等因素对初始结果进行重排序可以达到比较好的效果。在行人再识别中，通过计算每个查询样本和图片库中每个图片的相似度可以得到关于这个查询样本的初始排序，进一步地，如果我们考虑其它一些信息，比如，图片库中图片之间的相似度关系，然后进行重排序，最终结果会在初始排序的基础上有所提高。

Jorge Garcia 等[44]在 2015 年提出了基于上下文信息分析的排序优化算法。一般地，在查询样本的初始排序中的前几个样本和它具有相当高的相似度，那些排序靠前的负样本（和查询不是同一个人的样本）之所以排序靠前是因为它们的特征中含有与查询样本相似的成分。那么，通过除去这些容易造成混淆的特征成分理论上可以减少错误的排序。如图 2-5 所示，查询样本的前几个匹配中某些特征

成分（如橘黄色的上衣）造成了匹配的混淆，使得排序 1 造成了错误匹配（红色虚线框为正确匹配），因此在这种情况下，我们希望在特征空间中去掉造成混淆的相关特征，这样可以降低排序靠前的匹配错误的概率。

图 2-5 错误匹配示例

具体地，（1）首先确定查询样本x_p初始排序前 k 个样本作为处理的范围。k 值按如下方式确定：通过将初始排序按与查询样本的相似度进行聚类，取平均相似度最大的那个聚类，然后在这个聚类排序中找到其中与查询样本的相似度差最大的两个相邻样本，在这个最大相似度差前面的样本作为考虑的对象，这 k 个样本称为 content 信息C_p^{cn}。（2）针对查询样本的这 k 个排序靠前的样本，我们可以给每一个找到对应的 content 信息。将所有这些 content 信息放到一起，然后去取出现次数最多的 K 个组成查询样本的 context 信息C_p^{cx}。（3）将查询样本和它对应的 content 信息和 context 信息放到一起组成$D_p = \{x_p, C_p^{cn}, C_p^{cx}\}$，作者认为，在$D_p$中所包含的前 j 个主成分中包含了主要的混淆正确匹配的成分，所以通过主成分分析将它除去。具体地，通过公式：

$$D_p^* = D_p - PP^T D_p \qquad (2\text{-}20)$$

其中，p 代表D_p前 j 个主成分。通过计算处理之后的特征之间的距离在重排序时可以取得比初始排序更好的匹配性能。

2.5 本章小结

本章介绍了行人再识别领域的相关工作和技术。首先，阐述了行人再识别任务中两个关键内容，特征表示和度量学习的相关理论和现有工作；然后，结合本

研究的特点，介绍了无监督学习的概念和目前无监督行人再识别方法的发展情况；最后，介绍了重排序算法的作用和相关工作。

第三章 基于局部度量学习的行人再识别算法

上一章介绍了本研究的相关工作与技术背景。本章从具体的问题入手，首先描述和分析问题，然后提出整体算法框架，详细分析具体环节和算法的理论依据，最后进行实验结果展示和分析。

3.1 问题描述及分析

在视频监控领域中，行人再识别技术作为自动化监控手段相比于人工监视方法有着明显的优势。然而，现有的大多数行人再识别主要基于监督学习方法。和其他任务比如行人检测相比，行人再识别任务标注训练数据的代价是巨大的。比如：（1）行人检测在某个场景中的标定数据训练得到的模型，一般也适用于其他场景的行人检测，而行人再识别由于学习到的模型参数是针对特定场景的，在其他场景下会有很严重的性能损失。（2）行人检测训练数据的标定相对比较简单，只需要在一张图片中给出正例的位置，而行人再识别需要标注的是跨摄像头行人样本对，设想一下在两个相聚甚远的摄像头对中进行标注，在一个摄像头中出现的行人大概率不出现在另一个摄像头中，这给标注带来了巨大的挑战。

图 3-1 行人再识别标注示例

另一问题，现有的大部分度量学习方法学习的是全局的度量，即在整个训练

集上学习得到一个度量。然而，训练集是不完备的，在实际场景中有太多复杂多变的样本，导致全局的度量无法处理所有可能出现的情况。如图 3-2 所示，度量方法 1 能正确匹配橘黄色上衣查询样本，但错误匹配了白色上衣查询样本；度量方法 2 正好相反。说明全局的度量方法具有局限性。当然，针对这个问题，有些研究者提出了模型融合方案，将多个不同的子模型通过 adaboost 等方法融合为一个模型，充分发挥了各个子模型的优点。但这样的模型往往比较复杂，在实际运用中存在子模型难获取、计算复杂度高等问题。

图 3-2 全局度量的表现

因此，以上两个问题一定程度上限制了行人再识别技术在实际场景的应用。本研究针对两个问题提出了基于无监督局部度量学习或者叫基于样本的度量学习（SBML）的行人再识别算法。所谓无监督学习，即我们只需要大量的无标注训练样本，所谓局部度量，即我们为每一个查询样本学习一个度量矩阵，然后进行查询排序，充分考虑每一个样本的特殊性。

