中山大學
SUN YAT-SEN UNIVERSITY

本科生毕业论文（设计）

Undergraduate Graduation Thesis（Design）

题目 Title：基于卷积神经网络的行人再标识研究

院 系
School (Department)：电子与信息工程学院

专 业
Major：智能科学与技术专业

学生姓名
Student Name：吴尚轩

学 号
Student No.：12350063

指导教师(职称)
Supervisor（Title）：郑伟诗 教授

时间： 2016 年 4 月 24 日

Date: Month Day Year

附表一、毕业论文开题报告

论文（设计）题目：

（简述选题的目的、思路、方法、相关支持条件及进度安排等）

学生签名：　　年　　月　　日

指导教师意见：

1、同意开题（ ） 2、修改后开题（ ） 3、重新开题（ ）

指导教师签名：　　年　　月　　日

附表二、毕业论文过程检查情况记录表

指导教师分阶段检查论文的进展情况（要求过程检查记录不少于 3 次）：

第 1 次检查

学生总结：

指导教师意见：

第 2 次检查

学生总结：

指导教师意见：

第 3 次检查

学生总结：

指导教师意见：

第 4 次检查

学生总结：

指导教师意见：

学生签名：　　2016 年 4 月 25 日

指导教师签名：　　2016 年 4 月 25 日

总体完成情况	指导教师意见： 一切按原计划进行，研究进度很好。 1、按计划完成，完成情况优（ ） 2、按计划完成，完成情况良（ ）

	3、基本按计划完成，完成情况合格（ ） 4、完成情况不合格（ ） 指导教师签名：　　2016 年 4 月 25 日

附表三、毕业论文答辩情况

答辩人		专 业	
论文（设计）题目			
答辩小组成员			
答辩记录：			

记录人签名：　　年　月　日

学术诚信声明

本人所呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。本毕业论文的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

本人签名：　　日期：　　2016 年 4 月 25 日

论文题目：基于卷积神经网络的行人再标识研究
专　　业：智能科学与技术专业
学生姓名：吴尚轩
学　　号：12350063
指导教师：郑伟诗 教授

摘 要

行人再标识问题是一种在不同摄像头场景下发现同一行人并进行匹配的问题，目前在视频监控与智能安防领域的应用越来越广泛。然而，传统特征提取与度量学习的方法已经到达识别率的瓶颈。近两年来，随着卷积神经网络模型的普及，基于深度学习的方法正在在计算机视觉与机器学习领域受到越来越多的重视。

本文首先提出了一种基于卷积神经网络的行人再标识图像特征提取方法 Feature Fusion Net（FFN）。利用其反向传播、自我学习的特性，结合手工剪裁特征，我们得到表达能力更强的行人图像特征。在 VIPeR、CUHK01、PRID450s 三个通用行人图像数据集上的实验结果表明，FFN 特征能够有效改善不同分类器下行人图像匹配的效果。

其次，本文结合了传统的度量学习方法，提出了一个高效的行人图像匹配模型。这个模型将 FFN 图像特征、KMFA 度量学习方法和针对行人图像优化的 Mirror 特征转换表示结合在一起。实验证明，在上述三个行人再标识数据集中，我们的模型匹配准确率分别超出当前最好结果 8.09%，7.98%和 11.20%。

第三，本文研究了一个行人再标识领域的新问题：基于不完整行人图像的匹配问题。针对此问题，本文提出了一种基于卷积神经网络的模型 PartialNet。在 Partial REID 不完整行人图像数据集中，PartialNet 模型表现出比较优秀的匹配效果。

本文的前两项贡献已整理为会议论文，发表于 IEEE WACV2016，本人为第一作者。我们正在继续研究基于不完整图像的行人再标识问题，并希望于毕业前得到令人满意的研究结果，投稿至 IEEE ACCV2016 会议。

关键词：行人再标识；卷积神经网络；图像特征；度量学习；监控摄像头

Title: Person Re-identification Based on Convolutional Neural Network
Major: Intelligence Science and Technology
Name: Shangxuan Wu
Student ID: 12350063
Supervisor: Prof. Wei-Shi Zheng

Abstract

Person re-identification has been studied extensively in the past ten years. This problem aims at more precise matching of person images under different views of surveillance camera. Feature Representation and metric learning are two critical components in person re-identification models.

In this paper, we first focus on the feature representation of images and claim that hand-crafted histogram features can be complementary to Convolutional Neural Network (CNN) features. We propose a novel feature extraction model called Feature Fusion Net (FFN) for pedestrian image representation. Utilizing color histogram features (RGB, HSV, YCbCr, Lab and YIQ) and texture features (multi-scale and multi-orientation Gabor features), we get a new deep feature representation that is more discriminant and compact.

Secondly, we utilize different metric learning methods in order to achieve better matching accuracy on person re-identification problem. In our second model, we explored Kernel Marginal Fisher Analysis (KMFA) classifier and combined it with the Mirror trick, which is specifically designed for person re-identification tasks. Experiments on three challenging datasets (VIPeR, CUHK01, PRID450s) validates the effectiveness of our proposal.

Thirdly, we focus on the complex environment of actual-scene person re-identification and propose a new deep learning method (PartialNet) that could achieve better performance on partial person re-identification images. Our model is tested on a publicly available dataset (Partial REID) and achieve outstanding performance.

Keywords: Person Re-identification, Convolutional Neural Network, Feature Representation, Metric Learning, Surveillance Camera

目 录

第 1 章 引言

1.1 行人再标识的研究背景与意义

随着现代社会的快速发展和城市化进程的加快，城市人口密度不断增加。随之而来的是飙升社会治安问题。人口密集的公共场所，往往会出现意想不到的危险。因此，对行人的研究受到了学者的广泛关注。行人再标识模型作为行人研究的重要组成部分，逐渐成为视频监控领域的研究热点。在人工智能技术迅速发展的今天，利用计算机视觉，自动进行行人的识别跟踪，提高搜索效率和准确度，节省资源，对安防工作有很大的实用价值和研究意义。

行人再标识，就是在大规模分布式监控系统中，在不同摄像头之间，进行不同地点不同时间（一般是短期内）的目标行人匹配。图 1-1 是一个行人再标识系统的应用示例。首先在摄像头 A 拍摄到的场景中进行行人检测，然后确定要跟踪或搜索的目标行人，用行人再标识技术，将不同场景（如图中的摄像头 B）中的目标行人匹配，从而达到跟踪或搜索的目的。

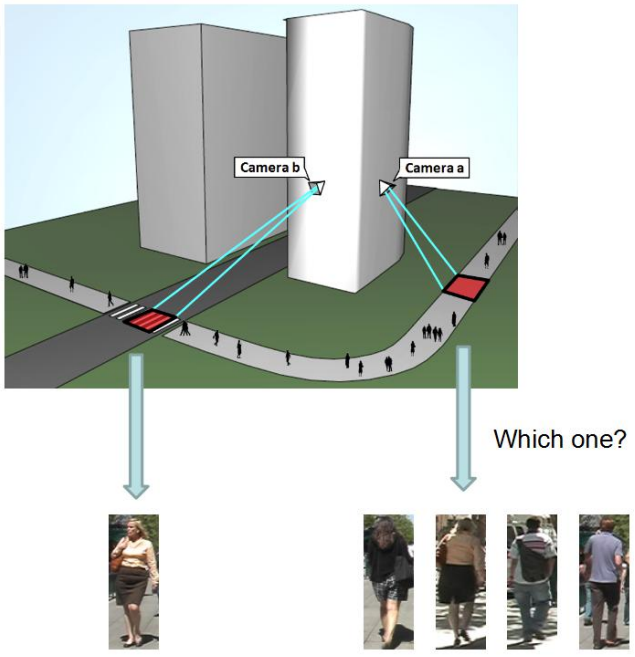

图 1-1 行人再标识系统示例

计算机科学、电子技术、图像处理、模式识别与人工智能的发展对行人识别模型的进步产生了推动作用，经过近十年的发展，行人识别领域积累了丰富的理论和大量的算法。现今的行人识别模型已经可以基本实现行人再标识，但是受行人所处环境的非可控性影响，识别率并不是十分理想。

在非可控环境下，同一行人在不同场景下存在的差异较大，给行人的准确识别带来了一定的困难。非可控因素一般存在于行人和背景两个方面。在行人方面，行人的动作和表情会随时发生变化；行人离摄像机距离远近不同，行人大小也会不同；不同的拍摄角度，会得到行人的不同姿势。在背景方面，由于行人不停移动导致背景是时变的，背景中遮挡物和光照的变化也较为明显。

图 1-2 直观地展示了非可控环境下的行人再标识问题的复杂性。图 1-2 左边的一组图片是身份已知的行人图片样本，中间的 20 张图片是身份未知的测试样本。行人再标识的任务就是，给定一张测试样本图片，在图库中寻找与测试样本相似度最高的样本，以决定测试样本的行人身份。图 1-2 中每一行的红色方框表示与左列正确匹配的行人图片。

图 1-2 行人图像匹配示意图[33]

1.2 行人再标识的通用模型与研究难点

行人再标识方法一般由以下两个步骤组成：

1) 提取行人图片的特征。从图片中提取颜色、纹理等低层次特征，通过统计直方图、协方差等方法得到复合的特征向量。将图像分块的特征向量

拼接，并以此作为该行人图片的特征表达。

2) 通过匹配模型，比如欧氏距离、马氏距离或通过度量学习方法求出的样本对间的距离，求出测试样本与样本库中每一张图片的相似度。以此，得出与测试样本最相似的 i 个样本。

针对以上两个步骤，行人再标识的研究方向主要集中在两点：

1) 提出区分度更大、更紧凑的图像特征表达。

2) 研究更有效的分类器，在度量学习方法中，通过学习，将不同类的特征样本尽量分开。

聚焦于特征提取与度量学习方法的行人再标识模型虽然可以获得比较高的识别成功率，但是也不可避免地有以下不足：

1) 行人图像成对比较，输出每对图片的概率，匹配过程耗时长；

2) 在小的数据集进行训练，容易得到过拟合的模型；

3) 模型的匹配成功率强烈依赖于图像特征的表达能力；

4) 模型对于行人图像光照、遮挡、分辨率等变化的鲁棒性不足；

5) 度量学习方法数学理论解释较为复杂，优化方法限制较多。

近年来，卷积神经网络方法的提出，能够通过更大的训练数据量和更复杂的模型与参数，有效地解决以上问题。在本文中，我们聚焦于利用卷积神经网络方法来解决行人再标识问题。

1.3 行人再标识的相关工作

1.3.1 行人再标识图像特征

颜色（Color）和纹理（Texture）特征是行人再标识问题中应用最广泛的两类基本特征。例如，HSV 和 LAB 色彩空间直方图可以用来描述图像的色彩信息。LBP 和 Gabor 描述子可以用来描述图像的纹理信息。

近来，专为行人再标识设计的手工剪裁特征组合（Hand-crafted Features）显著地提升了匹配算法的准确率[7,9,20,27,32,33]。Local Descriptors encoded by Fisher Vectors[19]使用 Fisher Vector 来构建图像描述子。Color Invariants[14]使用颜色的分布作为图像匹配的唯一线索。Symmetric-Driven Accumulation of Local Features[7]不仅证明了图像的对称分割可以显著提高特征的匹配准确率，而且提

出了累积求和的图像特征提取思想。Local Maximal Occurrence Features[18]分析了图像的横向局部特征，并使用最大池化的方法获得更稳定的行人再标识图像特征表达。

1.3.2 行人再标识图像匹配方法

建立完整的特征表达后，我们需要基于距离判断两幅行人图像的匹配程度。目前，度量学习（Metric Learning）是被广泛应用于人脸识别（Face Recognition）[38,39,40,41]与行人再标识（Person Re-identification）[3,4,5,12,17,23,28,33,34]等计算机视觉问题的图像匹配模型：

Guillaumin et al.提出了 Logistic Discriminant Metric Learning（LDML）模型[41]，通过对每一对测试图像计算最大对数似然来计算距离。紧接着发表的 KISSME 模型[12]使用了一个 KISS 距离映射矩阵来着重解决测度（Scalability）的问题。在一些其他学术论文中，Pairwise Constrained Component Analysis（PCCA）[56]、Local Fisher Discriminant Analysis（LFDA）[23]、Information-theoretic Metric Learning（ITML）[57]等也是基于度量学习的经典行人再标识方法。

1.3.3 基于卷积神经网络的行人再标识模型

深度学习（Deep Learning）是目前机器学习与人工智能领域炙手可热的数学模型。其中，卷积神经网络（Convolutional Neural Network）已经被广泛地使用于计算机视觉的各类问题中。然而，其中只有少数研究聚焦于解决行人再标识问题：

Li et al.首先提出了 Deep Filter Pairing Neural Network（FPNN）[16]，使用 Patch Matching 层和 Max-out Pooling 层来识别行人再标识图片中的姿势和视角变化。FPNN 也是第一个在行人再标识问题中应用深度学习的模型。Ahmed et al.使用特别设计的 Cross-input Neighbourhood Difference 层[1]来提升准确率。随后，Deep Metric Learning[26]使用对称的（Siamese）深度神经网络和 Cosine Loss Function 损失函数应对不同视角下图片的巨大差别。Hu et al.提出了 Deep Transfer Metric Learning（DTML）[10]，将跨视角的先验知识传递到测试数据集中。

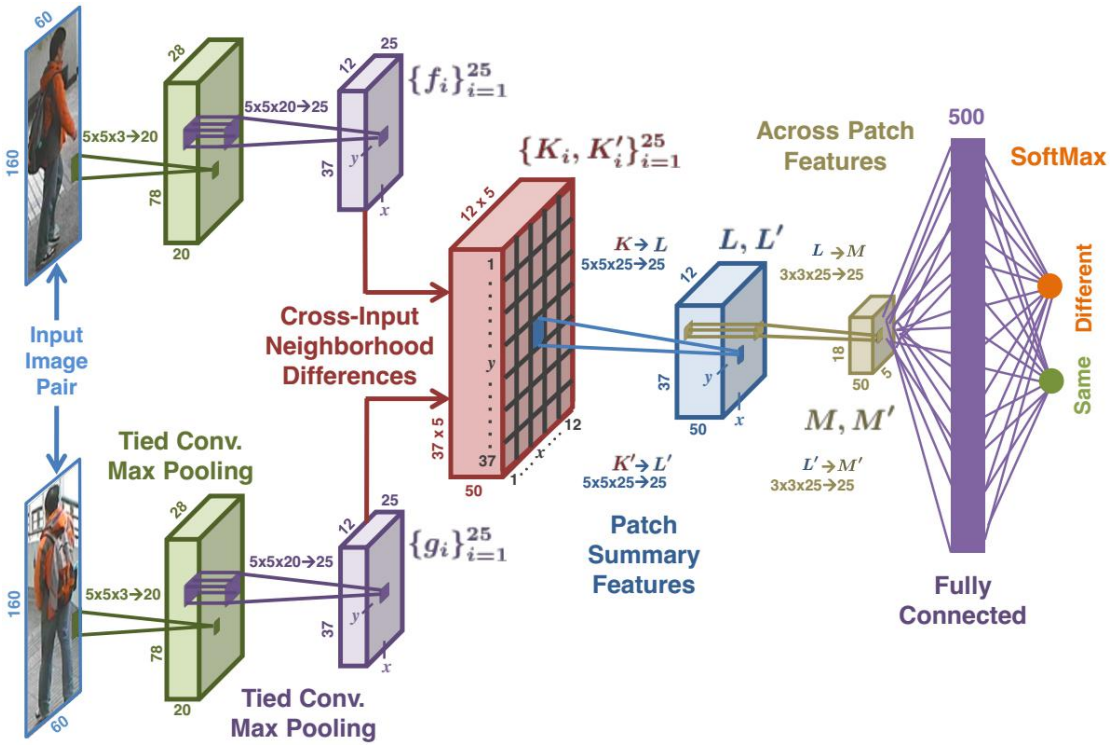

图 1-1 Ahmed et al.的神经网络结构示意图[1]

这些深度学习方法将特征提取和图像匹配两个问题放入同一个数学模型中予以解决。成对（Pairwise）图像的比较和对称的神经网络结构在这些深度学习方法中被广泛运用。成对学习（Pairwise Learning）导致训练样本数量剧增，也增加了神经网络训练的时间和难度。

1.4 论文研究工作与主要贡献

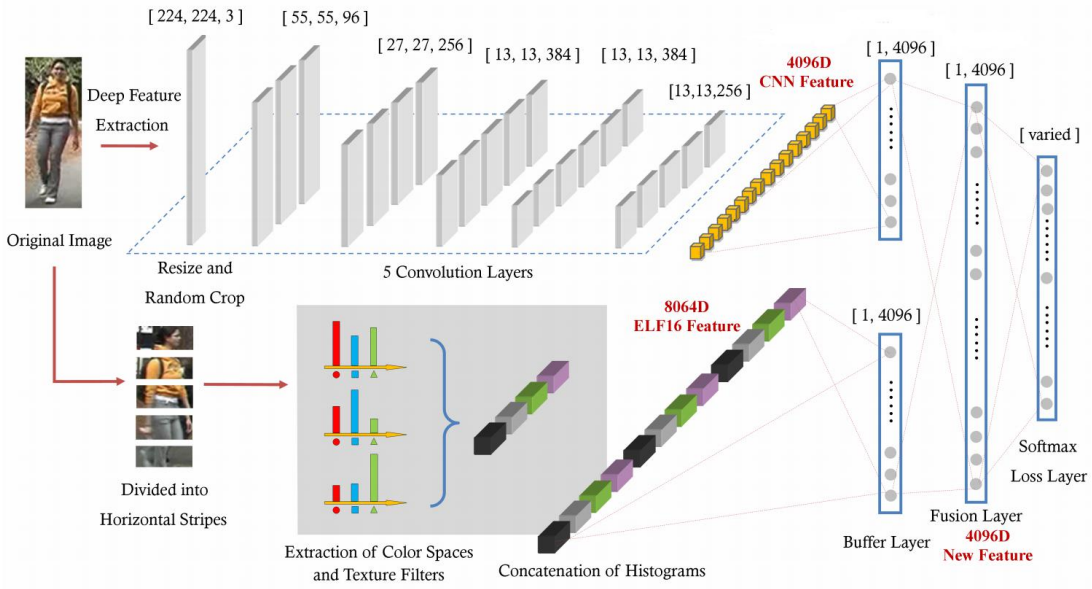

图 1-1 Feature Fusion Net 的网络结构示意图

本文首先提出了一种基于 CNN 的行人再标识图像特征提取方法（Feature Fusion Net，FFN，如图 1-1）。利用其反向传播、自我学习的特性，结合手工剪裁特征，我们得到了更有区分度的行人图像特征。在 VIPeR、CUHK01、PRID450s 三个通用行人图像数据集上的实验结果表明，我们提出的 FFN 特征能够有效改善不同分类器下行人图像匹配的效果。