3.2 局部度量学习算法框架

基于局部度量学习的无监督行人再识别框架是本研究的主要部分。本节先给出整体算法框架，然后对各个组成部分进行详细说明。

3.2.1 算法框架

本算法框架主要由行人检测、特征表示、度量学习、匹配排序等模块组成，如图 3-3 所示。其中，度量学习是算法的核心部分。

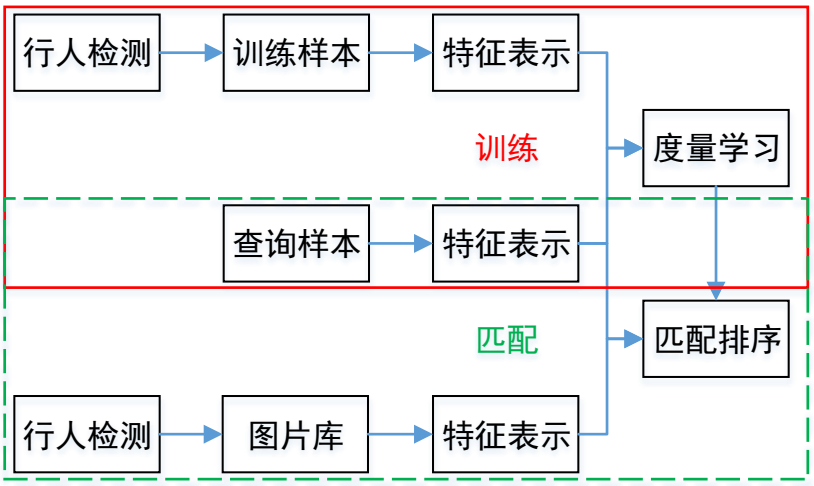


图 3-3 算法框架

1. 行人检测。该部分主要是通过检测算法从原始的监控视频中获取行人目标，去除背景。

2. 特征表示。该部分将获得的人体图片描述为特征向量，方便计算机处理及后续的训练和匹配。

3. 度量学习。学习距离度量，即特征的映射空间。

4. 匹配排序。通过学习得到的度量进行相似度计算，然后根据相似度大小进行排序。

3.2.2 行人检测

行人检测主要用来获取行人目标。由于监控视频中的图片包含整个场景，其中，我们感兴趣的目标只是行人，其他大部分属于背景信息，如果不经过处理直

接使用视频帧进行行人再识别，效果一般无法符合期望，常用的做法是获得行人目标区域后进行再识别。早期使用人工获取目标区域的方法虽然准确度高，但成本高，效率低，无法规模化拓展。现在一般使用行人检测算法来获取目标。然而，现有的行人检测算法，即使是性能最好的基于深度学习的方法，也会存在误检和漏检。因此，在实际使用时需要一定的人工干预。所以，检测算法和人工结合，以算法为主，人工筛选为辅的方法具有比较好的效果。

1. 训练样本的获取。监督方法需要获取跨摄像头的样本对作为训练数据，所以在使用检测算法获取每个摄像头下的行人图片后，还需要人工进行匹配，获取训练样本对。由于本研究使用的是无监督学习算法，不需要跨摄像头的样本对标注。因此，只需要使用检测算法如 DPM[45]自动获取监控视频中的人体图片作为负样本，不需要人工关联步骤。

2. 图片库数据获取。在匹配阶段之前，我们需要获得某个摄像头下一定时间内的行人图片库作为查找的范围，这一步可以直接使用检测算法获得。

实验中我们使用 DPM 检测视频中的行人目标。训练样本获取阶段，需要尽可能降低误检率，因为我们的目标是获得质量较好的训练样本图片。在图片库数据获取阶段，需要降低漏检率，因为我们需要确保查询图片对应的目标不被漏检。行人检测结果如图 3-4 所示。

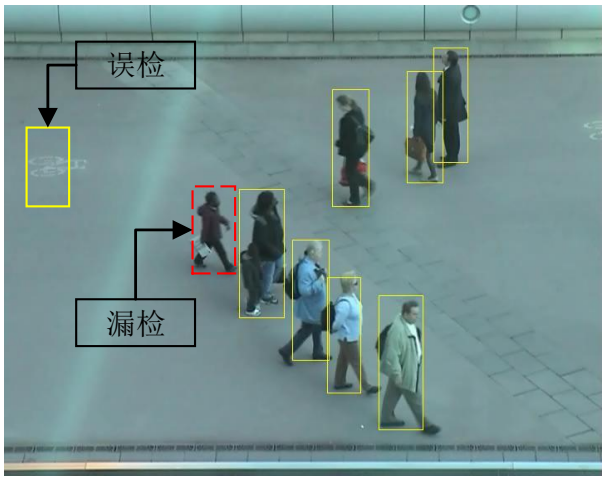


图 3-4 行人检测结果

3.2.3 特征表示

获得人体图片后，接下来需要将图片进行特征提取，向量化表示。为了进行相似度比对，直观上颜色、纹理、边缘等属性都是比较好的选择。实验证明，这些特征都能在一定程度上起到鉴别行人的作用。所以，我们充分利用这些特征，级联成最终的特征。人体图片样本中行人基本都是直立的，但由于视角不同，外貌会有所差异，针对这一问题，我们将图片分成不同的水平条带，然后以此为单位进行特征提取。本实验中，我们使用 WHOS 特征，WHOS 特征已在第二章中详细介绍，这里不再重复。

3.2.4 度量学习

度量学习的目的在于学习一个映射空间，使得原始特征在这个映射空间中能最大化类间距离，最小化类内距离。之所以有了原始的特征空间后还需要学习映射空间，是因为原始特征表示方法是一种无监督的方法，它没有考虑数据之间的类别关系。因此，通常的做法是通过度量学习获得一个具有判别性的映射空间。

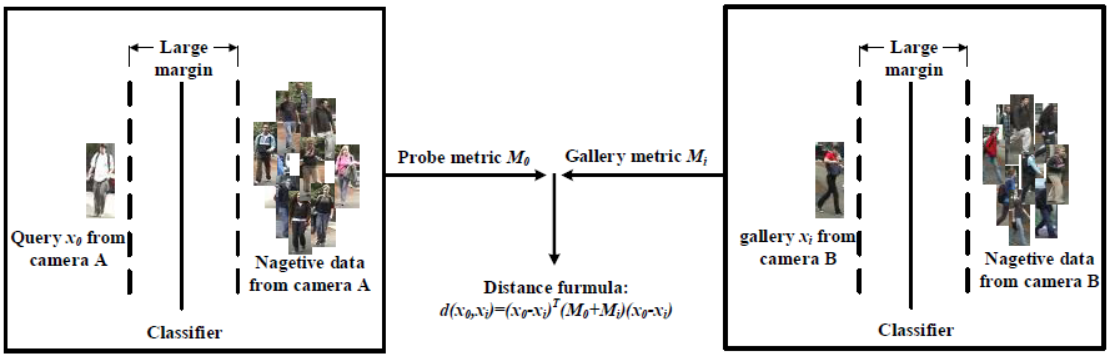


图 3-5 局部度量学习算法示意图

标准的度量学习方法是学习一个全局的马氏距离度量：

$$d(x,y)_M^2=(x-y)^T M(x-y) \qquad (3\text{-}1)$$

其中，x和y表示行人样本特征向量（如 color naming[46]，GoG[7]或 LOMO[6]等特征，或者是多种特征的组合）。

然而，在无监督学习框架下，我们考虑以查询样本为正样本，其他训练样本为负样本的学习方法。基于样本的无监督度量学习算法（如图 3-5 所示）是通过最大化查询样本x_0和反例样本集$x_1,\dots,x_n$中距离最小的样本的距离，从而学习得到一个针对于x_0的局部度量矩阵M：

$$M=\arg\max_{M\succ=0}(\min_{1\le i\le n}(x_i-x_0)^T M(x_i-x_0)) \tag{3-2}$$

其中，反例样本采集自和查询样本相同的摄像头场景，并且与查询样本非同一个行人。这样的无标定反例样本可以通过目标检测相关算法（DPM 等）自动获取。公式（3-2）是一个无约束的优化问题，我们归一化式（3-2），然后将方程重写成具有不等式约束的优化问题形式：

$$\begin{aligned} &M(x_0)=\arg\min_{M\succ=0}\frac{1}{2}\|M\|_2^2 \\ &subject\ to:\ (x_i-x_0)^T M(x_i-x_0)\ge c\ \ \forall i\in\{1,\dots,n\} \end{aligned} \tag{3-3}$$

其中，c为一个常数，为了计算方便，在本实验中我们将其设置为 2。同时，我们定义（详见后续合理性证明）：

$$\begin{aligned} &\tilde{x}_i=x_i-x_0 \\ &M=yy^T \end{aligned} \tag{3-4}$$

然后，公式（3-3）中的不等式约束可以重写为二次核函数的形式（黑点表示内积运算），如下：

$$\begin{aligned} &(x_i-x_0)^T M(x_i-x_0)=\tilde{x}_i^T yy^T\tilde{x}_i=\tilde{x}_i\tilde{x}_i^T\bullet yy^T \\ &=\varphi(\tilde{x}_i)\bullet\varphi(y)=k(\tilde{x}_i,y) \end{aligned} \tag{3-5}$$

同时，对于x_0定义$y_0=-1$且对于$x_i\ (1\le i\le n)$ 定义$y_i=1$，则公式（3-3）可以重写成 SVM 的形式，如下：

$$\begin{aligned} &M(x_0)=\arg\min_{M\succ=0}\frac{1}{2}\|M\|_2^2 \\ &subject\ to:\ (\langle M,\varphi(\tilde{x}_i)\rangle-1)\ge 1\ \forall i\in\{1,\dots,n\} \end{aligned} \tag{3-6}$$

基于公式（3-6），我们可以看到原问题等价于一个带核函数的 SVM 问题，因此便可以使用二次规划求解方法对其进行有效求解。

最后，距离度量M可以重写为：

$$M = \sum_{i=0}^{n} \alpha_i y_i \varphi(\tilde{x}_i),\ \alpha_i \geq 0 \tag{3-7}$$

得到x_0对应的度量M_0后，我们就可以计算x_0和 gallery 集中样本$x_1^g, \ldots, x_m^g$之间的距离。同样的策略，我们可以计算 gallery 集中的每一个样本并获取所其对应的度量矩阵M_i，最终的距离表达形式为：

$$d(x_0, \mathrm{x}_i)^2 = (x_i - x_0)^T (M_0 + M_i)(x_i - x_0) \tag{3-8}$$

合理性证明：由于$\varphi(x) \succcurlyeq 0$，$\varphi(\tilde{x}_0) = \varphi(0) = 0$，且 $y_i = 1 (i \geq 1)$，因此，我们可以得到：

$$M = \sum_{i=0}^{n} \alpha_i y_i \varphi(\tilde{x}_i) = \sum_{i=1}^{n} \alpha_i \varphi(\tilde{x}_i) \succ= 0 \tag{3-9}$$

因此公式（3-3）中的度量M是半正定的，所以可以被合理的写成公式（3-4）的形式。由上述公式可以分析得到，与现有的基于字典学习和稀疏特征编码的无监督度量学习方法不同，我们的方法基于判别模型。

3.3 实验结果及分析

本节首先介绍实验中使用的数据集，然后进行实验结果的展示，并与其他方法进行对比分析。

3.3.1 数据集介绍

为了实验的准确性和可比性，我们使用了行人再识别领域被广泛使用的公开数据：VIPeR[47]，CUHK01[48]和 PRID[49]，如图 3-7 所示。

VIPeR 数据集。由 632 个人在两个摄像头下的 1264 张图片组成，每个人在每个摄像头下只有一张图片，图片被归一化到 128*48 像素值。该数据集除了两

个摄像头的视角不同外，光照条件变化非常大，这给再识别带来了非常大的难度。试验中，我们使用 316 对样本作为训练集，剩下的作为测试集。

CUHK01 数据集。总共包含了 971 个人，每个人在每个摄像头下有两张图像。摄像头 a 摄取的是人的前后视角，摄像头 b 摄取的是人的侧视角。所有的图像都被归一化到了 160*60 像素值。该数据集场景是在室内，所以光照变化不大，相对比较简单。我们使用 485 对样本作为训练集，剩下的作为测试集。

PRID 数据集。该数据集摄像头 a 有 385 个样本，摄像头 b 有 749 个样本，并有 200 个同时出现在 a 和 b 中。PRID 数据集跨摄像头光照和视角变化明显，但背景相对单一。我们也像大多数学者一样，随机取 200 对中的 100 对样本作为训练集，剩下的 100 对中摄像头 a 中的 100 个作为 probe，所有摄像头 b 中的 649 个（100+549）作为 gallery。

实验中均取 10 次结果的平均值作为最终的结果。

图 3-6 VIPeR、CUHK01、PRID 数据集示意图

3.3.2 性能评测准则

为了进行算法性能的评测以及与其他方法的对比，我们选取在行人再识别中被广泛使用的评测方法：累计匹配特性曲线（cumulative matching characteristic，CMC）。行人再识别问题本质上是一个检索排序问题。在 CMC 曲线中，如图 3-8 所示，横坐标表示排名，纵坐标表示准确率。对于曲线上的每一个点(x, y)表示

在排名前x个样本中，包含正确匹配的查询样本占所有查询样本的比率，CMC 是一条非降的曲线。

图 3-7 CMC 性能曲线

也有一些学者使用其他的评测指标比如平均准确率（mean average precision，mAP），如 Liang Zheng[50]等在他们的实验中使用了这种评测方法。它是在准确率-召回率（precision-recall，PR）曲线的基础上计算出来的。准确率和召回率计算方式如下：

$$\mathrm{Pr}\,ecision = \frac{TP}{FP+TP} \tag{3-10}$$

$$\mathrm{Re}\,call = \frac{TP}{FN+TP} \tag{3-11}$$

AP 是在 PR 曲线的基础上计算曲线下面的面积，mAP 对所有的 AP 结果取平均。mAP 在每个行人具有多张图片的情况下能够比较好的反映同一个人的不同图片在排序中的位置情况。在我们的实验中，图片库中每个行人只有一张图片，所以我们使用更被广泛使用的 CMC 曲线作为最终的性能衡量标准。