其次，本文结合了传统的度量学习方法，提出了一个高效的行人图像匹配模型。这个模型利用了 FFN 图像特征，并结合了 Kernel Margin Fisher Analysis（KMFA）度量学习方法和针对行人图像优化的 Mirror 特征转换表示。实验证明，在上述三个行人再标识数据集中，Mirror KMFA 模型准确率分别超出当前最好结果 8.09%，7.98%和 11.20%。

第三，本文研究了一个行人再标识领域的新问题：基于不完整行人图像下的匹配问题。针对此问题，本文提出了一种基于 CNN 的网络模型 PartialNet（如图 1-2）。在 Partial REID 不完整行人图像数据集中，本文的模型表现出比较优秀的匹配效果。

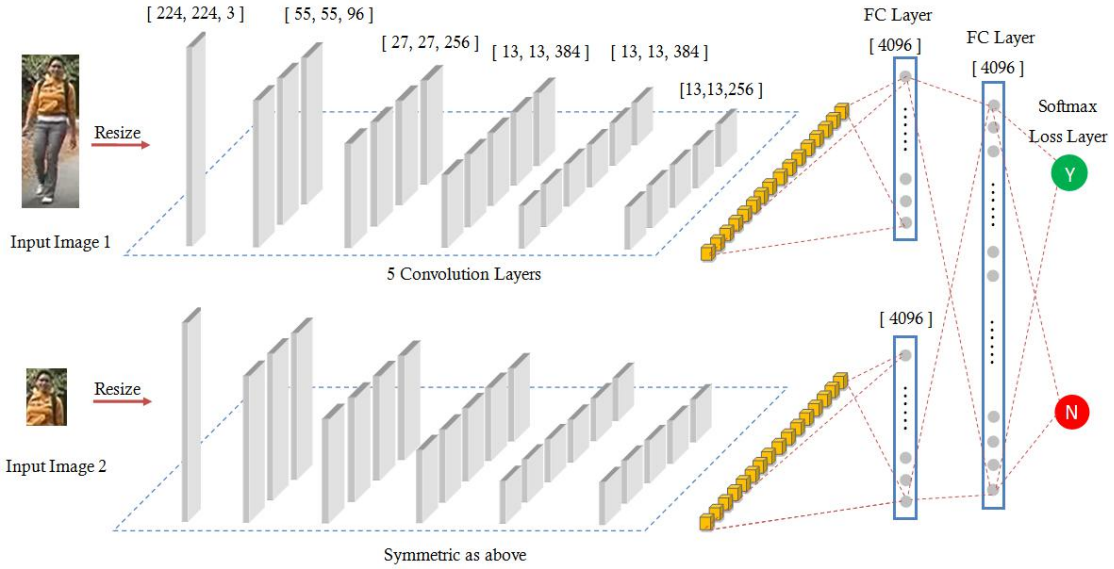

图 1-2 PartialNet 的网络结构示意图

1.5 论文结构与章节安排

本文一共分为五章，章节内容安排如下：

第一章为引言，简要介绍当前行人再标识（Person Re-identification）问题在学术界的研究进展。解决行人再标识问题的数学模型一般分为两个步骤：特征提取（Feature Extraction）与距离学习（Distance Learning）。本文的二至四章分别对这两个问题进行了深入探讨。

第二章着重研究了行人再标识问题中的图像特征提取方法。首先，本章介绍了几种常用的颜色、纹理特征，如 RGB、YCbCr、HSV 色彩空间，及 LBP、HOG、Gabor 纹理特征描述子。其次，本章分析了传统的行人再标识手工剪裁特征（Hand-crafted Feature），提出了一种新颖的卷积神经网络（Convolutional Neural Network）提取行人图片特征的方法 Feature Fusion Net。在实验环节中，FFN 特征被证明能够显著提升行人再标识问题的匹配率。

第三章详细介绍了几种常用的距离学习（计算）方法：L1-norm、LDA、LFDA 及 MFA。并根据[3]的方法，提出了一种适合 FFN 特征的度量学习（Metric Learning）方法 Mirror KMFA。在实验环节中，使用 Mirror KMFA 方法，配合第二章提出的 FFN 特征，匹配率可以超过当前最好结果 10%。

第四章探究了行人再标识领域的一个新问题：基于不完整图像的行人再标识问题。此问题由 Zheng et al.提出[43]，目前仍未得到深入研究。本章提出了一个基于卷积神经网络的模型 PartialNet，并在实验中证明其有效性。

第五章为结语。本章对本文进行了简要总结，并对未来的行人再标识研究工作进行了展望。

第 2 章 基于卷积神经网络的图像特征提取方法

图像特征表达是计算机视觉领域中大多数算法的第一个重要步骤。一个好的图像特征表达，在很大程度上影响着后续算法的准确度与时间空间复杂度。本章首先介绍了各种独立的图像特征表达，然后列举了目前在行人再标识领域比较常用的复合图像特征，最后结合卷积神经网络，提出了一种增强的行人图像特征提取方法。

2.1 传统图像特征提取策略

颜色和纹理是在图像特征表达中最有用的两项特性。例如，HSV 和 LAB 颜色直方图通常被用来衡量图像中的颜色信息。LBP 直方图[22]和 Gabor 滤波器描述了图像的纹理特征。近来，学术界倾向于使用这些特征的混合型特征[7,9,20,27,32,33]来获取更完整、准确的行人图像描述。下面，我们将分别介绍常用颜色特征、纹理特征与混合特征。

2.1.1 颜色特征

颜色特征是应用最为广泛的视觉特征，主要原因在于颜色往往和图像中所包含的物体和场景十分相关。同时，颜色特征对图像本身的尺寸、方向、视角的依赖性较小，具有较高的鲁棒性。此外，颜色特征计算量小，只需将图像中的像素点进行相应的转换，表现为数值即可。因此计算复杂度较低。颜色特征具有的这些优势使其能够很好的描述行人图像的颜色信息，特别适合用于行人再标识问题。

对于图像颜色特征的分析、理解与处理，是建立在特定的色彩空间之上的，图像在不同的色彩空间中的颜色特征表达不同。下面简要介绍本章方法中用到的三种色彩空间：RGB、YCbCr 与 HSV。

2.1.1.1 RGB 色彩空间

RGB 色彩空间是最直观、也是应用最为广泛的色彩空间。从颜色发光的原理解释，通过对红（Red）、绿（Green）、蓝（Blue）三个颜色通道的变化以及

它们相互之间的叠加，我们可以得到自然界任何一种色光。RGB 代表红、绿、蓝三个通道的颜色，这个标准几乎包括了人类视力所能感知的所有颜色，是目前运用最广的颜色系统之一。RGB 色彩模式已经成为工业界的一种颜色标准。256 级（256-bin）的 RGB 色彩总共能组合出 1600 万（$256\times256\times256\approx1.6\times10^{8}$）种色彩。

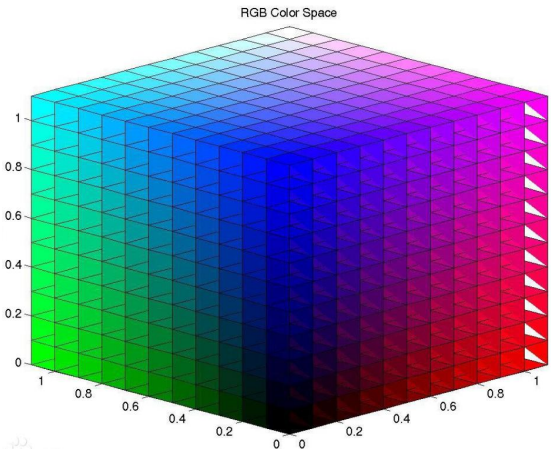

图 2-1 RGB 色彩空间示意图

2.1.1.2 YCbCr 色彩空间

RGB 色彩空间采用物理光学上的三种基本颜色表示，但不适应于所有的色彩应用。因而在学术界与工业界应用中，产生了其他不同的色彩空间表示法。YCbCr 与 HSV 就是其中应用较为广泛的两个。YCbCr 是用的色彩编码方案。其中 Y 是指亮度分量，Cb 指蓝色色度（Blue-difference）分量，Cr 指红色色度（Red-difference）分量[51]。YCbCr 色彩空间可以看作是 RGB 色彩空间的一个坐标转换，如图 2-2 所示。

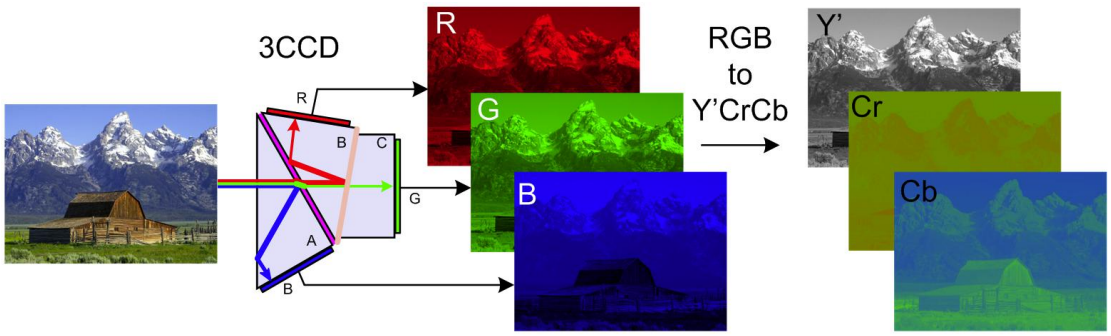

图 2-2 RGB 色彩空间转化为 YCbCr 色彩空间示意图[51]

2.1.1.3 HSV 色彩空间

HSV 色彩空间是最常用的圆柱体坐标（Cylindrical-coordinate）色彩空间[52]。在 HSV 色彩空间模型中，参数 H 表示色彩信息（Hue），即所处的光谱颜色的位置。H 用角度量表示，范围从 0° 至 360°，红、绿、蓝分别相隔 120 度，互补色分别相差 180 度。参数 S 表示饱和度（Saturation），范围从 0 到 1，表示所选颜色的纯度和该颜色最大的纯度之间的比率。当 S=0 时，只有灰度。参数 V 表示值（Value），表示色彩的明亮程度，范围从 0 到 1，但是它和光强度之间并没有直接的联系。HSV 是一种比较直观的颜色模型，在图像分析、计算机图像编辑中应用较为广泛。

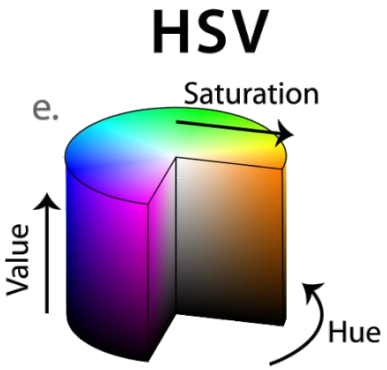

图 2-3 HSV 色彩模式示意图[52]

2.1.1.4 颜色直方图

颜色直方图（Color Histogram）通过统计图像像素值的分布情况来表示图像的颜色特征。其优点在于它能简单描述一幅图像中色彩空间信息的全局（Global）分布，即不同色彩在整幅图像中所占的比例，特别适用于描述那些难以界定边界的图像和不需要考虑内容的空间位置关系的图像。而颜色直方图的缺点是它无法描述颜色的局部分布特性，也不能够描述其纹理（Texture）信息。

2.1.2 纹理特征

纹理特征（Texture Feature）在图像分析中较为常用，其被用于描述图像或其中小块区域空间的纹理的粗细与疏密变化。纹理特征利用计算机技术从数字图像中计算出来的可以定量描述人对纹理的定性的感知的某些参数，对区域内部灰度变化或者色彩变化的某种规律进行量化。纹理特征具有旋转不变性，并且对于

噪声有较强的抵抗能力，但是其也有缺点，一个很明显的不足是当图像的光照和分辨率变化时，计算出来的纹理可能会有较大偏差。

2.1.2.1 LBP 描述子

局部二值模式（Local Binary Pattern，LBP）描述子[53]，是由 Ojala et al.在上世纪 90 年代提出的图像纹理特征描述方法。LBP 是一种通过统计法提取图像局部纹理特征的描述子，已成功地被运用于人脸检测、表情检测、动作识别等领域，也被广泛地运用于行人再标识研究。

LBP 描述子的相关提取方法简述如下：

1) 描述单个像素点的 LBP 描述子：以 3*3 窗口中心点为阈值（Threshold），将相邻 8 个点的灰度值与其进行比较，大于则记为 1，小于则记为 0。
2) 描述一张图的 LBP 特征：将检测窗口划分为 16*16 像素的 Cell，对于 Cell 中的每一个像素，计算 LBP 值。对于每个 Cell，得到直方图。将每个 Cell 的直方图串联在一起，得到总的 LBP 纹理特征向量。
3) 描述带有旋转不变性的 LBP 特征：原始的 LBP 描述子并不带有旋转不变性，而在实际的图像匹配应用中，特征的旋转不变性都是非常重要的环节。因此，Ojala et al.在改进的 LBP 描述子[53]中，通过不断旋转中心点圆形邻域，取使其达到最小 LBP 值的值为中心点 LBP 值，来达到描述子旋转不变性的目的。

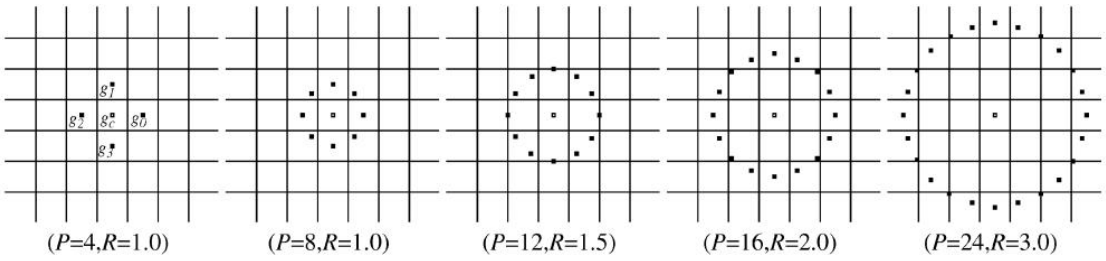

图 2-4 改进的 LBP 描述子关键点示意图[53]

2.1.2.2 HOG 特征

方向梯度直方图（Histogram of Oriented Gradients，HOG）[54]是 Dalal et al.提出的、广泛应用于计算机视觉和图像处理领域的一种特征描述子。在行人检测中，HOG 特征加 SVM 分类器的组合使用相当广泛。HOG 特征重要的思想是：

在一副图像中，局部目标的外观和形状能够被梯度或边缘的方向密度分布（直方图）很好地描述。HOG 特征描述了局部图像梯度的方向信息的统计值，通过对图像局部上的像素点提取梯度方向直方图，刻画了物体的表象和形状信息。将 HOG 特征用在行人图像的特征表达中，将光照与形变对图片的影响降至最低，可以对行人的信息有较好的描述（如图 2-5 所示）。

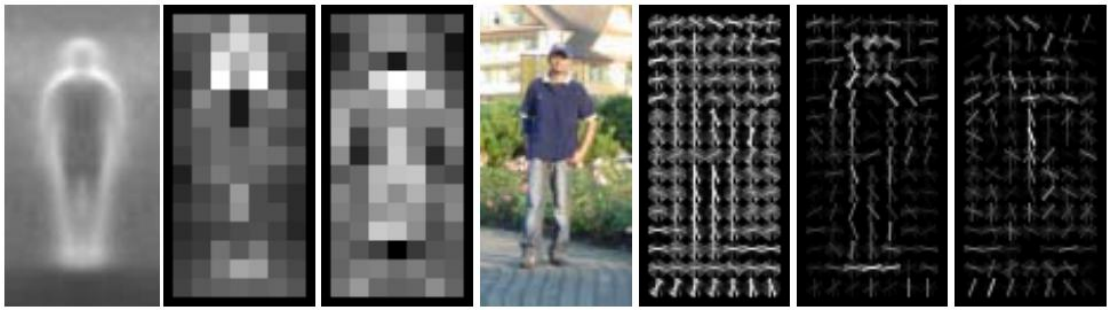

图 2-5 HOG 纹理特征示意图[54]

HOG 特征的提取方法简述如下：

1) 将图片进行灰度化与标准化：将彩色图片变换为灰度图，并使用 Gamma 变换对图片进行标准化操作，避免部分极大响应对 HOG 特征提取的影响。对图像中某一像素点 (x,y) 的 Gamma 变换的公式如下：

$$H(x,y)=H(x,y)^{\gamma} \tag{2.1}$$

工程实践中，通常去典型值 $\gamma=0.5$。

2) 计算每个像素的梯度（包括大小和方向）： 图像中某一像素点 (x,y) 的横向梯度 $G_x(x,y)$ 与纵向梯度 $G_y(x,y)$ 的计算公式为

$$\begin{aligned} G_x(x,y)&=H(x+1,y)-H(x-1,y)\\ G_y(x,y)&=H(x,y+1)-H(x,y-1) \end{aligned} \tag{2.2}$$

而梯度的大小和方向的计算公式为

$$\begin{aligned} G(x,y)&=\sqrt{G_x(x,y)^2+G_y(x,y)^2}\\ \alpha(x,y)&=\tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \end{aligned} \tag{2.3}$$

3) 将图像划分为小 Cell（例如 6*6 的像素为一个 Cell），统计每个 Cell 内

的梯度大小和方向，构成以方向为横坐标的直方图。

4) 将 Cell 合并成 Block（例如 3*3Cell 组成一个 Block），将一个 Block 内的所有 Cell 的特征串联起来。

5) 将所有 Block 的特征串联起来，构成一副图像的 HOG 特征。

注：在 HOG 特征提取的步骤中，没有方向矫正这一步。因此 HOG 特征没有旋转不变性。

2.1.2.3 Gabor 滤波器

Gabor 滤波器（Gabor Filter）属于加窗傅里叶变换，它实质上是在傅里叶变换中加入一个高斯窗函数，通过窗函数来实现时域、频域同时分析。Gabor 滤波器能够提取出不同尺度、不同频率的特征，能够很好地描述图像的局部信息。其可以写成如下形式：

$$g(x,y;\lambda,\theta,\varphi,\sigma,\gamma)=\exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\exp(i(2\pi\frac{x'}{\lambda}+\varphi)) \tag{2.4}$$