3.3.3 结果与分析

在实验中，我们使用颜色直方图，HOG 和 LBP 特征级联形成的 WHOS 特征原始维度为 5138 维，为了降低计算复杂度，在使用之前，我们使用 PCA 降维算法将其降到 400 维。对于每一个查询样本x_p和比对样本x_g，我们分别计算它们

各自为正样本时的局部度量即 probe metric 和 gallery metric，然后分别使用 probe metric、gallery metric 计算x_p和x_g之间的相似度，最后将这两个相似度相加。在 VIPeR 和 CUHK01 中我们每次实验都是随机选取一半的样本对作为训练集，剩下的一半作为测试集。由于 PRID 数据集的特殊性，我们训练时随机取 100 对样本。在测试时 probe 为 100 个样本，即 100 个查询样本，但是 gallery 集我们分别取 100（对应 probe 集的 100 个样本）和 649（100+549）个样本进行测试。

图 3-8 分别展示了 probe metric、gallery metric 和 combined metric 在 VIPeR、CUHK01 和 PRID 数据集上的 CMC 性能曲线，表 3-1 展示了三种 metric 在三个数据集上 rank-1 的匹配性能。图 3-9 展示了部分样本在局部度量学习算法下的排序示例，从左到右与查询样本的距离依次增大。表 3-1 和图 3-8 说明通常情况下融合的距离度量性能要好于单个的 probe metric 或者 gallery metric 的性能，这符合模型融合的通常规律。然而，在 PRID（gal:649）中，由于 gallery 中混有大量不相关的样本，使得使用基于样本的局部度量学习算法时，gallery metric 的性能远远大于 probe metric 的性能，这导致两个度量融合后的性能只是在较低的 probe metric 的性能的基础上有一个小的提升。通过比较 PRID（gal:100）和 PRID（gal:649）的结果说明一个问题，在查询过程中，gallery 中不相关（反例）样本会对正确匹配起到干扰作用。

为了缓解这个问题，我们提出了基于 KNN 交集的重排序算法，通过充分利用 gallery 中大量不相关样本的信息来减少其带来的负面影响。有关重排序算法将在下一章中详细阐述。

表 3-1 局部度量学习算法在三个数据集上的实验结果

Rank-1	Probe metric	Gallery metric	Combined metric
VIPeR	27.53	25.25	**29.91**
CUHK01	27.44	29.05	**32.82**
PRID（gal:100）	35.90	34.80	**40.90**
PRID（gal:649）	17.10	**25.30**	19.30

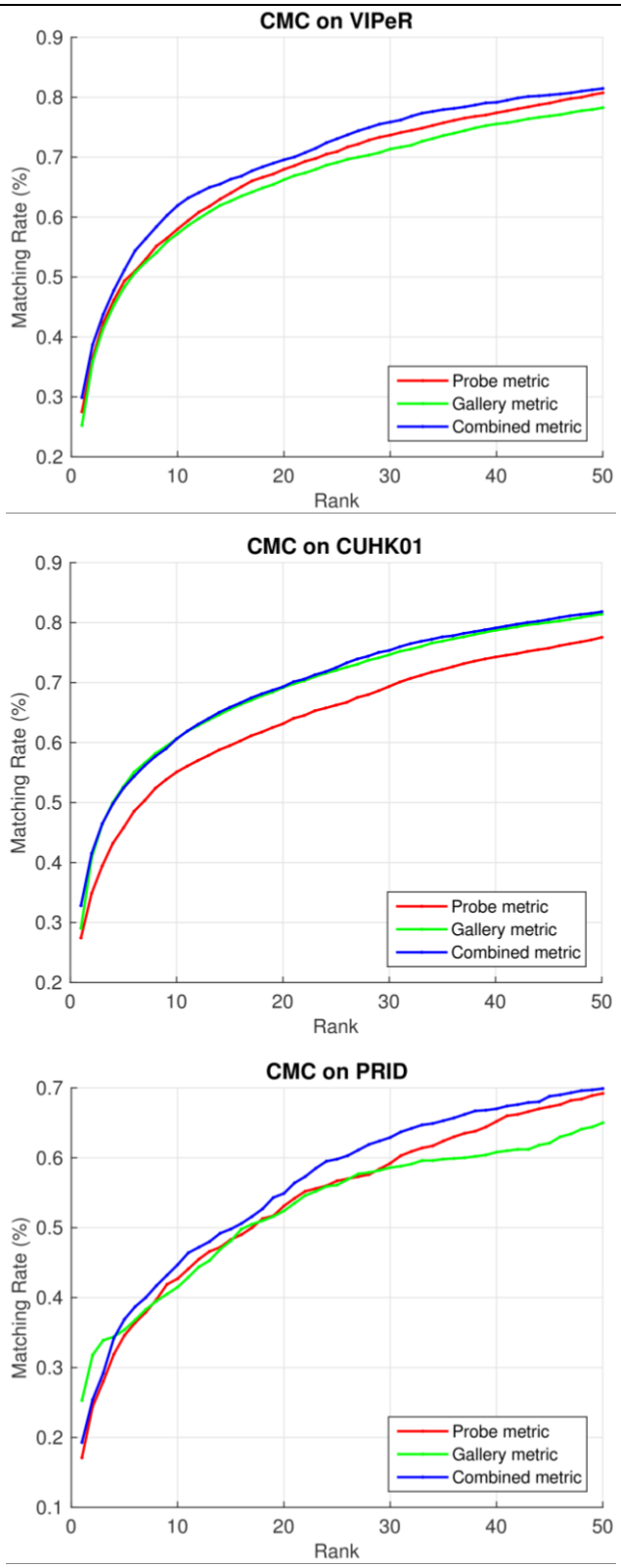


图 3-8 局部度量学习算法在三个数据集上的实验结果

图 3-9 局部度量学习算法的排序示例

3.4 本章小结

本章主要介绍了本研究的算法框架，包括行人检测、特征表示、度量学习、匹配排序等模块。然后详细描述了实验结果和分析。实验结果表明，基于无监督局部度量学习的行人再识别算法在三个数据集上都有不错的性能，融合 probe metric 和 gallery matric 后最终性能会得到提升。

然而，本算法也有不足之处，特别是在 gallery 集中有许多不相关的样本时，会显著影响最终的匹配性能。因此，针对这一问题，我们提出了一个基于 KNN 交集的重排序算法。

第四章 基于重排序的行人再识别算法

本章针对上一章中发现的问题，提出了基于 KNN 交集重排序（KIRR）的行人再识别算法。本章首先阐述引入重排序算法的原因，然后详细介绍算法细节，最后分析实验结果，总结重排序算法起到的作用。

4.1 问题描述及分析

在实际监控场景中，图片库的规模一般是巨大的，因为待查询样本在目标摄像头下是否出现以及出现的具体时间段实际上我们是无法得知的。一般情况下我们会估计一个时间段然后收集该时间段内所有出现的行人组成图片库，然后使用算法进行查询。查询结果没有出现对应的样本，那么有可能需要继续扩大查询时间段，图片库的规模将进一步扩大。因此，相对于感兴趣的那个样本图像而言，图片库中所有其他图片都是反例图片，这些图片中和正例图片相似度比较大的样本（难反例样本）将起到干扰匹配的反作用，通过实验表明，针对同一个查询样本，图片库中反例样本越多，正确匹配率就越低，如图 4-1 所示。

图 4-1 gallery 大小和匹配准确率的关系

相似样本干扰匹配的问题限制了数据规模的可扩展性，极大地影响了行人

再识别算法在实际场景的应用。针对这一问题，我们尝试合理利用图片库中的样本之间的相似度关系，然后通过在重排序阶段引入这些信息来降低难反例的影响，优化排序结果。

4.2 基于 KNN 交集的重排序算法

本小节提出了一种基于k近邻交集的重排序方法。该方法主要基于查询样本的初始排序进行重排序。初始排序基于查询样本和每一个 gallery 样本之间的距离，距离越小排名越靠前。该方法的核心思想是：查询样本在 gallery 集合中获取的初始排序可以看做是查询样本基于样本距离的特征描述，即如果某个查询样本在 gallery 样本集合中的k个最近邻（即前k个排序）中有一部分和某个 gallery 样本的k个 gallery 样本最近邻一样，则这两个样本一定在某种程度上比较相似。基于这个原则，我们可以通过两个样本k近邻中相同样本的数量以及后者在前者初始排序中的位置来重新衡量这两个样本之间的相似度。

我们用q表示一个查询样本，$G = \{g_i\}_{i=1}^{m}$表示 gallery set，其中m表示集合中样本的数目。通过计算距离$d(q, g_i)$，可以得到初始的 query-gallery 排序$R_q(G) = \{g_i^0\}_{i=1}^{m}$。其中，$\mathrm{d}(q, g_1^0) < \mathrm{d}(q, g_2^0) <, \ldots, < \mathrm{d}(q, g_m^0)$。

首先，定义一个排序得分$S_r(q, g_i^0) = 1/i$，表示查询样本和 gallery 样本的初始相似度。然后我们计算q和g_i^0之间k近邻中相同样本的数量。具体地，我们定义$n_k(q)$为q的k近邻，$n_k(g_i^0)$为g_i^0的k近邻。所以我们可以定义k近邻得分为：

$$S_{cn}(q, g^0{}_i) = | n_k(q) \cap n_k(g^0{}_i) | \tag{4-1}$$

最后，q和g_i^0之间新的相似度可以定义为：

$$S_n(q, g^0{}_i) = S_{cn}(q, g^0{}_i) \cdot S_r(q, g^0{}_i) = \frac{| n_k(q) \cap n_k(g^0{}_i) |}{i} \tag{4-2}$$

显而易见，$S_{cn}(q, g_i^0)$和$S_r(q, g_i^0)$越大，则S_n越大，两者之间的相似度就越大。使用公式（4-2）我们可以得到查询样本和 gallery 样本之间新的更加精确的相似度。尤其在 gallery 集包含比较多的不相关样本的情况下，因为S_{cn}充分利用了

gallery 集中所有样本之间的近邻关系和与查询样本之间的近邻关系来表示样本之间的相似度。算法细节如图 4-2 所示。

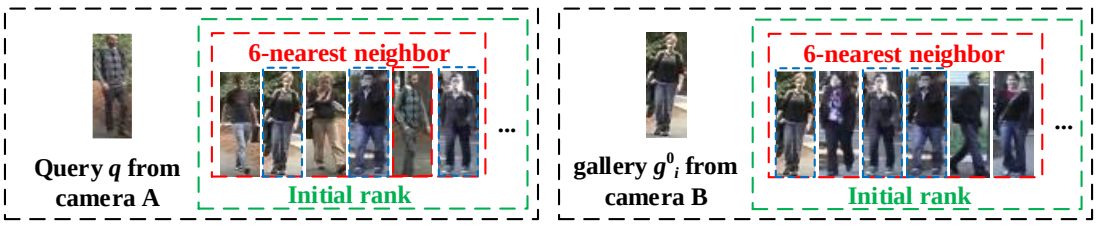


图 4-2 KNN 交集重排序算法

4.3 实验结果及分析

为了验证基于 KNN 交集的重排序算法的效果，我们在三个数据集上都做了重排序的实验，图 4-3 展示了基于样本的度量学习算法（SBML）和结合 KNN 交集重排序（KIRR）的 SBML 算法在三个数据集上的性能。为了充分展现 KNN 交集重排序算法的作用，我们增大 gallery 集样本数相对于查询样本数的大小，即增加了 gallery 集中反例样本的数量，具体的实验设置如下：VIPeR（probe: 158, gallery: 316）；CUHK01（probe:243, gallery: 486）；PRID（probe: 100, gallery: 649）。实验表明：合理利用 gallery 中的样本能有效提高性能，特别是在 PRID（probe: 100, gallery: 649）数据集上 gallery 比 probe 大得多的情况下（从 25.30%提到 38.50%），几乎接近 40.90%（probe: 100, gallery: 100）。说明 KNN 交集重排序算法能通过有效利用 gallery 中大量不相关样本的信息来减少其在匹配过程中带来的负面影响。图 4-4 展示了部分样本在 KNN 交集重排序算法下的排序示例。