其中，x 与 y 为坐标，λ、θ、φ、σ、γ 为有相应物理意义的参数。

2.1.3 专为行人再标识设计的手工裁剪特征

近来，专为行人再标识设计的手工剪裁特征组合（Hand-crafted Features）显著地提升了匹配算法的准确率。Local Descriptors encoded by Fisher Vectors[19]使用 Fisher Vector 来构建图像描述子。Color Invariants[14]使用颜色的分布作为图像匹配的唯一线索。Symmetric-Driven Accumulation of local features[7]不仅证明了图像的对称分割可以显著提高特征的匹配准确率，而且提出了累积（Accumulative）的特征提取思想。Local Maximal Occurrence Features[18]分析了图像的横向局部特征，并使用最大池化（Max Pooling）的方法获得更稳定的行人再标识图像特征表达。

图 2-6 是一种在行人再标识中常用的手工剪裁特征[32]。其将输入图像横向分为六个区域，并分别对这六个子图进行颜色、纹理等特征的提取与统计（直方图）操作。最后，将六块区域的特征组合在一起，形成行人图像的完整特征表达。

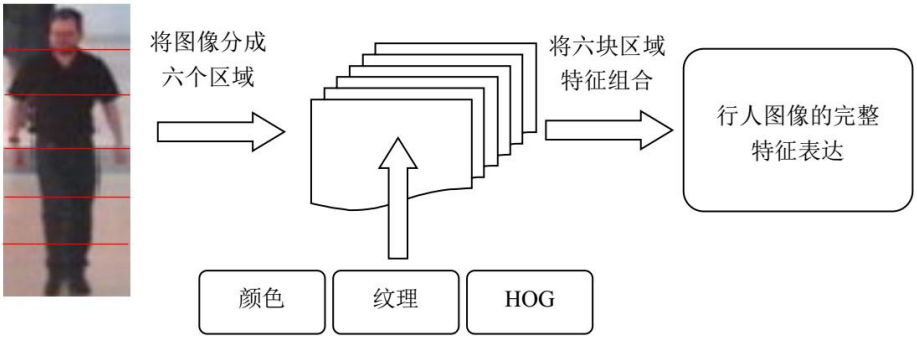

图 2-6 一种手工剪裁特征组合示意图[32]

2.2 基于神经网络的图像特征提取方法

随着神经网络（Neural Network）方法在近十年来的深入研究和计算机运算速度的快速提高，基于神经网络的图像特征提取在学术研究中逐渐成为可能。本节中，介绍两种实用的基于神经网络的图像特征提取方法：自编码器和卷积神经网络。

2.2.1 自编码器

自编码器（Autoencoder）是一种经典的无监督学习算法。利用反向传播（Back Propagation），尝试让网络输出值等于输入值。如图所示：

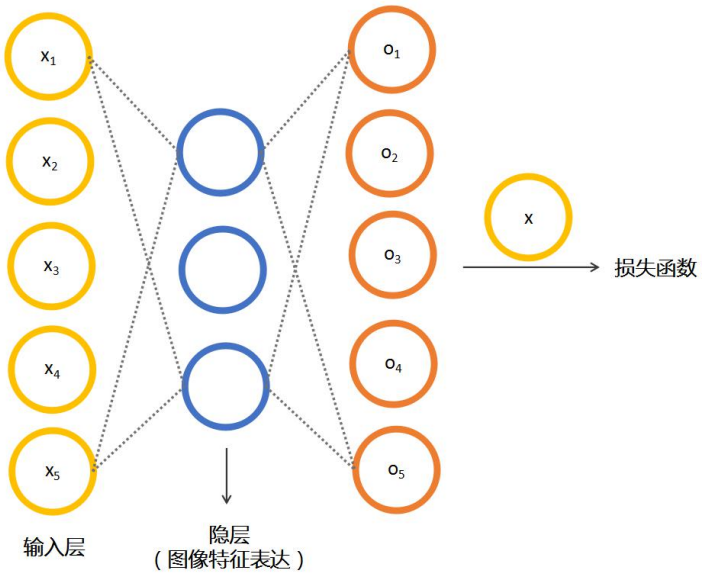

图 2-7 自编码器的网络结构示意图

对于自编码器，每一层（如图中的虚线）尝试学习一个 $Z(x)=H(W^Tx+b)$ 的函数。而到最后，其输出与输入进行比较并计算损失函数。即自编码器尝试通过反向传播学习网络权值，使输出 o 接近于输入 x。

为了使自编码器达到为图像“编码”的效果，需要加入约束条件（如限制隐层神经元数目），从而构造有意义的网络输出。设计良好的自编码器可以从数据中学习其压缩表示，即图像特征。例如，若输入图像尺寸为 100×100，隐层有 50 个神经元，则自编码器需要根据这 50 维数据重构出原图像。因此隐层的 50 维的数据必然包含输入数据相关性。原始的自编码器层数只有三层，而在实际应用中，多层自编码器能更好地提取图像特征表示。

在行人再标识领域，由于图像特征的复杂性，自编码器极少被应用于特征提取。[36]是一个利用自编码器来提取特征的例子。其使用了特殊训练方法，融合自编码器的无监督特点，构造了一种比较有效的图像特征表达（如图 2-8 中的 h^4 层）。

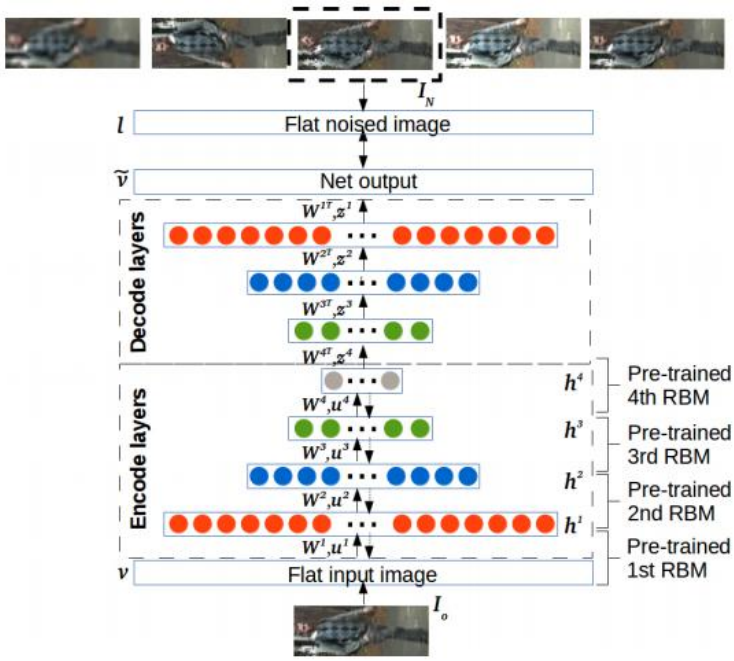

图 2-8 Franco et al.[36]提出的针对行人再标识问题的自编码器

2.2.2 卷积神经网络

卷积神经网络发展自传统（浅层）神经网络，通过加入卷积层、池化层和激活层，将传统网络中相对较为原始的模式识别方法扩展至图像识别领域。1998 年，LeCun et al.在[37]中使用卷积神经网络成功地将手写数字的识别错误率降低至 0.7%，并将此系统在商业领域成功运用。由此，人们对卷积神经网络方法愈发重视。2012 年，由 AlexNet[13]在 ImageNet 图像分类比赛中的突出表现，掀起了卷积神经网络在 AI 领域的又一次热潮。

图 2-9 为通过通用卷积神经网络进行特征提取的步骤：输入图像分为 R、G、B 三个通道，使用一组经过训练的滤波器（Convolution Filter）进行图像卷积操作。输入图像卷积后在第二层产生 32 个特征映射图（Feature Map）。然后对每个特征映射图进行池化（Pooling）操作，并

通过一个激活函数（通常为 Sigmoid 或者 ReLU 函数）产生第三层的特征映射图。这些映射图再经不断的卷积、滤波、激活操作，最后得到一个一维向量。此向量即可作为该图的特征表示。

使用生物学原理解释卷积神经网络，则卷积层的滤波器可看做人眼的一组局部感受野（Reception Field）。其作用为提取图像的某一局部特征。在提取局部特征的同时，其与其他特征间的位置关系也随之确定下来。池化层与激活层对应的是生物神经元的轴突，采用最大池化或平均池化操作来模拟神经细胞的抽象画过程，用激活函数来模拟神经细胞的非线性。

此外，卷积神经网络有共享权值的特性，因而减少了网络自由参数的个数，降低了网络参数选择的复杂度。卷积神经网络中不断的卷积、滤波、激活操作，使网络在识别时对输入样本有较高的畸变容忍能力。

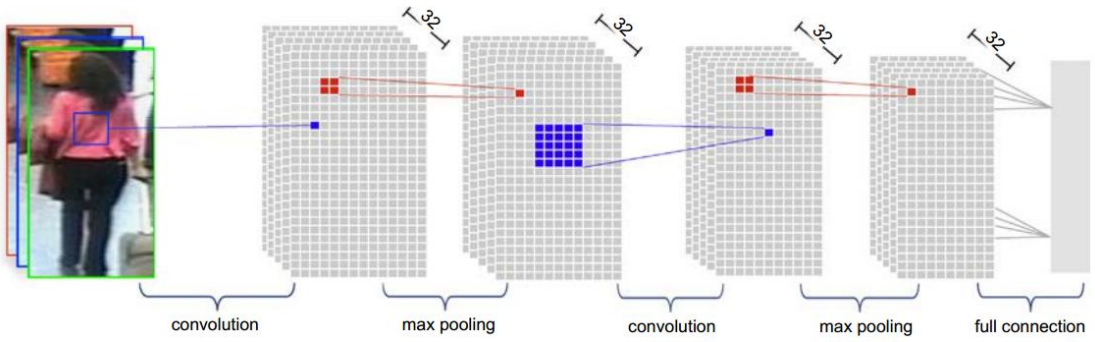

图 2-9 Ding et al.在[6]中提出的针对行人图像的卷积神经网络特征提取器

2.2.3 Feature Fusion Net

我们希望把 CNN 特征和手工选择的特征（颜色特征与纹理滤波器，简称为 ELF16 特征）同时映射至一个新的、统一的特征空间。因此，本章提出了 Feature Fusion Net（FFN）卷积神经网络结构。图 2-10 为该网络的示意图。

FFN 使用 ELF16 特征来影响 CNN 特征的提取，希望能让这两种特征更加互补。在我们的 FFN 框架中，通过反向传播（Back Propagation）过程，神经网络优化了图像卷积滤波器的参数学习过程（见下证明）。总的来说，作为特征融合（Fusion）的结果，最后输出的新特征应该比 CNN 特征和 ELF16 特征更加有区分度。

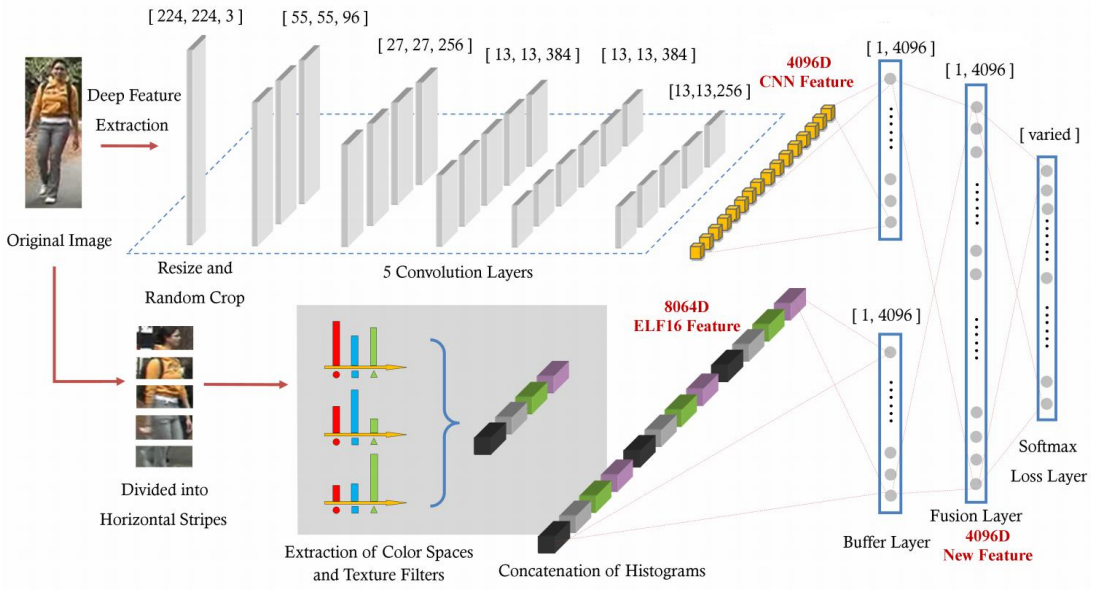

图 2-10 Feature Fusion Net 的网络结构示意图

表 2-1 Feature Fusion Net 网络结构参数表（详见附录 B）

层	类型	输入层	输出层	输出维度	参数个数
Data	图像输入层	-	Data	227×227×3	-
Conv1	卷积层	Data	Conv1	55×55×96	11×11×3×96
ReLU1	激活层	-	-	-	-
Pool1	池化层	Conv1	Pool1	27×27×96	-
Norm1	正则层	Pool1	Norm1	-	-
Conv2	卷积层	Pool1	Conv2	27×27×256	5×5×48×256
ReLU2	激活层	-	-	-	-

Pool2	池化层	Conv2	Pool2	13×13×256	-
Norm2	正则层	Pool2	Norm2	-	-
Conv3	卷积层	Norm2	Conv3	13×13×384	3×3×256×384
ReLU3	激活层	-	-	-	-
Conv4	卷积层	Conv3	Conv4	13×13×384	3×3×384×384
ReLU4	激活层	-	-	-	-
Conv5	卷积层	Conv4	Conv5	13×13×256	3×3×384×256
ReLU5	激活层	-	-	-	-
Pool5	池化层	Conv5	Pool5	6×6×256	-
FC6	全连接层	Pool5	FC6	4096	6×6×256×4096
ReLU6	激活层	-	-	-	-
Drop6	激活层	-	-	-	-
FeatureData	输入层	-	FeatureData	1×8064	-
FeatureData-FC	全连接层	FeatureData	FeatureData-FC	4096	8064×4096
FeatureData-ReLU	激活层	-	-	-	-
FeatureData-Drop	激活层	-	-	-	-
Concat	辅助层	FC6、FeatureData-FC	Concat	8192	-
FC7	全连接层	Concat	FC7	4096	8192×4096
ReLU7	激活层	-	-	-	-
Drop7	激活层	-	-	-	-
FC8	全连接层	FC7	FC8	变动	变动
SoftmaxLoss	损失函数层	FC8	SoftmaxLoss	-	-

表 2-1 中，带“-”的表示本地操作：直接对数据块内的每一个数据进行改变，而不会将其复制至新的数据层。此操作不会产生新的参数。这种层次化的框架设计有利于减少程序内存消耗、提高神经网络的运算速度。

2.2.4 Fusion 层和 Buffer 层

我们将 Concat 辅助层之后的 FC7 全连接层称为 Fusion 层。Fusion 层使用全连接来增强对 CNN 特征和 ELF16 特征的自适应能力。

如果将 Fusion 层的输入表示为

$$x = [ELF16, \mathrm{CNN}\_Feature] \tag{2.5}$$

那么这一层的输出可以表示为

$$Z_{Fusion}(x)=h(W_{Fusion}^{T}x+b_{Fusion}) \tag{2.6}$$

其中，$h(\cdot)$ 代表激活函数。ReLU 和 Dropout 层紧跟于 Fusion 层后面，Dropout 层的随机参数（Dropout Ratio）设置为 0.5.根据反向传播算法，在进行一次反向传播之后，第 l 层的参数可以更新为：

$$\begin{aligned} W_{new}^{(l)} &= W^{(l)}-\alpha[(\frac{1}{m}\Delta W^{(l)})+\lambda W^{(l)}] \\ b_{new}^{(l)} &= b^{(l)}-\alpha[\frac{1}{m}\Delta b^{(l)}] \end{aligned} \tag{2.7}$$

其中，α 称为学习率（Learning Rate），λ 称为动量（Momentum）。这两个参数设置见 2.3.4 节。

现有的使用卷积神经网络解决行人再标识问题的模型，一般使用 Deviance Loss[10]

$$J_{devian}(x,y,l)=\ln(e^{-\cos(B_1(x),B_2(y))l}+1) \tag{2.8}$$

或者 Hinge Loss[1]

$$J_{hinge}(x)=\max(0,\ 1-lx) \tag{2.9}$$

作为损失函数。而在 FFN 中，我们要解决的是一个多分类问题。因此，Softmax Loss 损失函数被应用在网络结构的最后一层。对于一个输入向量 x 和一个 FC8 层的单个输出结点 j，每个类别的预测概率可以用以下公式计算：

$$p(y=j\mid x;\theta)=\frac{e^{\theta_j^{T}x}}{\sum_{k=1}^{n}e^{\theta_k^{T}x}} \tag{2.10}$$

FFN 网络的最后一层被设计为最小化 Softmax Loss。令 J 表示对于单个样本的在所有输出节点损失函数之和，则：

$$J_{soft\max}(\theta)=-[\sum_{k=1}^{m}y^{(i)}\log h_\theta(x^{(i)})+(1-y^{(i)})\log(1-h_\theta(x^{(i)}))] \tag{2.11}$$

式中的 n 指 FC8 层的输出节点个数。n 的选择见 2.3.4 节。Softmax Loss 损失函数的导数为

$$\nabla_{\theta^{(k)}}J(\theta)=-\sum_{i=1}^{m}[x^{(i)}(1\{y^{(i)}=k\}-P(y^{(i)}=k\mid x^{(i)};\theta)] \tag{2.12}$$

用式（2.12）来更新 FC8 层的参数，并使用链式求导规则逐层更新整个网络的参数。

2.2.5 手工挑选的特征如何影响卷积神经网络特征？

如果 FFN 网络的参数被手工挑选的特征（即 ELF16 特征）$\tilde{x}$ 影响，即网络参数的梯度会根据 $\tilde{x}$ 来进行调整，那么我们就可以说，在 FFN 网络结构图（图 2-10）中，

$$\delta_i^n = \frac{\partial J}{\partial Z_i^n} \tag{2.13}$$

记 CNN 特征（FC7 层特征）为 x，记 ELF16 特征为 $\tilde{x}$，把连接第 n 层的第 j 个节点和第 $n+1$ 层第 i 个节点的权值记为 W_{ij}^n。则

$$\begin{aligned} Z_j^n &= \sum\nolimits_j W_{ji}^{n-1} a_i^{n-1} \\ a_i^{n-1} &= h(Z_i^{n-1}) \end{aligned} \tag{2.14}$$

记 $Z_j^8 = \sum_j W^{n-1} x_i^{n-1}$.我们要证明，使用反向传播算法后，$\frac{\partial j}{\partial W_{ij}^7}$ 被 $\tilde{x}$ 影响。因此，CNN 特征将会学习到与 ELF16 特征互补的特征。记

$$\begin{aligned} \delta_i^8 &= \left(\sum_j W_{ji}^8 \delta_j^9 \right) h'\left(Z_i^8\right) \\ \delta_i^9 &= \left(\sum_k W_{kj}^9 \delta_k^{10} \right) h'\left(Z_i^9\right) \\ \frac{\partial J}{\partial W_{ij}^7} &= x_j \delta_i^8 \end{aligned} \tag{2.15}$$

那么，$\tilde{x}$ 会通过两个方面影响 δ_j^9。首先，

$$\begin{aligned} \tilde{a}_j^8 &= h(\sum_i \tilde{W}_{ji}^7 \tilde{x}_i) \\ Z_k^9 &= \sum_j W_{kj}^8 a_j^8 + \sum_j \tilde{W}_{kj}^8 \tilde{a}_j^8 \end{aligned} \tag{2.16}$$

换句话说，ELF16 特征 $\tilde{x}$ 中包含的信息可以通过 $h'(Z_j^9)$ 传播。因此，FFN 中，Conv1

至 Conv5 层的滤波器会通过反向传播适应 $\tilde{x}$。

其次，SoftmaxLoss 层的输出会受 $\tilde{x}$ 的影响，因此 δ_k^{10} 也会受 $\tilde{x}$ 影响。

2.2.6 激活函数

在最初神经网络数学模型定义中，Kohen et al[50]为了模拟生物神经网络结构，增加神经网络的非线性（Non-linearity）性质，引入了激活函数（Activation Function）。理想中的激活函数是 0-1 阶跃函数，其将输入值映射为 0 或 1 的输出值。0 代表神经元抑制，而 1 代表神经元兴奋。

然而，0-1 阶跃函数具有不连续、不可导等数学性质，使其无法很好地适应神经网络的反向传播（Back Propagation）更新算法。因此，数学家通过经验得出了较为通用的一些激活函数。这些函数既能模拟神经元的兴奋、抑制状态，又具有光滑可导的优秀数学性质。激活函数的应用，大大增加了神经网络的适应性，提高了其在大部分分类问题上的准确率。

以下为在卷积神经网络中常用的四个激活函数：

1) Sigmoid 激活函数

$$f(z)=Sigmoid(z)=\frac{1}{1+e^{-z}} \tag{2.17}$$

2) TanH 激活函数

$$f(z)=TanH(z)=\frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{2.18}$$

3) ReLU 激活函数

$$f(z)=\mathrm{Relu}(z)=\max(0,z) \tag{2.19}$$

4) PReLU 激活函数

$$f(z)=\mathrm{PRelu}(z)=\begin{cases}\alpha\, z, & z>0\\ z, & z\le 0\end{cases} \tag{2.20}$$

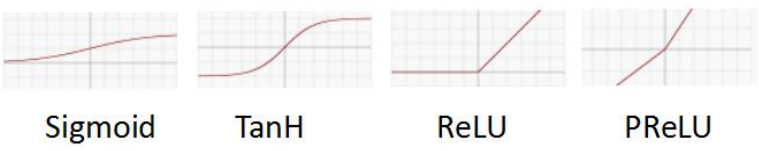

图 2-11 四种常用激活函数示意图

Sigmoid 和 TanH 激活函数在 20 世纪末的浅层神经网络方法中较为流行。目前，随着深度神经网络方法的大规模应用，ReLU 和 PReLU 激活函数正在逐渐得到运用。本章 FFN 模型方法中，大量使用了 ReLU 激活函数。

2.3 Feature Fusion Net 网络训练设置

2.3.1 训练数据集

Market-1501 是清华大学于 2015 年公开的行人再标识数据集。它由 1501 个 ID 下的 38195 幅行人图像组成，是一个多摄像头混合的数据集。相比于传统行人再标识数据集（VIPeR、CUHK 系列、ETHZ、iLIDS 等），它的优势是图像数量多，适宜用于深度神经网络的训练。我们使用 Market-1501 数据集对 Feature Fusion Network 进行训练，然后使用这个神经网络来提取测试数据集（VIPeR、CUHK01、PRID450s）的行人图像特征。

表 2-2 数据集基本信息

数据集名称	Market-1501
总图像数量	39185
总 ID 数量	1501
训练集 ID 数量	1501
摄像头数目	多摄像头混合，不分视角
每个 ID 总的图像数目	~30，不定

图 2-12 Market-1501 数据集部分样本

2.3.2 训练策略

分批随机梯度下降（Mini-batch Stochastic Gradient Descent）[2]以一批样本为单位，计算每个样本输入的损失（Loss）的平均值，并以此作为反向传播的损失。相对于传统的以单个样本输入为单位进行反向传播的梯度下降算法，Mini-batch 方法能得到更快的反向传播速度和更平滑的收敛。

在 FFN 训练时，每个循环（Iteration，一批样本的一次完整前向传导和反向传播称为一个循环）中，25 张行人图片组成一批。每批样本中的每一幅图像均被前向传导（forward）至最后的 Softmax 损失层（Softmax Loss Layer）。初始学习率（Learning Rate）被固定在$\gamma_{initial}=1e-5$。这是一个远远小于其他 CNN 模型的初始学习率。随着 FFN 训练过程的进行，每经过 20000 次循环，学习率根据$\gamma_{new}=0.1\times\gamma_{old}$降低。我们使用 ImageNet[13]的参数初始化 FFN 中 *Conv*1 至 *FC*6 层的参数，并按照以上训练规则进行训练。经过大约 50000 次循环后，FFN 网络开始收敛。在单块 NVIDIA Tesla K20m GPU 的实验机器上，这个过程大约花费 4 个小时。为了提高 FFN 模型的表达能力，我们进一步使用 Market-1501 数据集中的困难样本（Difficult Samples）对其进行微调。

2.3.3 困难样本微调

困难样本微调（Hard Negative Mining）[1]给我们提供了一种增强 CNN 网络针对不同分类目标迁移能力的方法。这个增强的训练策略本意是为了解决二分类训练数据集中的正负样本对不均匀问题。我们在 FFN 网络的训练中应用这个策略。在前一步训练中，被错分的样本（大概 630 个 ID 的 12000 张图片）被人工挑出，并放入 FFN 中继续微调。

微调时，把网络最后一层 FC8 的输出节点数降至 630，并使用更低的学习率（$\gamma'_{initial}=1e-6$）和更少的循环数（约 10000 次）。整个微调过程花费大约 1-2 小时，至网络的损失（Loss）降至一个合理的水平（典型值为~0.05）。

2.4 基于行人再标识特征的实验

2.4.1 测试数据集介绍

本节实验基于三个通用行人再标识数据库（VIPeR[8]，CUHK01[15]和

PRID450s[24]）。本节简要介绍这三个数据库。在图片展示中，上下两行分别代表两个不同的摄像头视角（Camera-A, Camera-B）。

2.4.1.1 VIPeR

Viewpoint Invariant Pedestrian Recognition（VIPeR）数据集[8]是学术界最通用的行人再标识基准测试数据集。该数据集共 1264 张图片，包含 2 个摄像头下、每个摄像头 632 个人（ID）、每个人 1 张的剪裁过的图片。图片大小为 128×48。VIPeR 数据集中，图像背景集中于校园街道，人体轮廓不清晰，视角变化幅度较大。此数据集目前的算法准确率（以通用测试规则计）在 30%-40%之间。图 2-13 展示了 VIPeR 数据集下的一些图像。每列的两幅图像属于同一个 ID。

图 2-13 VIPeR 数据集部分样本

2.4.1.2 CUHK01

CUHK 数据集[16]是由香港中文大学的 Zhao et al.近年来发表的一系列行人再标识基准测试数据集。本文使用其中应用最广泛的 CUHK01 数据集。该数据集共 1264 张图片，包含 2 个摄像头下、每个摄像头 971 个 ID、每个人 2 张的经过统一剪裁的图片。图片大小为 160×60。CUHK01 数据集拍摄于较暗环境下的行人天桥，整体图片亮度较低。此数据集目前的算法 Rank-1 准确率（以行人再标识通用测试规则计）在 30%-45%之间。图 2-14 展示了 CUHK01 数据集下的一些图像。每两列图像属于同一个 ID。

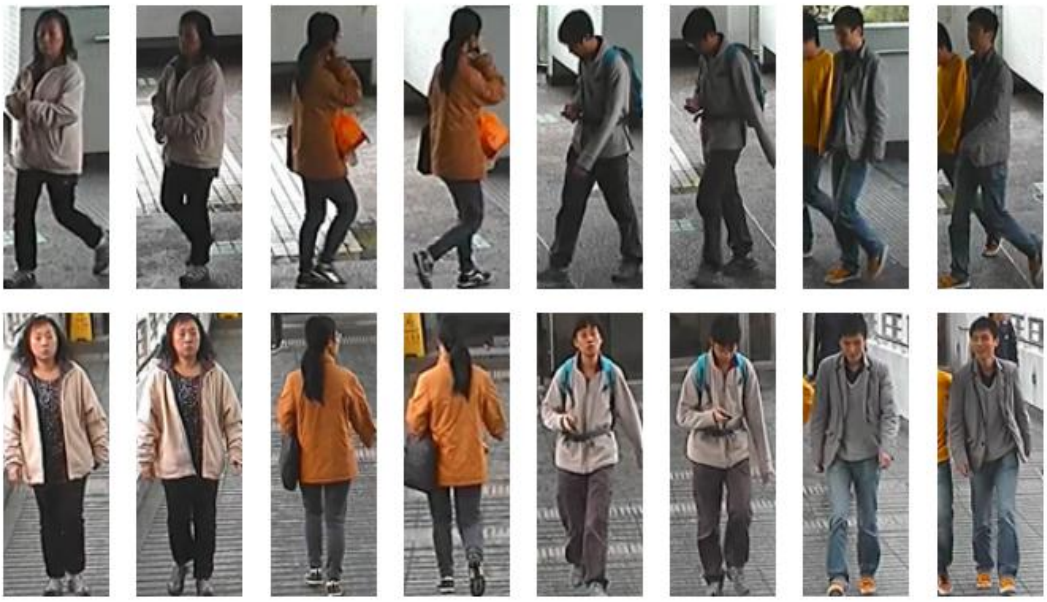

图 2-14 CUHK01 数据集部分样本

2.4.1.3 PRID450s

Person Re-ID（PRID）数据集[24]是由 Graz et al.建立的行人再标识基准测试数据集系列，包括 PRID450s 静态数据集和 PRID2011 视频数据集。本文中使用 PRID450s 数据集。该数据集共 900 张图片，包含 2 个摄像头下、每个摄像头 450 个 ID、每个 ID 包含 1 张的剪裁过的图片。图片大小在 160×80 附近变动。PRID450s 数据集拍摄环境比较统一，识别匹配成功率较高，Rank-1 准确率（以行人再标识通用测试规则计）在 50%-60%之间。图 2-15 展示了 PRID450s 数据集下的一些图像。每两列图像属于同一个 ID。

图 2-15 PRID450s 数据集部分样本（注：该数据集图像尺寸不统一）

2.4.1.4 总结

针对行人再标识问题，不同的数据集有不同的通用测试策略，而策略的微小变动可以导致实验结果的巨大差异。表 2-3 总结了 VIPeR、CUHK01、PRID450s 三个数据集的基本信息与通用测试规则。

表 2-3 测试数据集基本信息

数据集名称	VIPeR	CUHK01	PRID450s
总图像数量	1264	3884	900
总 ID 数量	632	971	450
训练集 ID 数量	316	485	225
摄像头数目	2	2	2
每个 ID 在单个摄像头下的图像数目	1	2	1

2.4.2 实验规则

在每个独立实验中，随机地选取一半 ID 作为训练数据集（Training Set），另一半则作为测试数据集（Testing Set）。在度量学习方法中，我们使用训练数据集来学习并得到投影矩阵 W。在测试数据集中，使用 $x'=W^Tx$ 来得到测试样本 x 的最终投影向量，并对每一对输入图像距离进行计算。

为了测试的准确性和可靠性，每个实验重复十次，并取平均的 Rank-i 值记录于 CMC 曲线中。

2.4.2.1 Rank-i 值

Rank-i 是行人再标识与人脸识别（Face Recognition）等问题中常用的准确率衡量方法。下面给出 Rank-i 的定义：

对于测试数据集，给定一幅 Probe 图像，将 Gallery 图像集中的每一幅图像按照与其关联得分（即距离 $D(x_i,x_j)$ 的大小）升序排序。取前 i 个 Gallery 图像样本，若 Probe 图像与这 i 个样本中的某个样本属于同一 ID，则判断其被正确匹配。

2.4.2.2 CMC 曲线

CMC 曲线即累积匹配特性曲线（Cumulative Match Characteristic），表示某个模型的 Rank-1 至 Rank-i 匹配率。本文中，令 $i=20$，统一使用此 CMC 曲线来对模型匹配率进行对比。

2.4.2.3 分类器

本章实验重点在于对 FFN 特征和其他常用行人再标识特征进行比较。因此，在分类器的选择上，使用一种非监督式算法（L1-norm 距离）与一种监督学习算法（LFDA，行人再标识领域常用度量学习算法）。用 $D(x_i,x_j)$ 代表使用分类算法得到的两幅行人图像之间的距离，并得出 Rank-i 值和 CMC 曲线。对分类器的详细介绍，见 3.1 节。

2.4.2.4 特征维度

对于每种行人图像特征，本章实验均采用其默认维度（见表 2-10）。

2.4.3 实验结果与结果分析

2.4.3.1 非监督式算法

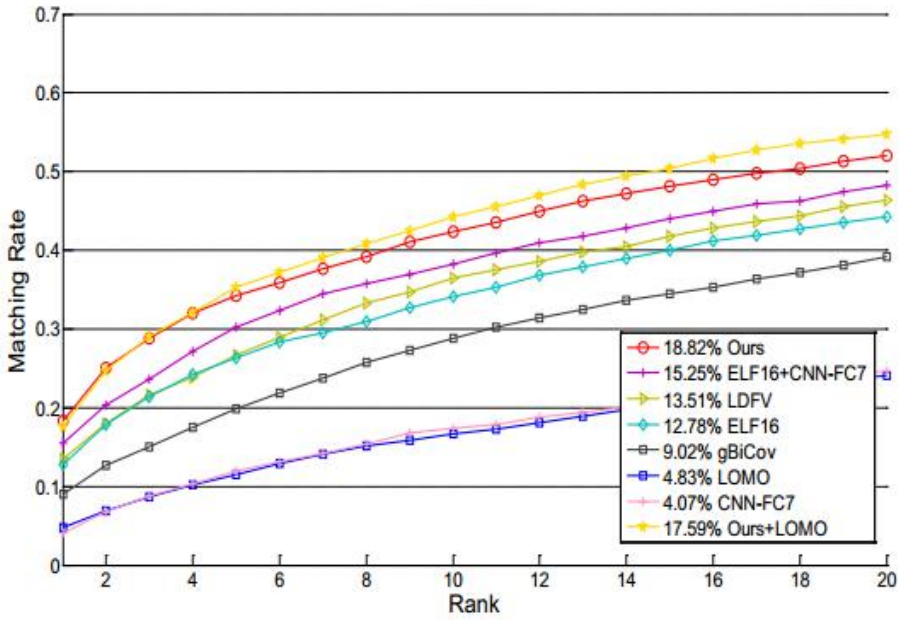

图 2-16 VIPeR 数据集、L1-norm 分类器下各种特征的 CMC 曲线

表 2-4 各种特征在 VIPeR 数据集、L1-norm 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	**18.82**	**34.18**	**42.37**	**52.02**
ELF16+CNN-FC7	15.25	35.25	44.21	54.78
LDFV	13.51	26.31	36.49	46.42
ELF16	12.78	26.31	34.06	44.22
gBiCov	9.02	19.90	28.76	39.20
LOMO	4.83	11.47	16.65	24.05
CNN-FC7	4.07	11.91	17.37	24.68

我的特征+LOMO 特征	17.59	35.25	44.21	54.78

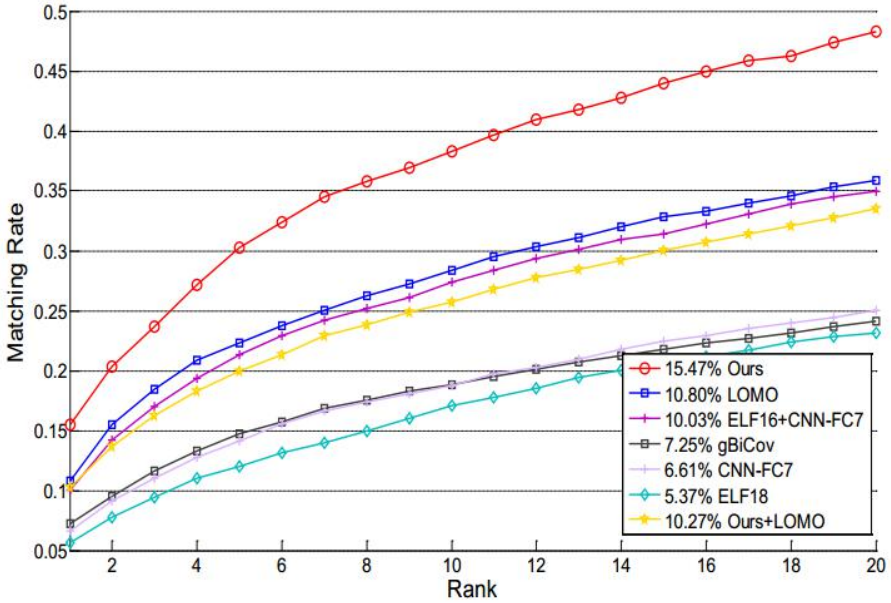

图 2-17 CUHK01 数据集、L1-norm 分类器下各种特征的 CMC 曲线

表 2-5 各种特征在 CUHK01 数据集、L1-norm 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	**15.47**	**30.25**	**38.29**	**48.26**
LOMO	10.80	22.27	28.40	35.88
ELF16+CNN-FC7	10.03	21.35	27.39	34.59
gBiCov	7.25	14.74	18.75	25.03
CNN-FC7	6.61	14.09	18.75	25.03
ELF16	5.37	11.99	17.10	23.17
我的特征+LOMO 特征	10.27	19.98	25.74	33.52