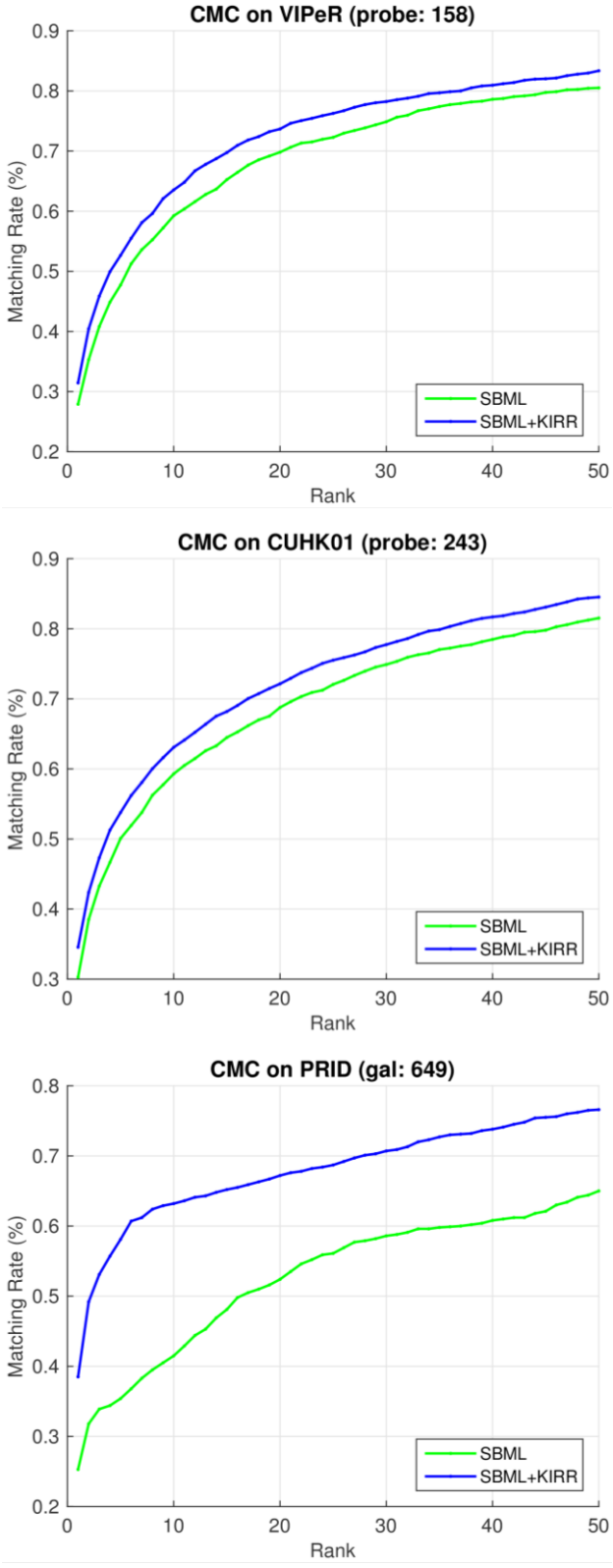


图 4-3 基于 KNN 交集重排序算法在三个数据集上的实验结果

图 4-4 基于 KNN 交集重排序算法的排序示例

表 4-1 与 state-of-the-arts 的结果比较

Rank-1	VIPeR	CUHK01	PRID
ISR[33]	27.0	-	17.0
GTS[51]	25.2	-	-
DLILR[35]	29.6	28.4	21.1
UCDTL[52]	**31.5**	27.1	24.2
Ours-SBML	29.91	32.82	25.30
Ours-SBML+KIRR	29.94	**32.82**	**38.50**

在表 4-1 中，我们选择了一些经典的无监督度量学习算法进行实验结果比较，包括：基于图模型的 GTS[51]，基于稀疏表示分类的 ISR[33]，迭代的拉普拉斯正则字典学习 DLILR[35]和无监督跨数据集迁移学习 UCDTL[52]。这些方法中大

部分原理是学习一个新的特征空间，使得在这个特征空间中的距离具有比较好的判别性。然而，它们在学习的过程中还是需要以一定概率假设样本之间的类别关系，这就引入了一定比例的错误类别标号，进而影响最终的匹配性能。与目前性能比较好的 UCDTL[52]相比，我们的算法在 VIPeR 上取得了可比的性能，在 CUHK01 和 PRID 上性能有大幅度提高。值得注意的是，我们的算法只使用了目标数据集的训练数据而没有使用其它的数据集进行训练。

4.4 本章小结

本章提出的基于 KNN 交集重排序的行人再识别算法是为了降低图片库中反例样本对正确匹配造成的干扰，使得在图片库规模不断扩大时查询的正确率不会有比较明显的下降。实验表明，我们的方法基本上达到了预期的目标。

基于 KNN 交集重排序的行人再识别算法是基于初始排序进行的一种重排序方法，不需要监督信息，即不需要标注样本，这使得本方法可以和其他的包括基于监督学习和基于无监督学习方法结合使用，在原来的基础上进一步提高性能。

虽然我们的方法在一定程度上解决了匹配问题，但在图片库中反例样本比较少的时候重排序的效果不是很明显，目前的方法过于直观是造成这个问题的原因之一，之后我们将不断改进，使之能适应更多不同的场景。

第五章 结论与展望

跨摄像头的行人再识别是目前智能视频监控领域的热点研究问题，它的主要目的是获取特定目标在一个特定摄像头网络覆盖区域特定时间内的行动轨迹。这在目前大数据时代、视频监控自动化、平安城市建设等背景下具有重要意义，在刑事侦查、走失儿童查找、智能交通管理等领域具有很大的实用价值。如何利用各种技术手段如计算机视觉、机器学习、深度学习等有效地解决行人在跨摄像头过程中产生的各种变化，从而使得最终的查询结果符合人们的期望是现在学者们的主要研究方向。目前，行人再识别面临的主要挑战有：背景复杂多变、光照变化、视角变化、姿态变化、行人遮挡、相似行人干扰、摄像头参数变化等影响因素造成的干扰问题。这些因素极大地影响了行人再识别技术在实际场景中的使用。针对这些问题，研究者们提出了许多算法和解决方案。这在一定程度上使得上述问题有所缓解。

5.1 总结

本文首先介绍了行人再识别领域近几年来国内外最新研究进展，总结了目前行人再识别技术的常用算法框架，并详细介绍了系统框架中的每一个模块。分析了现有方法中存在的缺点，并在现有技术的基础上，针对其中的不足之处，给出了改进办法，并提出了自己的解决方案。