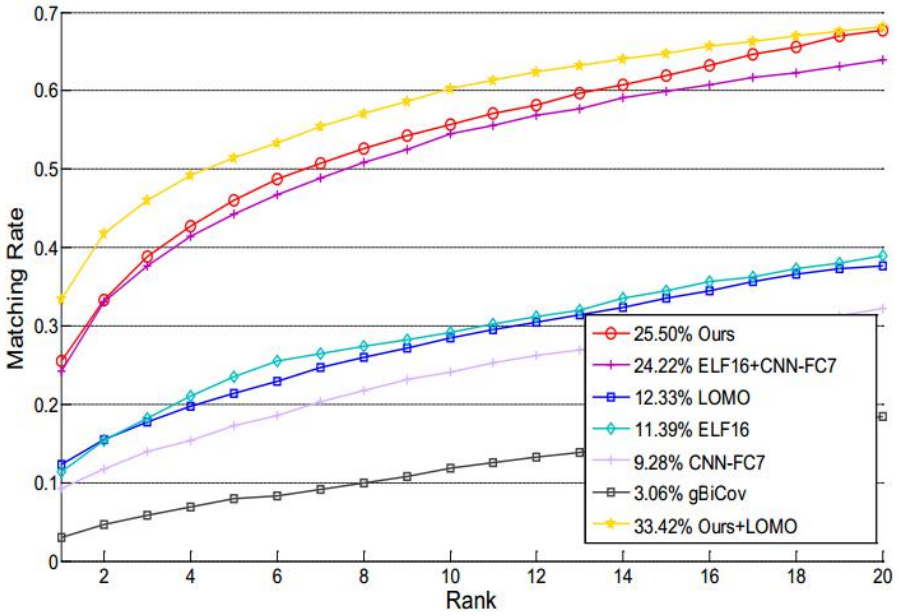

图 2-18 PRID450s 数据集、L1-norm 分类器下各种特征的 CMC 曲线

表 2-6 各种特征在 PRID450s 数据集、L1-norm 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	**25.50**	**51.42**	**60.22**	**68.18**
ELF16+CNN-FC7	24.22	46.04	55.64	67.64
LOMO	12.33	23.56	29.22	38.94
ELF16	11.39	21.39	29.22	37.67
CNN-FC7	9.28	17.22	24.11	32.28
gBiCov	3.06	7.94	11.83	18.50
我的特征+LOMO 特征	33.42	51.42	60.22	68.18

图 2-16 至图 2-18 展示了在 L1-norm 距离下，我们的特征与其他特征的比较。L1-norm 距离是最基本的无监督分类算法，用于衡量各种特征最直观的区分度。实验结果有力地说明，FFN 特征所表现出的区分度大大超越了其他的特征。这说明，FFN 特征中提供的原始信息比其他特征更多、更有用。

ELF16+FC7 特征在 L1-norm 距离下的识别准确率仅次于 FFN 特征，超过了 ELF16 和 CNN-FC7 两个单独的特征。这从结果上证明了我们的假设，即传统特征和 CNN 特征是在某种程度上是互补的。FFN 特征也大大地超越了 ELF17+FC7 联合特征的准确率。这可能基于以下几个原因：

1) 在 FFN 网络的训练中，我们有意识地引导 CNN 特征与 ELF16 特征的融合过程，希望 CNN 特征的卷积核（Filters）通过反向传播过程，学习到与 ELF16 特征互补的特性。而在 ELF16+CNN-FC7 特征中，这两个特征

只是简单地连接在一起，而这种连接并不是最优化的。

2) Buffer 层和 Fusion 层的使用可以使网络自动地调整相应全连接层中的 w 和 b 参数。这使得输出的 FFN 特征在表达性上更胜一筹。

LOMO 特征是 Liao et al.针对行人再标识问题在[18]中提出的一种新的特征。然而，在 VIPeR 数据库上，它的准确率排名第七；在 CUHK01 数据库上，它的准确率排名第三。这种波动的排名说明，LOMO 特征在 L1-norm 下的准确率表现不够稳定。

2.4.3.2 监督学习

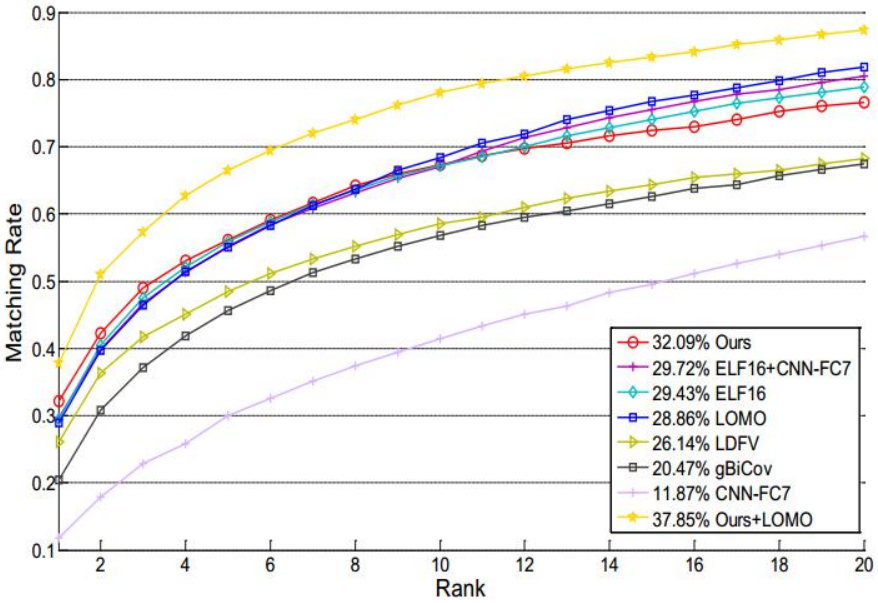

图 2-19 VIPeR 数据集、LFDA 分类器下各种特征的 CMC 曲线

表 2-7 各种特征在 VIPeR 数据集、LFDA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	32.09	55.85	67.18	76.65
ELF16+CNN-FC7	29.72	55.03	67.18	80.44
ELF16	29.43	55.85	67.18	78.96
LOMO	28.86	55.03	66.55	81.90
LDFV	26.14	48.48	58.54	68.26
gBiCov	20.47	45.60	56.84	67.44
CNN-FC7	11.87	30.00	41.42	56.71
我的特征+LOMO 特征	37.85	66.46	78.13	87.34

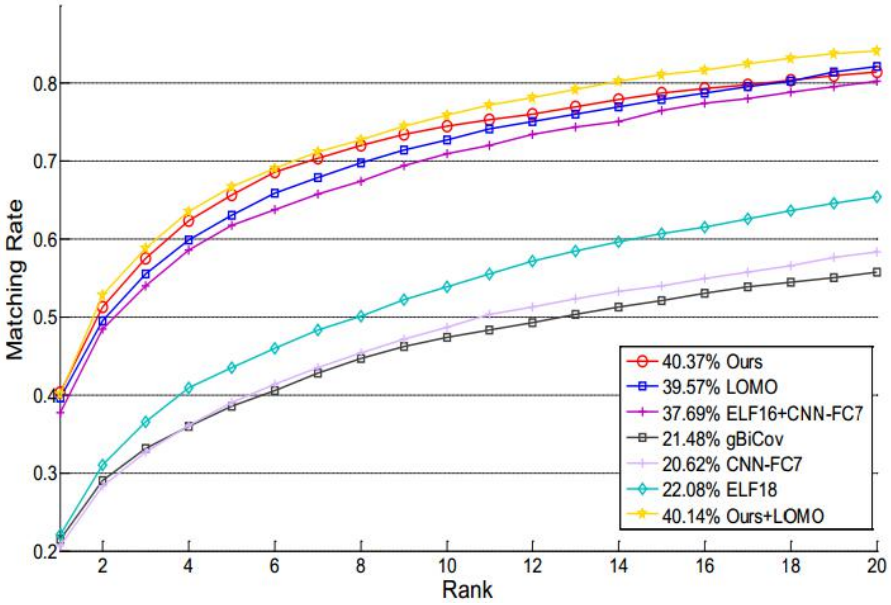

图 2-20 CUHK01 数据集、LFDA 分类器下各种特征的 CMC 曲线

表 2-8 各种特征在 CUHK01 数据集、LFDA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	**40.37**	**65.61**	**74.48**	**81.42**
LOMO	39.57	63.11	72.72	82.14
ELF16+CNN-FC7	37.69	61.73	71.02	80.25
gBiCov	21.48	38.54	47.37	55.72
CNN-FC7	20.62	39.03	47.37	55.72
ELF16	22.08	43.56	53.85	65.41
我的特征+LOMO 特征	40.14	66.67	75.86	84.22

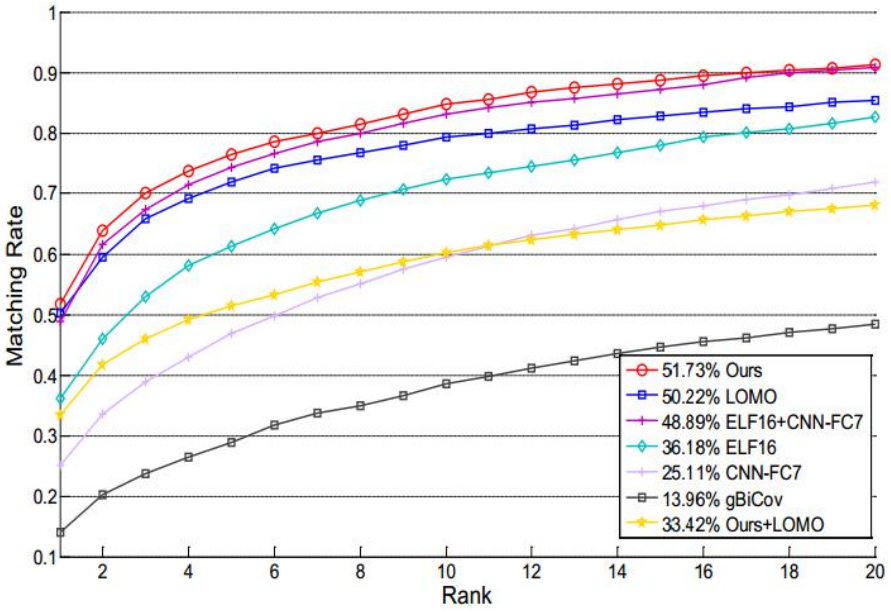

图 2-21 PRID450s 数据集、LFDA 分类器下各种特征的 CMC 曲线

表 2-9 各种特征在 PRID450s 数据集、LFDA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
我的特征	**51.73**	**76.44**	**84.76**	**90.89**
LOMO	50.22	71.87	79.24	85.42
ELF16+CNN-FC7	48.89	74.27	83.11	90.89
ELF16	36.18	61.33	72.40	82.67
CNN-FC7	25.11	51.42	60.22	68.18
gBiCov	13.96	28.93	38.62	48.40
我的特征+LOMO 特征	33.42	51.42	60.22	68.18

为了展示出 FFN 特征的最大有效性，我们把它放入两种监督式学习方法中进行实验（LFDA[23]和 Mirror KMFA[3]）。本文中，LFDA 在第二章进行实验，Mirror KMFA 在第三章进行实验。这两种方法均是度量学习（Metric Learning）方法。同时进行比较的还有其它在行人再标识模型中广泛使用的图像特征（如 LDFV、gBiCov、LOMO 等）。对于每种特征，用训练集来学习一个映射矩阵 W。将测试集用投影矩阵 W 映射至新空间，并通过计算每对图片之间的相对距离来测试他们的准确率。

图 2-19 至图 2-21 展示了在三个数据集上的 CMC 曲线。曲线上标注了每种特征的 Rank-1 识别准确率。

实验结果清楚地展示出 FFN 特征的强大区分能力。在 VIPeR 和 CUHK01 两个库中，它超过了其他所有的独立特征。对比 ELF16 特征和 CNN-FC7 特征，FFN 特征有更好的表现。同时，这两个特征的简单拼接（ELF16+CNN-FC7）也不能

在度量学习方法中，FFN 特征在大多数情况下比 LOMO 特征表现得好。PRID450s 数据集上，因为 LOMO 特征使用 HSV 颜色直方图和 SILTP 纹理算子，对于特定的光照环境，它的表现比 FFN 特征要好。而在 VIPeR 和 CUHK01 数据集中，FFN 特征表现出了绝对的优势。

2.4.3.3 特征提取时间

表 2-10 评估了每一种行人再标识图像特征的提取时间，其中的数据基于提取一张 VIPeR 数据集中 128×48 图像的平均时间。在第一行中，我们已经把提取 ELF16 特征的时间计算在内。

可以看到，FFN 特征的提取速度甚至比某些手工裁剪特征（例如 gBiCov）

更快。这打破了人们对于深度神经网络方法时间复杂度的刻板印象。

对比 LOMO 特征，我们的特征拥有更低的维数，因此在接下来的分类器步骤中可以获得更快的运算速度。

FFN 在提取速度和特征维度上取得了很好的平衡，因此可以在实际应用中发挥很大作用。除此之外，对比其他 CNN 模型，我们的 FFN 网络不需要在测试数据集上进行微调，因此可以在便利性上取得优势。

表 2-10 各种特征的提取时间、输出维度比较

特征	提取时间	默认维度
FFN 特征	**0.1796s+0.5720s**	**4096**
ImageNet CNN-FC7 特征	0.1773s	4096
ELF16 特征	0.5720s	8064
LOMO 特征	0.2610s	26960
gBiCov 特征	13.6152s	5940

2.5 本章小结

本章首先介绍了行人再标识领域通用的几种常用图像特征：RGB、YCbCr、HSV 色彩空间；LBP、HOG、Gabor 纹理特征以及专为行人再标识设计的手工剪裁特征。

其次，本章根据卷积神经网络的自我学习特性，提出了将手工剪裁特征与 CNN 特征相融合的 Feature Fusion Net（FFN）网络。经过特殊的神经网络训练方法训练后，由该网络提取出的行人再标识图像特征，在 $L1-norm$ 和 LFDA 两种分类器下的能得到非常高的匹配准确率。

最后，本章比较了不同图像特征的提取时间与输出维度。FFN 特征在提取时间与特征维度中取得了很好的平衡，进一步证明了其优秀的工程特性。

第 3 章　基于度量学习的行人图像匹配模型

建立完整的特征表达后，我们需要基于距离判断两幅行人图像的匹配程度。目前，度量学习（Metric Learning）是被广泛应用于人脸识别（Face Recognition）[38,39,40,41]与行人再标识（Person Re-identification）[3,4,5,12,17,23,28,33,34]等计算机视觉问题的图像匹配模型。一个好的图像特征表达，需要在合适的度量学习方法下匹配相应图像。本章首先介绍目前在行人再标识领域比较常用的度量学习方法，然后聚焦于各种提高度量学习在行人再标识问题中准确率的提升技巧（如核技巧、Mirror 方法）。最后，结合第 2 章中提出的基于卷积神经网络的行人图像特征提取方法，本章提出了一种增强的行人图像匹配模型。

3.1 行人再标识领域常用分类器

在行人再标识问题上，图像特征表达（Feature Representation）和分类器（Classifier）是算法模型中最关键的两个步骤。本节介绍几个常用的分类器，包括 L1-norm、L2-norm、LFDA、以及本模型中最后采用的高效度量学习方法：MFA。

3.1.1 直接距离比较方法：L1 与 L2 距离

L1 距离，记作 L1-norm，是指向量中各个元素绝对值之和。L2 距离，记作 L2-norm，指各元素的平方和。这两种直接距离比较方法均无“学习”的步骤，因此无需训练过程。也因此，这两种方法得到的匹配准确率很低。在实际应用中，L1-norm 与 L2-norm 距离常用于衡量输入特征的原始区分度。对于两个输入特征向量 x_1 与 x_2，L1-norm 与 L2-norm 距离的计算公式如下：

$$\begin{aligned} \rho_{L1-norm} &= \sum_{i=1}^{n} \left| x_1^i - x_2^i \right| \\ \rho_{L2-norm} &= \sum_{i=1}^{n} \left| x_1^i - x_2^i \right|^2 . \end{aligned} \tag{3.1}$$

3.1.2 度量学习方法：LFDA

度量学习（Distance Metric Learning）[47]是被广泛应用于机器学习领域的降维与分类方法，在人脸识别[38,39,40,41]与行人再标识[3,4,5,12,17,23,28,33,34]等领域有着广泛应用。

定义两个 d 维样本 x_i 与 x_j 之间的马氏距离为

$$dist_{mah}^2(x_i,x_j)=(x_i-x_j)^T M(x_i-x_j)=\left\|x_i-x_j\right\|_M^2. \tag{3.2}$$

其中 M 称为度量矩阵。则度量学习则是对 M 进行学习。

3.1.2.1 线性判别分析（Linear Discriminant Analysis, LDA）

线性判别分析由 Fisher[48]提出，亦称为 Fisher Discriminant Analysis，是最经典的线性度量学习方法之一。其基本思想在于：设法将样本点投影至一个超平面上，并使同类样本点之间的距离尽可能近、异类样本点之间的距离尽可能远。图 3-1[49]为 LDA 在二维样本点下降维与分类的示意图。

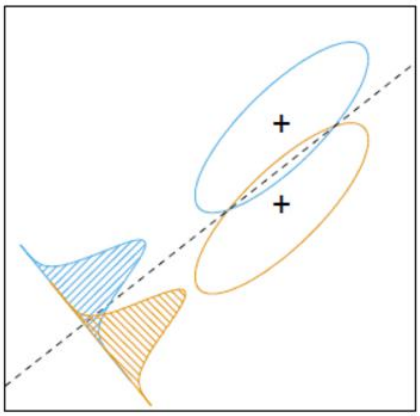

图 3-1 LDA 在二维样本点下的降维与分类示意图

记 u_i 为第 i 类的样本中心。我们使用类内散度矩阵（Within Class Scatter Matrix） S_W 与类间散度矩阵（Between Class Scatter Matrix） S_B 来刻画类内、类与类之间的样本耦合度：

$$\begin{aligned} S_W &= \sum_{i=1}^{c}\sum_{x_k\in class\ i}(u_i-x_k)(u_i-x_k)^T \\ S_B &= \sum_{i=1}^{c} n_i(u_i-u)(u_i-u)^T. \end{aligned} \tag{3.3}$$

定义广义瑞利商（Generalized Rayleigh Quotient）$J(W)$ 为类间距离和类内距

离的比值

$$J(W)=\frac{W^T S_B W}{W^T S_W W}, \tag{3.4}$$

则 LDA 算法的目标函数可写成

$$W_{opt}=\arg\max_W J(w)=\arg\max_W \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}. \tag{3.5}$$

在求出 W_{opt} 后，样本 x 超平面上的投影 x^* 可表示为

$$x^*=W_{opt}x. \tag{3.6}$$

3.1.2.2 局部线性判别分析（Local Fisher Discriminant Analysis, LFDA）

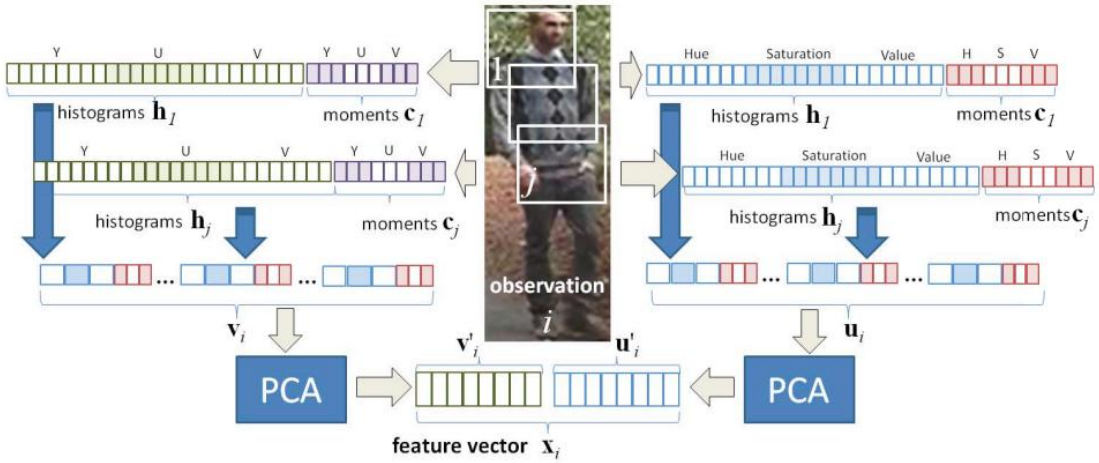

图 3-2 LFDA 方法[23]的原始图像特征提取步骤

LFDA[23]是 Pedagadi et al.在 2013 年提出的一种新的度量学习方法，它将 LPP 特征的特性与 Fisher 判别分析（Fisher Discriminant Analysis）相结合：类内散布矩阵 S_W 和类间散布矩阵 S_B 都根据关联矩阵 A（Affinity Matrix）加上了一个权重。亲密度矩阵 A 被应用于计算局部（Local）样本点的特性。它的应用，使得属于同一类但距离较远的类内样本对可以被忽略。

LFDA 是最有效的行人再标识度量学习方法之一。定义关联矩阵（Affinity Matrix） $A^w_{i,j}$ 与 $A^b_{i,j}$ 为

$$A_{i,j}^{w}=\begin{cases}A_{i,j}/n_c & if\ \ y_i=y_j=c\\ 0 & if\ \ y\neq y\end{cases}$$
$$A_{i,j}^{b}=\begin{cases}A_{i,j}/(\frac{1}{n}-\frac{1}{n_c}) & if\ \ y_i=y_j=c\\ 0 & if\ \ y\neq y\end{cases}. \tag{3.7}$$

则经改进的类内散度矩阵 S_W 与类间散度矩阵 S_B 可表示为

$$S_W=\frac{1}{2}\sum_{i=1}^{N}A_{i,j}^{b}(x_i-x_j)(x_i-x_j)'$$
$$S_B=\frac{1}{2}\sum_{i,j=1}^{N}A_{i,j}^{w}(x_i-x_j)(x_i-x_j)'. \tag{3.8}$$

类比于 LDA 方法，LFDA 的优化目标函数可以写为

$$W_{opt}^{lfda}=\arg\max_{W} r(\frac{W'S_WW}{W'S_BW}). \tag{3.9}$$

在求出 W_{opt} 后，样本 x 超平面上的投影 x^* 可表示为

$$x_i^*=W_{opt}^{lfda}x_i. \tag{3.10}$$

最后，使用欧氏距离计算公式来计算两个样本之间的距离：

$$D(i,j)=|z_i-z_j|. \tag{3.11}$$

对一个测试数据集中的一张 Probe 图像（记为 i ）和 Gallery 中的每一张图像（记为 j ），计算经 LFDA 投影矩阵 T'_{lfda} 投影后的两个向量之间的距离 $D(i,j)$。对 $D(i,:)$ 进行排序，取其中距离最小的前 k 个样本库（Gallery）图像，即为最有可能与测试图像（Probe）ID 相同的前 k 个图像。

3.1.5 边缘 Fisher 判别分析（Marginal Fisher Analysis, MFA）

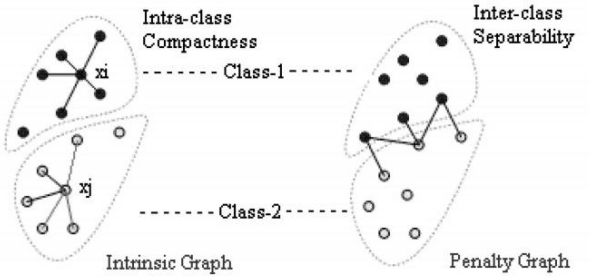

图 3-3 MFA 方法[42]的类内矩阵 W 与类间矩阵 W^P 示意图

作为对 LDA 部分不足的改进，Marginal Fisher Analysis（MFA）[42]是由 Yan et al.于 2007 年提出的有效的特征降维方法。MFA 不仅利用了 LDA 中的类内信息和类间信息，更加入了 Margin 的思想，对每个样本的邻域中的样本点信息进行利用。其基本步骤如下：

1) 进行 PCA 降维投影：将样本投影至 PCA 子空间，维数降至 $N^* = N - N_c$。记 PCA 转换矩阵为 W_{PCA}。

2) 构造类内矩阵和类间矩阵：类内的区分度由一个散度矩阵 $\tilde{S}_c$ 表示：

$$\tilde{S}_c = 2w^T X(D-W)X^T w \tag{3.12}$$

类间的区分度由一个惩罚矩阵 $\tilde{S}_p$ 表示：

$$\tilde{S}_p = 2w^T X(D^P - W^P)X^T w \tag{3.13}$$

其中，

$$\begin{aligned} W_{ij} &= \begin{cases} 1, \ if \ i \in N_{k_1}^+(j) \ or \ j \in N_{k_1}^+(i) \\ 0, \ else. \end{cases} \\ W_{ij}^p &= \begin{cases} 1, \ if \ (i,j) \in P_{k_2}(c_i) \ or (i,j) \in P_{k_2}(c_j) \\ 0, \ else. \end{cases} \end{aligned} \tag{3.14}$$

及

$$D_{ii} = \sum_{j \neq i} W_{ij}, \quad \forall i. \tag{3.15}$$

注：$N_{k_1}^+(i)$ 代表与样本 x_i 最接近的 k_1 个同类样本的集合；$P_{k_2}(c)$ 代表与样本 x_i 最接近的 k_2 个不同类样本的集合。

3) 使用 Marginal Fisher 准则：

$$w^* = \arg\min_w \frac{w^T X(D-W)X^T w}{w^T X(D^p - W^p)X^T w} \tag{3.16}$$

求得 MFA 的投影矩阵 w^*。

4) 结合 PCA 转换矩阵 W_{PCA}，输出最后的投影矩阵

$$w = W_{PCA} w^* \tag{3.17}$$

3.1.6 卷积神经网络（Convolutional Neural Network）

卷积神经网络（Convolutional Neural Network）已经被广泛地使用于计算机视觉的各类问题中。然而，其中只有少数研究聚焦于解决行人再标识问题。

Li et al.首先提出了 Deep Filter Pairing Neural Network（FPNN）[16]，使用 Patch Matching 层和 Max-out Pooling 层来识别行人再标识图片中的姿势和视角变化。FPNN 也是第一个在行人再标识问题中应用深度学习的模型。Ahmed et al.使用特别设计的 Cross-input Neighbourhood Difference 层[1]来提升准确率。随后，Deep Metric Learning[26]使用对称的（Siamese）深度神经网络和 Cosine Loss Function 损失函数应对不同视角下图片的巨大差别。Hu et al.提出了 Deep Transfer Metric Learning（DTML）[10]，将跨视角的先验知识传递到测试数据集中。

聚焦于行人再标识问题的卷积神经网络虽然可以获得相当高的识别成功率，但是也不可避免地有以下深度学习方法共有的不足：

1) 行人图像成对比较，网络输出二值概率，匹配过程耗时长；
2) 需要相当多的数据进行训练，而目前行人再标识领域极少此类大数据库；
3) 训练时间长、参数调整复杂、极度依赖于网络初始化参数；
4) 相对于传统分类方法，卷积神经网络方法的数学理论解释较为缺乏。

基于以上不足，在本章提出的行人再标识方法中，卷积神经网络只作为特征提取器的一部分。我们选择传统的度量学习方法，以进一步提高匹配准确率。

3.2 核技巧在距离学习中的应用

对解线性分类问题，线性的分类器是一种非常有效的方法。然而，行人再标识问题属于典型的非线性（Non-linear）问题，且行人图像中的背景环境、光照、角度等干扰急剧增加了非线性程度。本节叙述核技巧在非线性度量学习方法，并结合 3.1.5 节的 MFA 度量学习方法，提出其在核技巧下的运用。

用线性分类的方法求解非线性分类问题分为两步：首先使用一个变换将原空间（欧式空间，Euclidean Space）数据映射到新的空间；然后在新的空间（希尔伯特空间，Hilbert Space）中用线性分类学习方法从训练数据集中学习分类模型。核技巧（Kernel Trick）就属于这样的方法。

$$K(x,z)=\phi(x)\bullet\phi(z) \tag{3.18}$$

核技巧的中心思想是，在学习与预测中只定义核函数 $K(x,z)$，而不显式地定义映射函数 $\phi(\cdot)$。通常，直接计算 $K(x,z)$ 比较容易，而通过计算 $\phi(x)$ 和 $\phi(z)$ 来得到 $\phi(x)\bullet\phi(z)$ 并不容易。

以下为在支持向量机（Support Vector Machine，SVM）中常用的核函数（Kernel Function）：

1) 卡方核函数（Chi-Square Kernel）

$$K(x,y)=1-\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{\frac{1}{2}(x_i+y_i)} \tag{3.19}$$

2) RBF 核函数（Gaussian Radial Basis Function Kernel）

$$K(x,y)=\exp(-\frac{\sum_{i=1}^{n}|x_i-y_i|^2}{2\sigma^2}) \tag{3.20}$$

3) 多项式核函数（Polynomial Kernel）

$$K(x,y)=(x\bullet y+1)^p \tag{3.21}$$

在度量学习方法中，核函数的选择比较依赖经验，而很难用完整的数学理论解释。一系列实验表明，在本章的最终模型中，卡方核函数下的行人图像分类效果最好。

3.3 用于增强行人图像特征表达能力的 Mirror 方法

Mirror 方法是 Chen et al 在[3]中提出的行人再标识特征增强方法。通过研究不同视角下的行人图像特征的失真问题，这篇文章提出了适用于大多数度量学习方法的 Mirror 特征表示。对于适合使用核技巧（Kernel Trick）的度量方法，这篇文章将 Mirror 方法拓展至了核空间（Kernel Space）。实验证明，Mirror 方法能够在我们最终模型的三个测试数据库上，分别提升准确率 1%-2%。

3.3.1 特征层次的 Mirror 方法

因此，R 和 M 被设计成如下模式：

$$R=\frac{1+r}{z}I,$$
$$M=\frac{1-r}{z}I.$$

其中，z 是正则项。[3]提到，按照经验来说，当 z 取 L_2 正则项，即 $z=\sqrt{(1-r)^2+(1+r)^2}$ 时，模型的表现最好。

$$f_a(X^a)=U^TX^a_{aug}=(R^TU_1^T+M^TU_2^T)X^a,$$

$$f_b(X^b)=U^TX^b_{aug}=(R^TU_1^T+M^TU_2^T)X^b.$$

3.3.2 映射层次的 Mirror 方法

$\Lambda^{-\frac{1}{2}}P^TX_{aug}$ 叫做特征的镜像表示。在 Mirror 表示中，原特征和它的镜像被同时转换至一个更鲁棒的空间，所以能在行人再标识问题中取得更好的表示效果。

若我们把增强的特征 X_{aug} 转换为 $\Lambda^{-\frac{1}{2}}P^TX_{aug}$，则 H 可以用标准的度量学习方法表示。

3.3.3 核技巧（Kernel Trick）下的 Mirror 方法

上述的 Mirror 方法提供了一个在线性特征下的映射，然而，通常的特征数据分布以及失真程度往往是非线性的。为了解决这个问题，[3]中借用机器学习领域的核技巧（Kernel Trick），推导出核技巧下的 Mirror 方法。

因此，在加入核技巧后，映射函数可以定义如下：

$$f_a(x^a)=\alpha^Tk(X[R,M],x^a)$$
$$f_b(x^b)=\alpha^Tk(X[R,M],x^b)$$

其中，R 和 M 是在 3.3.1 节中定义的 $n\times n$ 矩阵，$k(\cdot,\cdot)$ 代表核函数。

$$\alpha^*=\arg\min_{\alpha}\frac{\alpha^TK(D-W)K\alpha}{\alpha^TK(D^p-W^p)K\alpha}$$

3.4 基于行人再标识分类器的实验

3.4.1 测试数据集与规则设置

本章测试数据集与规则设置与第 2 章实验相同。

记 FFN 特征为$[FFN]$，LOMO 特征为$[LOMO]$，则在总模型中，$FFN+LOMO$特征的归一化串联可以表示为

$$[FFN, \frac{LOMO}{LOMO_{\max}} \times FFN_{\max}].$$

归一化可以避免某种单一特征的过度突出，使两种特征的融合更加平顺。

LDFV 特征在使用卡方核（Chi-Square Kernel）时的准确率极低，因此在 Mirror KMFA 分类器实验中，我们使用了线性核（Linear Kernel）。

3.4.2 实验结果与分析

3.4.2.1 Mirror KMFA 分类器下各特征的表现

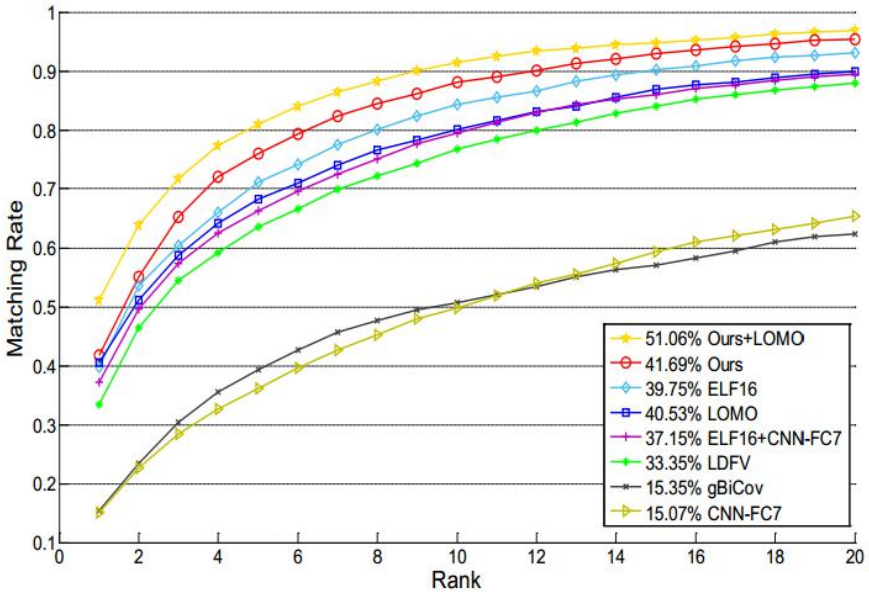

图 3-4 VIPeR 数据集、Mirror KMFA 分类器下各种特征的 CMC 曲线

表 3-1 各种特征在 VIPeR 数据集、Mirror KMFA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
FFN 特征+LOMO 特征	**51.06**	**81.01**	**91.39**	**96.90**
FFN 特征	**41.69**	**75.91**	**88.05**	**95.33**
ELF16+CNN-FC7	37.15	66.30	79.40	89.53
LDFV	33.35	63.53	76.78	87.97
ELF16	39.75	71.12	84.26	93.12
gBiCov	15.35	39.32	50.67	62.30
LOMO	40.53	68.24	80.14	89.95
CNN-FC7	15.07	36.12	49.76	65.43

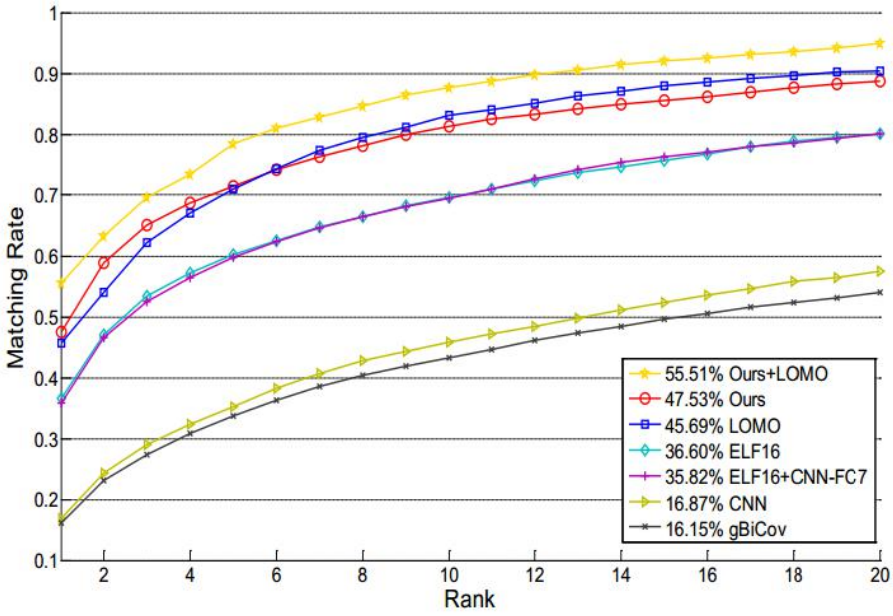

图 3-5 CUHK01 数据集、Mirror KMFA 分类器下各种特征的 CMC 曲线

表 3-2 各种特征在 CUHK01 数据集、Mirror KMFA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
FFN 特征+LOMO 特征	55.51	78.40	87.62	95.00
FFN 特征	47.53	71.50	81.26	88.77
LOMO	45.69	70.91	83.05	90.33
ELF16+CNN-FC7	35.82	59.73	69.65	80.25
gBiCov	16.15	33.68	43.25	53.99
CNN-FC7	16.87	35.29	45.83	57.48
ELF16	36.60	60.24	69.49	80.00