1. 算法框架。目前比较常用的行人再识别算法框架一般由：视频关键帧行人检测、背景去除、特征表示、度量学习构成，大量工作集中在特征表示和相似性度量学习上。

2. 现有方法总结。由于摄像头采集到的是视频数据，无法直接使用。因此，首先需要通过行人检测的方法获取关键帧中的感兴趣目标，目前常用的检测方法有 DPM 和基于深度网络的检测方法。在获得了人体图片之后，需要进行背景去除，但这一步不是必须的，因为通过行人检测之后的人体图片中背景部分所占比重已经很小，可以直接进行特征提取。在特征提取阶段，大量的特征表示方法

被提出，其中，跨摄像头的光照、视角和姿态变化是主要考虑的因素，针对光照变化，可以对图片直方图做均衡化、MSR（Multi Scale Retinex）算法等，针对视角和姿态变化，一般我们会把人体图片进行水平划分后基于划分条带为单位进行滑窗，然后提取特征。当然，基于深度学习的特征表示也取得了很好的性能，在图片预处理和网络结构设计过程中也需要考虑光照、视角和姿态变化的因素。最后是度量学习阶段，度量学习本质上是学习一个特征的映射空间，使得对于不同的场景，可以通过数据学习适合特定场景的映射子空间。现有的度量学习方法存在的问题有：（1）迁移能力。在一个场景中学习的度量无法很好地在另一个场景中使用。（2）数据标定。现有的大部分度量学习方法需要标定数据，这限制了算法的实际应用，由于迁移能力弱，对每一个新场景都需要重新标定训练数据，进一步降低了算法的实用性。（3）全局度量。由于训练数据有限，实际场景中不同行人之间差异巨大，通过有限的数据集训练得到的全局度量无法对场景中出现的所有行人都保持较好的判别性。

3. 本研究总结。针对上述目前行人再识别方法中存在的问题，本文尝试使用无监督学习的方法，提出了基于局部度量学习的行人再识别算法。针对数据标定问题，本算法基于无监督学习，只需要相关场景中的无标定数据进行训练，可行性和实用性大大提高。针对全局度量的缺陷，本算法为每一个查询样本训练局部度量，相当于为单个样本进行了调优，保证距离度量的精准性。现有的无监督行人再识别方法大多使用字典学习和稀疏重构方法学习新的特征表示，然而这种在重构误差和一范数正则最小的基础上加入一些模糊的标签信息的做法不具有很好的判别性，而且在大数据量情况下模型的计算复杂度也是一个问题。本方法使用基于支持向量的判别模型，只使用和查询样本不是同一个行人的负样本以及查询样本训练模型，不加入其他假设。具有较好的判别性和较低的时间复杂度。

针对实际场景中图片库中相似样本的干扰，本研究提出了基于样本k近邻交集的重排序算法，通过引入样本之间的k近邻关系来减少在查询匹配时图片库中

和查询样本相似的负样本（难样本）的干扰。使得本方法具有较强的实用性。

5.2 展望

随着机器学习、计算机视觉技术的快速发展，行人再识别领域受到了人们的广泛关注，各种新的方法被提出。然而，目前的性能还是难以达到人们的预期目标，离实际运用还有很大的差距。在行人再识别的研究过程中，其它相关领域的优秀理论成果可以被借鉴比如：人脸识别、信息检索、图像分类等，在借鉴的同时充分考虑行人再识别问题的特殊性，就有可能做出好的研究工作。作者认为，目前在如下几个方面值得进行深入研究：

1. 基于深度学习的行人再识别。深度学习在诸多领域比如：计算机视觉、自然语言处理、语音识别等发挥了不可比拟的作用。在行人再识别中，基于手工设计的特征具有非常大的局限性，具有多个隐层的深度网络可以自动学习图像的底层和高层特征表示。而且深度网络可以同时处理检测、特征表示、分类等多个任务，这给端到端的行人再识别提供了合适的解决方案。

2. 度量学习研究。度量学习是一个基础性问题。虽然目前在这个领域已经有许多不错的研究成果，但是针对特定问题比如再识别问题的改进优化还需要继续探究。在行人再识别问题中，如何获得泛化性好的度量是一个重要问题，即一次训练到处使用。在一个场景中训练得到的度量如果可以在其他场景使用，或者做简单的迁移学习可以使之适应不同的场景，这具有非常高的实用价值。

3. 无监督的行人再识别。无监督是未来机器学习重点研究方向[53]。目前在行人再识别中还没有在性能上可以和监督学习媲美的无监督学习方法。因此，在这个方面还有很长的路要走。

所以，在行人再识别领域，虽然目前已经取得了一些成果，但面对现实环境中复杂多变的因素，导致现在的方法还不能在真实场景中使用。但随着技术的不断发展，这些问题将被慢慢解决，最终行人再识别技术一定可以像人脸识别一样被人们有效利用，解决实际问题。

参考文献

[1] Bedagkar-Gala A, Shah S K. A survey of approaches and trends in person re-identification[J]. Image and Vision Computing, 2014, 32(4): 270-286.

[2] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//European Conference on Computer Vision. Springer Berlin Heidelberg, 2008: 262-275.

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

[4] Ma B, Su Y, Jurie F. Local descriptors encoded by fisher vectors for person re-identification[C]//European Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 413-422.

[5] Zhao R, Ouyang W, Wang X. Learning mid-level filters for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 144-151.

[6] Liao S, Hu Y, Zhu X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2197-2206.

[7] Matsukawa T, Okabe T, Suzuki E, et al. Hierarchical gaussian descriptor for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1363-1372.

[8] Köstinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2288-2295.

[9] Zheng W S, Gong S, Xiang T. Reidentification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 653-668.

[10] Xiong F, Gou M, Camps O, et al. Person re-identification using kernel-based metric learning methods[C]//European Conference on Computer Vision. Springer International Publishing, 2014: 1-16.

[11] He D C, Wang L. Texture unit, texture spectrum, and texture analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 1990, 28(4): 509-512.

[12] Fogel I, Sagi D. Gabor filters as texture discriminator[J]. Biological Cybernetics, 1989, 61(2): 103-113.

[13] Schwartz W R, Davis L S. Learning discriminative appearance-based models using partial least squares[C]//Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on. IEEE, 2009: 322-329.