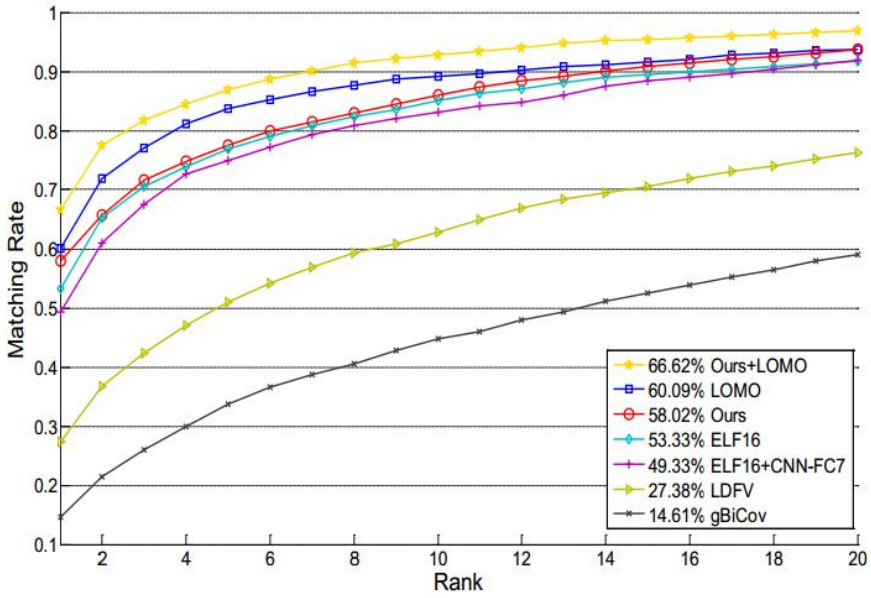

图 3-6 PRID450s 数据集、Mirror KMFA 分类器下各种特征的 CMC 曲线

表 3-3 各种特征在 PRID450s 数据集、Mirror KMFA 模型下的性能比较

Rank-i	i=1	i=5	i=10	i=20
FFN 特征+LOMO 特征	**66.62**	**86.84**	**92.84**	**96.89**
FFN 特征	**58.02**	**77.56**	**85.96**	**93.64**
LOMO	60.09	83.73	89.11	93.64
ELF16+CNN-FC7	49.33	74.98	83.11	91.91
ELF16	53.33	76.89	85.11	91.91
CNN-FC7	25.11	51.42	60.22	68.18
gBiCov	14.61	33.67	44.72	58.94

3.4.2.2 与 State-of-the-Art 方法的比较

这个实验对比了本章提出的模型与当前准确率最高（State-of-the-Art）的行人再标识模型之间的准确率。本章提出的模型基于 Mirror KMFA 分类器，使用的特征为第二章中提出的 FFN 特征+LOMO 特征的归一化串联。

表 3-4 至表 3-6 总结了在 VIPeR、CUHK01、PRID450s 三个库上当前准确率最高的一些行人再标识模型，包括 LOMO＋XQDA[18]、Mirror KMFA[3]、Ahmed's Improved Deep ReID[1]、Mid-level Filter[30]等。

注：表 3-4 至表 3-6 按年代顺序排列，红色标注为本章提出的模型，蓝色标注为应用深度学习方法的模型。

表 3-4 各种模型在 VIPeR 数据集下的 Rank-i 准确率

Rank-i	i=1	i=5	i=10	i=20
我们的模型	**51.06**	**81.01**	**91.39**	**96.90**
Deep Feature Learning	40.50	60.80	70.40	84.40
LOMO+XQDA	40.00	67.40	80.51	91.08
Mirror KMFA(R_{χ^2})	42.97	75.82	87.28	94.84
mFilter+LADF	43.39	73.04	84.87	93.70
mFilter	29.11	52.10	67.20	80.14
SalMatch	30.16	52.31	65.54	79.15
LFDA	24.18	52.85	67.12	78.96
LADF	29.34	61.04	75.98	88.10
RDC	15.66	38.42	53.86	70.09
KISSME	24.75	53.48	67.44	80.92
LMNN-R	19.28	48.71	65.49	78.34
PCCA	19.28	48.89	64.91	80.28
L2-norm	10.89	22.37	32.34	45.19
L1-norm	12.15	26.01	32.09	34.72

表 3-5 各种模型在 CUHK01 数据集下的 Rank-i 准确率

Rank-i	i=1	i=5	i=10	i=20
我们的模型	**55.51**	**78.40**	**83.68**	**92.59**
Mirror KMFA(R_{χ^2})	40.40	64.53	75.34	84.08
Ahmed's Deep Re-id	47.53	72.10	80.53	88.49
mFilter	34.30	55.12	64.91	74.53
DeepReID	27.87	64.01	82.50	87.36
ITML	15.98	35.22	45.60	59.81
eSDC	19.67	32.72	40.29	50.58
LFDA	22.08	41.56	53.85	64.51
KISSME	14.02	32.20	44.44	56.61
LMNN-R	13.45	31.33	42.25	54.11
L2-norm	5.63	16.00	22.89	30.63
L1-norm	10.80	15.51	37.57	35.57

表 3-6 各种模型在 PRID450s 数据集下的 Rank-i 准确率

Rank-i	i=1	i=5	i=10	i=20
我们的模型	**66.62**	**86.84**	**92.84**	**96.89**
Mirror KMFA(R_{χ^2})	55.42	79.29	87.82	93.87
Ahmed's Deep Re-id	34.81	63.72	76.24	81.90
ITML	24.27	47.82	58.67	70.89
LFDA	36.18	61.33	72.40	82.67
KISSME	36.31	65.11	75.42	83.69
LMNN-R	28.98	55.29	67.64	78.36
L2-norm	11.33	24.50	33.22	43.89
L1-norm	25.50	25.33	51.73	53.07

以 Rank-1 准确率计，本文提出的行人再标识模型可以比其他模型的最高准确率高 10%左右。在 VIPeR、CUHK01、PRID450s 这三个库中，我们的模型分别超过当前准确率最高的模型 8.09%，7.98%和 11.20%。

本实验也对比了三个使用卷积神经网络方法的行人再标识模型（DeepReID[16]、Ahmed's Deep Re-id[1]、Deep Feature Learning[6]）。它们在表 3-4 至表 3-6 中用蓝色粗体字着重标出。这些深度学习模型均使用修改过的卷积神经网络结构来做每对行人图像之间（Pairwise）的对比，并使用一些新创造的层来满足行人再标识模型中的独特需求。而本模型把卷积神经网络当作特征提取

器，然后把提取出的特征放入不同度量学习方法中计算每对图片之间的相对距离。这种模型结构不仅提高了准确率，而且也允许我们使用更大的训练数据集来训练 CNN 模型。

3.5 本章小结

本章重点介绍了行人再标识领域通用的几种常用分类器：$L1-norm$、LFDA、MFA，并根据行人图像的实际特点，结合[3]的研究工作，提出了 Mirror KMFA 分类器。

我们在三个通用的行人图像数据库上进行了实验，并与传统的度量学习行人再标识模型和当前已公开发表的几个基于卷积神经网络的行人再标识模型进行了比较。通过实验，得到以下结论：

1) FFN 特征能在 Mirror KMFA 分类器下达到相当好的分类准确率。
2) FFN+LOMO 特征在 Mirror KMFA 分类器下，$Rank-1$准确率超过当前最好的行人再标识模型 8.09%、7.98%、11.20%。
3) 对比目前已发表的基于卷积神经网络的行人再标识模型，我们的模型更有效。这证明了神经网络特征提取方法与传统度量学习分方法结合的可行性。

第 4 章　基于不完整图像的行人再标识研究

传统行人再表示问题的研究日趋深入，同时也衍生出了一些开放性的行人再标识问题[43, 44, 45]。本章研究了一个全新的行人再标识问题：基于不完整图像的行人再标识问题，探究了基于稀疏编码与字典学习的传统解决方法，并使用卷积神经网络，提出了一个可行的图像匹配模型。

4.1 行人再标识问题在实际应用中的复杂性

传统行人再标识问题解决了理想环境下，双摄像头图像之间的匹配。而在实际应用中，行人再标识问题远比实验环境复杂。光照、遮挡[43]、分辨率[44]等问题大大制约了传统行人再标识算法在实际中的应用。另外，基于监控摄像头视频的行人再标识问题于[45]中也得到了研究。

图 4-1 低分辨率下的行人再标识问题[44]

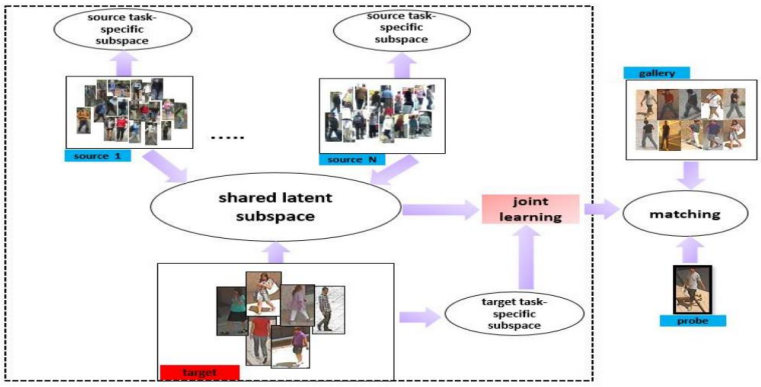

图 4-2 跨数据集的行人再标识迁移学习问题[25]

图 4-3 基于视频的行人再标识问题[45]

基于不完整图像的行人再标识问题（Partial Person Re-identification）是由 Zheng et al.于[43]提出的一个新问题。与传统行人再标识问题不同的是，新问题中只使用一张不完整的测试图片（Probe）匹配样本库（Gallery）中的完整行人图片。由于测试图片的不完整，图片匹配的难度大大增加。传统行人再标识方法均不能在测试数据集上取得很好的效果。图 4-4 是一个基于不完整图像的行人再标识问题示例。

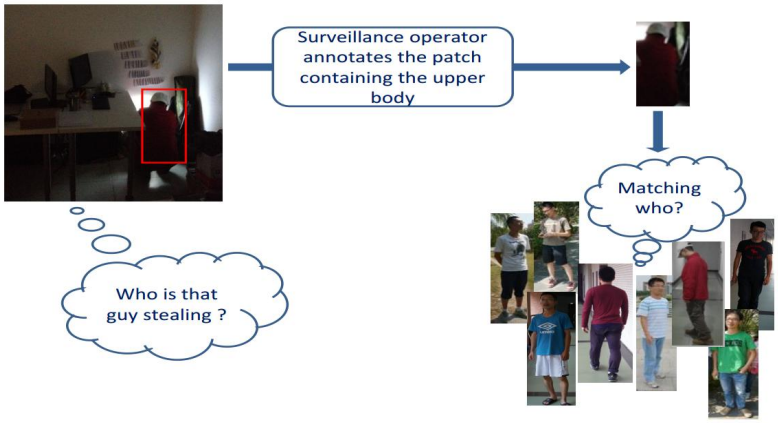

图 4-4 基于不完整图像的行人再标识问题[43]

4.2 基于不完整图像的行人再标识模型

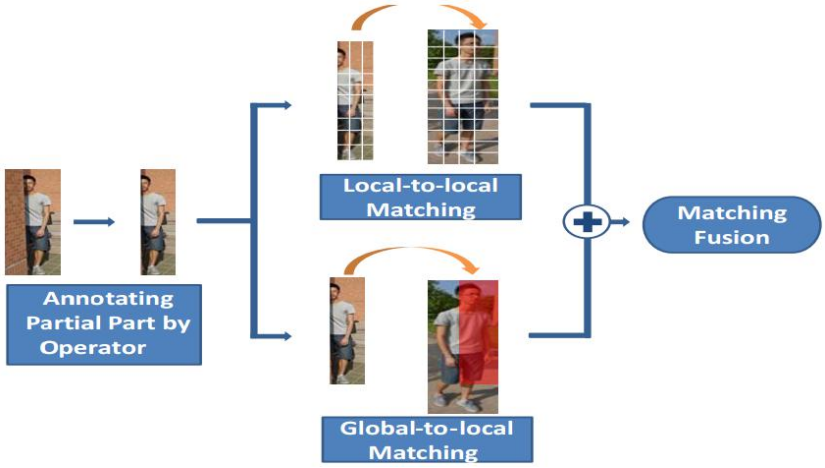

图 4-5 AMC-SWM 模型结构示意图[43]

[43]中提出的基于不完整图像的行人再标识模型由以下两个子模型组成：局部匹配局部模型（Local-to-local Model）与局部匹配整体模型（Global-to-local Model），结构如图 4-5。在局部匹配局部的过程中，将不完整图像分解为小的图像分块（Patches），对图像分块进行匹配。同时，在局部匹配整体模型中，使用滑动窗口搜索（Sliding Window Search）方法，将不完整图片与样本库中的图片进行整体匹配。在[43]中，局部匹配局部模型被称为 Ambiguity-Sensitive Matching（AMC）；局部匹配整体模型被称为 Sliding Window Matching（SWM）。下面分别介绍这两个子模型。

4.2.1 局部匹配局部模型 AMC

局部匹配局部模型基于图像分块成对比较的稀疏编码（Sparse Coding）方法。首先，将样本库中的一幅图分为 k_c 个图像分块（Patches），将这些图像分块分别提取特征向量，得到第 c 个人的字典（Dictionary）

$$D_c=[d_{c_1},d_{c_2},\ldots,d_{c_{k_c}}] \tag{4.1}$$

则样本库的总体字典（Gallery Dictionary）可以表示为：

$$D=[D_1,D_2,\ldots,D_C] \tag{4.2}$$

总体字典中共有 $K=\sum_{c=1}^{C}k_c$ 个图像分块的特征向量。

将 AMC 方法中的优化目标项写成

$$\min_{x_i}\|y_i - Dx_i\|_2^2 + \alpha pa_i^T x_i + \beta\|x_i\|_1,\quad i=1,\ldots,n \tag{4.3}$$

其中，n 为单张测试图片图像分块数；$pa_i=[pa_{i_1}, pa_{i_2},\ldots pa_{i_k}]^T$，$i=1,\ldots,n$ 是一个控制图像分块之间空间紧密性的 K 维向量；$\|x_i\|_1$ 是稀疏编码 x_i 的稀疏度控制项；α 与 β 分别为控制两个约束项影响力大小的参数。

使用 Feature-Sign Search 算法[46]解公式 4.3，得到 x_i 的最优解。将此最优解作为第 i 张测试图片的稀疏编码表示。用公式 4.4 预测此测试图片 Y 的类标 $\hat{c}$：

$$\hat{c}=\arg\min_{x_i} r_c(Y)=\frac{1}{n}\sum_{i=1}^{n}\|y_i - D_c\delta_c(x_i)\|_2^2 \tag{4.4}$$

4.2.2 局部匹配整体模型 SWM

AMC 模型不能很好地捕捉图像分块之间的空间关系，因此，[43]使用滑动窗口搜索方法来构建局部图像匹配全身图像的模型。在样本库图像中，对于一个与局部图像大小相同的滑动窗口，每一个窗口可以提取出一个图像特征向量。计算测试图片与窗口的最小距离 l_c。因此，对于 C 个类的最小距离向量可记作

$$L_{dist}=[l_1,l_2,\ldots,l_C]^T \tag{4.5}$$

一张测试图片的 SWM 模型预测类标 $\hat{c}$ 即为：

$$\hat{c}=\arg\min_{c} l_c,\quad c=1,2,\ldots,C \tag{4.6}$$

4.2.3 模型融合

对于单张给定的测试图片，使用公式 4.4 得到 C 个类分别的距离 $R_{dist}=[r_1,r_2,\ldots,r_C]^T$。与公式 4.5 进行融合，算出 AMC-SWM 模型下到每一个类别的总距离

$$S_{dist}=\gamma R_{dist}+(1-\gamma)L_{dist}. \tag{4.7}$$

则一张测试图片的 AMC-SWM 模型预测类标 $\hat{c}$ 即为：

$$\hat{c} = \arg\min_{c} S_c, \quad c = 1,2,\ldots,C. \tag{4.8}$$

4.3 基于卷积神经网络的模型 PartialNet

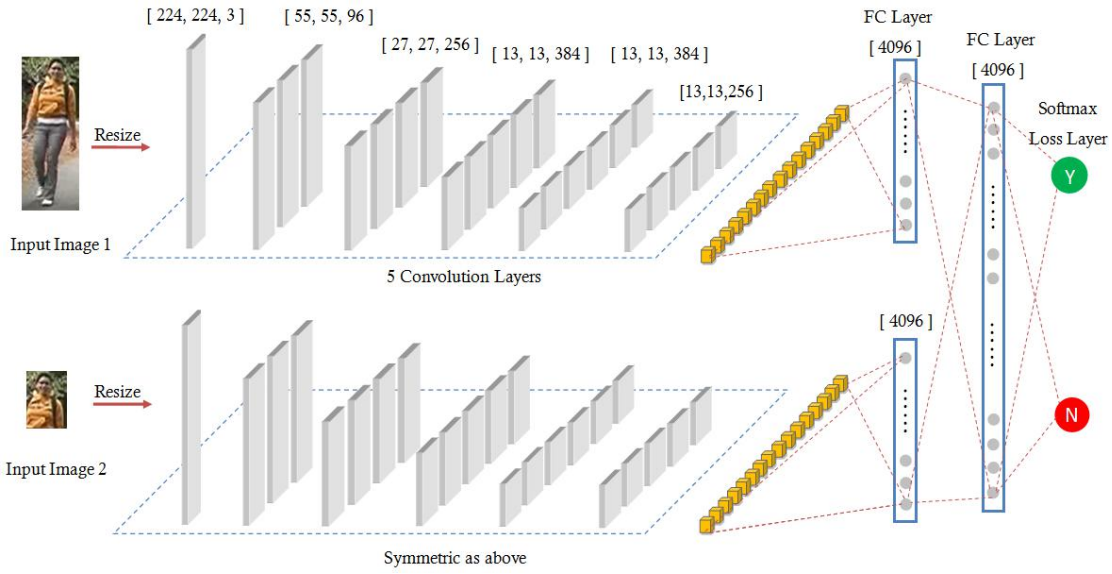

图 4-6 PartialNet 的网络结构示意图

我们希望把完整行人再标识图片（如图 4-6 中的 Full Image）和不完整图片（如图 4-6 中的 Partial Image）同时输入神经网络，而在神经网络的最后一层（SoftmaxLoss 层）输出这两张图片为同一个 ID 的概率。因此，本章提出了 PartialNet 卷积神经网络结构。图 4-6 为该网络的示意图，表 4-1 为该网络结构的参数表。