[14] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2005, 1: 886-893.

[15] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[16] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[17] Layne R, Hospedales T M, Gong S. Person Re-identification by Attributes[C]//BMVC. 2012, 2: 3.

[18] Layne R, Hospedales T M, Gong S. Towards person identification and re-identification with attributes[C]//European Conference on Computer Vision. Workshops and Demonstrations. Springer Berlin Heidelberg, 2012: 402-412.

[19] Layne R, Hospedales T M, Gong S. Re-id: Hunting attributes in the wild[C]//BMVC, 2014.

[20] Chen H, Gallagher A, Girod B. Describing clothing by semantic attributes [M]//European Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 609-623.

[21] Li A, Liu L, Wang K, et al. Clothing attributes assisted person reidentification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(5): 869-878.

[22] Zhao R, Ouyang W, Wang X. Learning mid-level filters for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 144-151.

[23] Zhao R, Ouyang W, Wang X. Unsupervised salience learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3586-3593.

[24] 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, 2015, 37(8): 1629-1642.

[25] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 152-159.

[26] Xing E P, Jordan M I, Russell S, et al. Distance metric learning with application to clustering with side-information[C]//Advances in Neural Information Processing Systems. 2002: 505-512.

[27] Weinberger K Q, Blitzer J, Saul L K. Distance metric learning for large margin nearest neighbor classification[C]//Advances in Neural Information Processing Systems. 2005: 1473-1480.

[28] Zheng W S, Gong S, Xiang T. Person re-identification by probabilistic relative distance comparison[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011: 649-656.

[29] Pedagadi S, Orwell J, Velastin S, et al. Local fisher discriminant analysis for pedestrian re-

identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3318-3325.

[30] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2288-2295.

[31] Paisitkriangkrai S, Shen C, van den Hengel A. Learning to rank in person re-identification with metric ensembles[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1846-1855.

[32] Zhang L, Xiang T, Gong S. Learning a discriminative null space for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1239-1248.

[33] 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, 2015, 37(8): 1629-1642.

[34] Wang H, Gong S, Xiang T. Unsupervised learning of generative topic saliency for person re-identification[J]. British Machine Vision Association Bmva, 2014.

[35] Kodirov E, Xiang T, Gong S. Dictionary Learning with Iterative Laplacian Regularisation for Unsupervised Person Re-identification[C]//BMVC. 2015, 3: 8.

[36] 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, 1997, 6(7): 965-976.

[37] Liao S, Zhao G, Kellokumpu V, et al. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 1301-1306.

[38] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[39] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. 2012: 1097-1105.

[40] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[41] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[42] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[43] Guo Y F, Wu L, Lu H, et al. Null foley–sammon transform[J]. Pattern Recognition, 2006, 39(11): 2248-2251.

[44] Garcia J, Martinel N, Micheloni C, et al. Person re-identification ranking optimisation by discriminant context information analysis[C]//Proceedings of the IEEE International

Conference on Computer Vision. 2015: 1305-1313.

[45] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[46] Yang Y, Yang J, Yan J, et al. Salient color names for person re-identification[C]//European Conference on Computer Vision. Springer International Publishing, 2014: 536-551.

[47] Gray D, Brennan S, Tao H. Evaluating appearance models for recognition, reacquisition, and tracking[C]//Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS). 2007, 3(5).

[48] Li W, Zhao R, Wang X. Human reidentification with transferred metric learning[C]//Asian Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 31-44.

[49] Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification[C]//Scandinavian Conference on Image Analysis. Springer Berlin Heidelberg, 2011: 91-102.

[50] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1116-1124.

[51] Wang H, Gong S, Xiang T. Unsupervised learning of generative topic saliency for person re-identification[J]. British Machine Vision Association Bmva, 2014.

[52] Peng P, Xiang T, Wang Y, et al. Unsupervised cross-dataset transfer learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1306-1315.

[53] http://yann.lecun.com/

个人简介、在学期间发表的论文与研究成果

个人简介

姓名：赵恒　性别：男　出生日期：1991.10.19　政治面貌：中共党员

2010 年 9 月至 2014 年 7 月　浙江大学　自动化　学士

2014 年 9 月至 2017 年 7 月　中国科学院大学　计算机技术　硕士

已发表（录用）论文

1. **Heng Zhao**, Zhenjun Han, Zhaoju Li, Fei Qin, Unsupervised Person Re-identification via Re-ranking Enhanced Sample-specific Metric Learning[C]//IEEE International Conference on Image Processing. IEEE, 2017.

致　谢

在中国科学院大学攻读硕士学位的三年学习生活时光即将结束。回首这段日子，我经历了许多挫折，付出了许多努力，同时也收获了许多成长。在毕业论文即将完成之际，在此由衷地感谢三年来帮助过我的老师、同学、朋友和家人。

本研究课题及学位论文受到了韩振军副教授和叶齐祥教授的悉心指导。首先，我要感谢导师韩振军副教授在我攻读硕士学位期间学习和生活上的关心和指导，感谢他在我论文选题、框架设计、论文撰写和修改中倾注的大量时间和心血。感谢叶齐祥教授在我科研的起步阶段给予的启蒙指导以及三年科研过程中的讨论和帮助。感谢秦飞副教授给我的科研提供的指导建议和对我论文的修改帮助。感谢焦建彬教授为我提供了优越的科研生活环境，让我可以排除其他困难，专注学习和科研。恩师们对学术的满腔热情，对治学的严谨态度，对学生的诲人不倦深深地感染了我。

感谢李策师姐、高山师兄、魏朋旭师姐、柯伟师兄、李兆举师兄等实验室其他师兄师姐在我学习和科研过程中提供的耐心指导。感谢同届的黄显淞、刘嫣然和施梦楠等同学三年来的互相照应，他们耐心的帮助和无私的关心是我强大的精神依靠，给我的硕士生涯增添了不少色彩，让我觉得这三年生活的无比温暖和快乐。

感谢我的家人特别是我的父母，感谢他们二十多年来的养育和陪伴，在我遇到困难时他们永远是我最安全的避风港。希望他们永远快乐安康。

感谢参加开题及中期评阅的各位老师和专家们，他们利用自己广博的学识和丰富的经验无私地帮助我把握论文方向和研究进度。感谢参加论文评审和答辩的各位老师。

赵恒

2017 年 5 月