PartialNet 单纯地使用了 CNN 图像特征（如图中的黄色小方块），并采用了上下对称、参数共享的结构来增强神经网络参数学习的效率。PartialNet 通过反向传播（Back Propagation）来进行参数学习。总的来说，经过大量的训练图片训练，PartialNet 应该能够优化图像卷积滤波器的参数学习过程，使得最后输出的 0-1 判断更加准确。

表 4-1 PartialNet 网络结构参数表

层	类型	输入层	输出层	输出维度	参数个数
Data	图像输入层	-	Data	227×227×3	-
Conv1	卷积层	Data	Conv1	55×55×96	11×11×3×96

ReLU1	激活层	-	-	-	-
Pool1	池化层	Conv1	Pool1	27×27×96	-
Norm1	正则层	Pool1	Norm1	-	-
Conv2	卷积层	Pool1	Conv2	27×27×256	5×5×48×256
ReLU2	激活层	-	-	-	-
Pool2	池化层	Conv2	Pool2	13×13×256	-
Norm2	正则层	Pool2	Norm2	-	-
Conv3	卷积层	Norm2	Conv3	13×13×384	3×3×256×384
ReLU3	激活层	-	-	-	-
Conv4	卷积层	Conv3	Conv4	13×13×384	3×3×384×384
ReLU4	激活层	-	-	-	-
Conv5	卷积层	Conv4	Conv5	13×13×256	3×3×384×256
ReLU5	激活层	-	-	-	-
Pool5	池化层	Conv5	Pool5	6×6×256	-
FC6	全连接层	Pool5	FC6	4096	6×6×256×4096
ReLU6	激活层	-	-	-	-
Drop6	激活层	-	-	-	-
Data-p	图像输入层	-	Data-p	227×227×3	-
Conv1-p	卷积层	Data-p	Conv1-p	55×55×96	11×11×3×96
ReLU1	激活层	-	-	-	-
Pool1-p	池化层	Conv1-p	Pool1-p	27×27×96	-
Norm1-p	正则层	Pool1-p	Norm1-p	-	-
Conv2-p	卷积层	Pool1-p	Conv2-p	27×27×256	5×5×48×256
ReLU2	激活层	-	-	-	-
Pool2-p	池化层	Conv2-p	Pool2-p	13×13×256	-
Norm2-p	正则层	Pool2-p	Norm2-p	-	-
Conv3-p	卷积层	Norm2-p	Conv3-p	13×13×384	3×3×256×384
ReLU3	激活层	-	-	-	-
Conv4-p	卷积层	Conv3-p	Conv4-p	13×13×384	3×3×384×384
ReLU4	激活层	-	-	-	-
Conv5-p	卷积层	Conv4-p	Conv5-p	13×13×	3×3×384×

				256	256
ReLU5	激活层	-	-	-	-
Pool5-p	池化层	Conv5-p	Pool5-p	6×6×256	-
FC6-p	全连接层	Pool5-p	FC6-p	4096	6×6×256×4096
ReLU6	激活层	-	-	-	-
Drop6	激活层	-	-	-	-
Concat	辅助层	FC6、FC6-p	Concat	8192	-
FC7	全连接层	Concat	FC7	4096	8192×4096
ReLU7	激活层	-	-	-	-
Drop7	激活层	-	-	-	-
FC8	全连接层	FC7	FC8	2	4096×2
SoftmaxLoss	损失函数层	FC8	SoftmaxLoss	-	-

4.4 基于不完整图像的行人再标识模型实验

4.4.1 神经网络训练设置

4.4.4.1 训练数据集

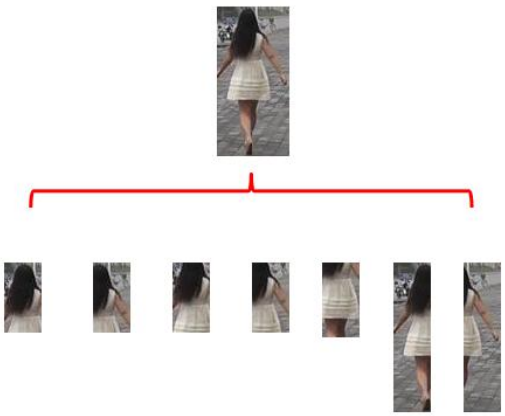

图 4-6 Partial REID 数据集部分训练样本

目前，学术界尚无适合于训练神经网络的不完整行人再标识数据库。因此，我们从 Market-1501 数据库中人工截取相应数据，手工构造一个不完整行人再标识数据库。Market-1501 是清华大学于 2015 年公开的行人再标识数据集。它由 1501 个 ID 下的 38195 幅行人图像组成，是一个多摄像头混合的数据集。相比于传统行人再标识数据集（VIPeR、CUHK 系列、ETHZ、iLIDS 等），它的优势是图像数量多，适宜用于深度神经网络的训练。

对于 Market-1501 数据库中的每一张图片，我们分别截取其不同部位的 7 个图像分块（如图 4-6），并且利用这些图像分块组成不完整图像的图像数据库。我们使用新构造的数据集对 PartialNet 进行训练（至 Loss 值稳定在 0.001 左右），然后使用这个神经网络来对测试数据集（Partial REID、P-CAVIAR、P-iLIDS）进行测试。

表 4-2 训练数据集基本信息

数据集名称	Market-1501
完整图像数量	25259
不完整图像数量	25259×7 = 176813
总 ID 数量	1501
摄像头数目	多摄像头混合，不分视角
每个 ID 总的图像数目	7

4.4.2 测试数据集与规则设置

本章测试数据集与规则设置与第二章实验相同。表 4-1 汇总了三个不完整行人再标识测试数据集（Partial REID、P-CAVIAR、P-iLIDS）的基本信息。图 4-6 为这三个测试数据集的部分样本。在图片展示中，上下两行分别代表测试图像（Probe）和样本库（Gallery Image），并按分辨率大小比例显示。

表 4-3 测试数据集基本信息

数据集名称	Partial REID	P-CAVIAR	P-iLIDS
完整图像数量	300	720	235
不完整图像数量	300	720	107
总 ID 数量	60	72	107
测试集 ID 数量	42	50	75
摄像头数目	多摄像头	多摄像头	多摄像头
每个 ID 的不完整图像数目	5	10	1

图 4-7 Partial REID、P-CAVIAR 与 P-iLIDS 数据集部分样本

由于时间与篇幅问题，本章只对 Partial REID 数据集进行测试。

4.4.2.1 Single-shot 与 Multi-shot

Single-shot 测试意味着在测试样本库（Gallery）中，每一个 ID 只有一张图片（记作 N=1）。Multi-shot 测试意味着，在测试样本库（Gallery）中，每一个 ID 有多张图片（N>1）。经验表明，大多数行人再标识模型在 Single-shot 规则与 Multi-shot 规则下的准确率表现不甚相同。

本章的三个测试数据集均可以进行 Multi-shot 测试，不过由于时间和篇幅问题，我们只进行 Single-shot 测试。

4.4.2 实验结果与分析

图 4-8 高斯模糊（r=100）后的部分 Partial REID 数据集样本

针对 PartialNet 网络的两个不同输入，我们设计了五种不同的图像配对方法。

对于每对测试图片，取 PartialNet 输出为 0 的概率作为两张图片之间的距离，并以此统计 Rank-i 值。

在不完整图片 VS 不完整图片这一设定中，我们分别对每对测试图片的 7 个图像分块取距离均值与最大值进行匹配。

在不完整图片 VS 完整图片这一设定中，我们尝试用高斯模糊来减少图片边缘背景部分对匹配造成的影响。我们分别取 $r=30$ 与 $r=100$ 进行实验。高斯模糊的公式如下：

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{4.9}$$

从表 4-4 可以看出，直接使用不完整图片 VS 完整图片的设定，在 PartialNet 神经网络结构中的匹配准确率表现最高。

表 4-4 几种模型在 Partial REID 数据集、 PartialNet 模型下的性能比较

Rank-i（Single-shot）	i=1	i=5	i=10	i=20
不完整图片 VS 完整图片	**17.90**	**47.57**	**65.04**	**83.47**
不完整图片 VS 不完整图片（取不完整图概率均值）	12.09	39.04	57.04	82.14
不完整图片 VS 不完整图片（取不完整图概率最大值）	9.47	29.42	46.67	71.28
不完整图片 VS 完整图片（完整图高斯模糊，r=30）	9.81	31.23	46.95	69.09
不完整图片 VS 完整图片（完整图高斯模糊，r=100）	11.76	37.23	53.95	77.57

4.5 本章小结

本章首先介绍了行人再标识领域的几个新问题：基于不完整图片的行人再标识、基于视频的行人再标识以及迁移学习行人再标识。其次，本章分析了基于不完整图片的行人再标识问题中，[43]提出的 AMC-SWM 模型。

在本章中，我们提出一种新颖的、基于不完整图像的卷积神经网络行人再标识模型。此神经网络同时输入完整行人图片和不完整行人图片，并判断是否为同一 ID，输出 0-1 判断的概率。实验结果表明，本章所提出的匹配模型具有比较直观的网络结构和比较理想的匹配结果。

我们认为，此模型有很大的提升空间，但由于时间限制，未能在毕业论文完稿之前将此模型完善至最佳状态。在接下来的时间里，我将继续完善此模型，争取于毕业前结束此工作，并投稿至 ACCV2016 会议。

第 5 章 总结与展望

5.1 工作总结

深度学习是近五年来在机器学习领域最火的新技术。由于卷积神经网络方法在各类计算机视觉问题上的高准确率表现，传统计算机视觉方法正在迅速地被深度学习方法所取代。

本文研究工作重点在于将卷积神经网络方法应用于行人再标识问题，并在相关实验中取得相当好的结果。对本文工作进行总体概括，主要有以下三方面的贡献：

1) 以提取更有区分度的行人图像特征为目标，提出了 Feature Fusion Net 卷积神经网络架构，并从理论和实验两方面证明了 FFN 特征的有效性。在 L1-norm 和 LFDA 两个常用行人再标识算法的实验中，FFN 特征均取得比传统特征更好的效果。
2) 以构建更有区分度的行人再标识分类器为目标，结合[3]的工作，提出了 Mirror KMFA 分类器。结合 FFN+LOMO 特征，我们构建了完整的行人再标识模型。相对于目前学术界已发表的行人再标识模型，在三个通用行人图像数据集上，本文模型的 Rank-1 准确率分别有 8.09%、7.98%及 11.20%的提升。
3) 研究了基于不完整行人图像的行人再标识问题，尝试了采用稀疏矩阵与字典学习的传统分类方法。并引入卷积神经网络算法，提出一种有效的基于不完整行人图像的行人再标识模型。

本文的前两项贡献已作为会议论文的形式发表于 IEEE WACV2016，本人为第一作者。

5.2 研究展望

本文的工作虽然在方法上取得一定突破，但仍存在许多不足。在将来的深度行人再标识研究工作中，我们将围绕以下几个方面，进行深入研究：

1) 行人特征表达的完善。行人特征表达的优劣直接关系到分类识别的效果。在接下来的工作中，我将会认真研究 LOMO 行人图像特征，并将其完整融入至 FFN 卷积神经网络架构中。
2) 不完整图像下的行人再标识问题。目前，不完整图像下的行人再标识问题在学术界中的研究工作相对较少。但这个具体问题对于实际生产环境中的监控摄像头行人再识别有相当重要的意义。我希望用接下来的两个月时间进行此研究，并于毕业之前得到令人满意的研究结果。

总结来说，我希望在毕业之前，能在基于卷积神经网络的行人再标识研究中，再取得一定突破。希望于毕业前，新的研究成果能投稿至 IEEE ACCV2016 会议。

参考文献

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In IEEE CVPR, 2015.

[2] L. Bottou. Stochastic gradient descent tricks. In Neural Networks: Tricks of the Trade. 2012.

[3] Y.-C. Chen, W.-S. Zheng, and J. Lai. Mirror representation for modeling view-specific transform in person reidentification. In IJCAI, 2015.

[4] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In ICML, 2007.

[5] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In ACCV. 2011.

[6] S. Ding, L. Lin, G. Wang, and H. Chao. Deep feature learning with relative distance comparison for person reidentification. Pattern Recognition, 2015.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In IEEE CVPR, 2010.

[8] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In IEEE PETS Workshop, 2007.

[9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In ECCV. 2008.

[10] J. Hu, J. Lu, and Y.-P. Tan. Deep transfer metric learning. In IEEE CVPR, 2015.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv:1408.5093, 2014.

[12] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In IEEE CVPR, 2012.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[14] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. IEEE TPAMI, 35(7):1622‐1634, 2013.

[15] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In ACCV, 2012.

[16] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In IEEE CVPR, 2014.

[17] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In IEEE CVPR, 2013.

[18] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In IEEE CVPR, 2015.

[19] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In ECCV, 2012.

[20] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. IVC, 32(6):379–390, 2014.

[21] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In IEEE CVPR, 2012.

[22] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI, 24(7):971–987, 2002.

[23] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In IEEE CVPR, 2013.

[24] P. M. Roth, M. Hirzer, M. Kostinger, C. Beleznai, and H. Bischof. Mahalanobis distance learning for person reidentification. In Person Re-Identification, PP247–267, 2014.

[25] X. Wang, W.-S. Zheng, X. Li, and J. Zhang. Cross-scenario transfer person re-identification. IEEE TCSVT, PP(99):1–1, 2015.

[26] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Deep metric learning for person re-identification. In IEEE ICPR, 2014.

[27] Y. Zhang and S. Li. Gabor-lbp based region covariance descriptor for person re-identification. In IEEE ICIG, 2011.

[28] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In IEEE ICCV, 2013.

[29] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In IEEE CVPR, 2013.

[30] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In IEEE CVPR, 2014.

[31] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, J. Bu, and Q. Tian. Scalable person re-identification: A benchmark. In IEEE ICCV, 2015.

[32] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In IEEE CVPR, 2011.

[33] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. IEEE TPAMI, 35(3):653–668, 2013.

[34] W.-S. Zheng, S. Gong, and T. Xiang. Towards open-world person re-identification by one-shot group-based verification. IEEE TPAMI, PP(99):1-1, 2015.

[35] G.D. Riccia, and A. Shapiro. Fisher discriminant analysis and factor analysis. IEEE TPAMI, PP(1):99-104, 1983.

[36] A. Franco and L. Oliveira. A Coarse-To-Fine Deep Learning for Person Re-Identification. In IEEE WACV, 2016.

[37] Y. LeCun, L. Bottou and Y. Bengio. Gradient-based learning applied to document recognition. IEEE TPAMI, PP86(11): 2278-2324, 1998.

[38] A. Pentland, B. Moghaddam and T. Starner. View-based and modular eigenspaces for face recognition. In IEEE CVPR, 1994.

[39] X. He, S. Yan and Y. Hu. Face recognition using Laplacianfaces. IEEE TPAMI, PP27(3): 328-340, 2005.

[40] K. Weinberger, J. Blitzer and L.-K., Saul. Distance metric learning for large margin nearest neighbor classification. In NIPS, 2005.

[41] M. Guillaumin, J. Verbeek and C. Schmid. Is that you? Metric learning approaches for face identification. In IEEE ICCV, 2009.

[42] S. Yan, D. Xu and B. Zhang. Graph embedding and extensions: a general framework for dimensionality reduction. IEEE TPAMI, PP29(1): 40-51, 2007.

[43] W.-S. Zheng and X. Li. Partial Person Re-identification. In IEEE CVPR, 2015.

[44] X. Li, W.-S., Zheng and X. Wang. Multi-scale Learning for Low-resolution Person Re-identification. In IEEE ICCV, 2015.

[45] J. You, A. Wu, X. Li and W.-S. Zheng. Top-push Video-based Person Re-identification. In IEEE CVPR, 2016.

[46] H. Lee, A. Battle and R. Raina. Efficient sparse coding algorithms. In NIPS, 2006.

[47] E. Xing, A Ng and M. Jordan. Distance metric learning with application to clustering with side-information. In NIPS, 2003.

[48] R. Fisher. The use of multiple measurements in taxonomic problems. Annals of eugenics, PP7(2): 179-188, 1936.

[49] Anzai Y. Pattern Recognition & Machine Learning. Elsevier, 2012.

[50] I. Aleksander I and H. Morton. An introduction to neural computing. London: Chapman and Hall, 1990.

[51] C. Poynton. Digital Video and HDTV. PP291-292, Morgan Kaufmann, 2003

[52] Agoston and K. Max. Computer Graphics and Geometric Modeling: Implementation and Algorithms. PP300-306, London: Springer, 2005.

[53] T. Ojala, M. Pietikäinen and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI, PP24(7): 971-987, 2004.

[54] N. Dalal and B. Triggs B. Histograms of oriented gradients for human detection. In IEEE CVPR, 2005.

[55] M. Guillaumin, J. Verbeek and C. Schmid. Is that you? Metric learning approaches for face identification. In IEEE ICCV, 2009.

[56] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In IEEE CVPR, 2012.

[57] J.-V. Davis, B. Kulis and P. Jain. Information-theoretic metric learning. In ACM ICML, 2007.

相关科研成果目录

[1] Shangxuan Wu, Ying-Cong Chen, Xiang Li, An-Cong Wu, Jin-Jie You, Wei-Shi Zheng, An Enhanced Deep Feature Representation for Person Re-identification, In IEEE WACV, 2016.

致 谢

感谢中山大学，特别是电子与信息工程学院的教授和老师们四年来对我谆谆教诲。木棉花开，时光匆匆，不变的是你们对学术的追求和对教书育人的坚持。

本论文的研究工作是在我的指导老师郑伟诗教授的严格要求下完成的。您手把手地指导我完成了从选题、查阅文献、研究算法到整个论文的撰写工作。在您的鼓励下，我已经将本文的前两项贡献以会议论文的形式正式发表于 IEEE WACV2016，并将在行人再标识领域继续研究下去。郑老师在候机厅读论文、在饭桌上推公式的场景历历在目，您严谨认真的学术风格以及实事求是、精益求精的研究态度鼓舞着我，是我在研究生阶段学习和未来工作中的榜样。

李翔、游瑾洁、陈颖聪、吴岸聪、黄东程等 iSEE 实验室的师兄师姐对我的研究工作给予了极大帮助。是你们将我领进了计算机视觉和机器学习的大门。你们永远是我的良师益友。

要最感谢我的父母和家人，感谢你们多年以来对我的鼓励，以及对我的梦想的大力支持。你们是我成长道路上最为珍视的人。

最后，对参加本论文评审和答辩工作的所有专家表示最诚挚的谢意！

吴尚轩

2016 年 4 月

附 录

附录 A 发表于 IEEE WACV2016 的文章原文