

密级： 保密期限：

北京邮电大学

硕士学位论文



题目：基于深度学习的步态识别关键技术研究

学 号：2013110604

姓 名：张 诚

专 业：计算机科学与技术

导 师：马华东

学 院：计算机学院

2016 年 1 月 21 日

独创性（或创新性）声明

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

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：_____

关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密在__年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

本人签名：_____ 日期：_____

导师签名：_____ 日期：_____

基于深度学习的步态识别关键技术研究

摘 要

生物识别技术是计算机视觉领域一个前沿的研究课题。在众多的生物特征当中，步态具有可远程获取、鲁棒性强、安全性强等优势。因此，在“以人中心”的现代智能监控系统中，步态识别技术获得了越来越多的关注。然而此问题存在着众多挑战，比如相同目标因不同视角、穿戴和行走速度带来的类内差异太大，以及不同目标之间的形态相似性带来的类间语义模糊。目前的步态识别技术大多基于手工设计的视觉特征来进行模式匹配，但是传统的人工特征已经无法满足步态精细识别的需求，所以很难打破特征提取和特征表示的瓶颈。

在本文中，我们围绕基于深度学习的步态识别问题，提出了一系列新模型和新方法。首先，我们研究设计了一个基于深度学习的步态识别技术框架。为了克服现有步态数据库样本容量小以及深度学习训练速度慢的挑战，我们将原始的步态序列进行融合，计算其步态能量图作为卷积神经网络的输入来对预训练的网络进行微调。然后，我们提出了基于 Siamese 神经网络的步态识别技术。基于 Siamese 的距离度量学习模型能够解决深度学习训练数据量不足以及分类和识别任务的领域鸿沟问题。最后，我们通过联合步态序列的三维卷积特征和 Siamese 结构在三维空间进行特征度量学习。该方法可以从连续的周期性步态序列中捕捉空间维度和时间维度的信息，进一步提高步态识别的准确率和实用性。经实验验证，本文提出的方法在步态属性分类和身份识别中都取得了理想的结果，在目前世界上最大的步态数据库中，本文算法比现有最好方法在身份识别任务中提高了约 5% 的正确识别率。

关键词：步态识别，深度学习，卷积神经网络，距离度量学习

RESEARCH ON KEY TECHNIQUES OF GAIT RECOGNITION BASED ON DEEP LEARNING

ABSTRACT

Biometrics based automatic human identification is one of the most fundamental and frontier research topic in computer vision. Among the massive biometric authentication traits, human gait gives more remarkable characteristics such as remote accessed, robust and security. Therefore, gait recognition, which aims essentially to discriminate individuals by the way they walk, has gained much attention in the human-centric modern intelligent video surveillance system. However, this problem remains many difficulties, such as the inconspicuous inter-class differences from different people, and the large intra-class variations from the same people (e.g., different viewpoints, clothing, belongings and walking speed), and still have large space to be improved. To solve these challenges, most existing methods mainly employ the handcrafted gait features, whereas, those features can extremely hard to break through feature extraction and representation bottleneck when facing with the requirement of fine-grained gait recognition task.

In this thesis, we propose a series of novel models and methods for deep learning based gait recognition problem. First of all, we design a conventional convolutional neural network (CNN) based gait recognition framework. The proposed framework uses Gait Energy Image (GEI) as the input of the CNN to fine-tune the pre-trained model, which is a good solution to solve the data limitation problem and speed up the convergence of new model. Next then, we exploit the Siamese neural network with a distance metric learning architecture to learn sufficient feature representations to tackle gait recognition for human identification. Finally, we propose a 3-Dimensional Siamese neural network to further improve the performance of the gait recognition. In this way, spatiotemporal deep features are extracted directly from sequential gait images via 3-Dimensional convolutional neural network without any

preprocessing, which is able to take great advantages of periodical dynamic and motion pattern of human gait. In the experiments, the evaluations on the world's largest and most comprehensive gait benchmark dataset demonstrate that our proposed methods can impressively beyond state-of-the-arts in gait-based gender classification problem and with nearly 5% improvement in human identification task.

KEY WORD: Gait recognition, deep learning, convolutional neural network, distance metric learning

目 录

第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 研究内容.....	4
1.3 论文结构.....	7
第二章 步态识别相关技术综述.....	9
2.1 引言.....	9
2.2 步态识别国内外研究进展.....	10
2.2.1 步态识别数据库.....	11
2.2.3 步态识别典型技术.....	12
2.3 深度学习.....	14
2.3.1 BP 神经网络基本原理.....	15
2.3.2 多层神经网络.....	16
2.3.3 深度卷积神经网络.....	16
2.4 距离度量学习.....	17
2.5 本章小结.....	19
第三章 基于传统卷积神经网络的步态识别技术.....	21
3.1 概述.....	21
3.2 步态能量图的生成.....	22
3.3 基于深度卷积神经网络的特征训练与提取.....	23
3.3.1 模型训练.....	23
3.3.2 特征与匹配.....	25
3.4 实验结果与分析.....	25
3.4.1 评测数据集.....	26
3.4.2 基于步态特征的性别分类.....	27
3.4.3 基于步态特征的身份识别.....	28
3.5 本章小结.....	29
第四章 步态识别中的距离度量学习模型.....	31
4.1 概述.....	31
4.2 基于 Siamese 神经网络的步态识别技术.....	32
4.2.1 网络结构.....	32
4.2.2 训练和测试策略.....	33

4.3	实验结果与分析.....	34
4.3.1	相同视角下的身份识别.....	35
4.3.2	不同视角下的身份识别.....	36
4.4	本章小结.....	37
第五章	基于时空联合深度神经网络的步态识别技术.....	38
5.1	概述.....	38
5.2	三维卷积神经网络.....	39
5.2.1	基于 CNN 的视频处理方法	39
5.2.2	基于三维卷积神经网络的步态序列特征提取.....	40
5.3	基于 3D-Siamese 网络的步态识别技术.....	42
5.3.1	特征训练.....	42
5.3.2	特征匹配与识别.....	43
5.4	实验结果与分析.....	43
5.5	本章小结.....	44
第六章	结束语.....	45
6.1	总结.....	45
6.2	展望.....	46
参考文献	48
致 谢	57
参加科研及获奖情况说明	58

第一章 绪论

1.1 研究背景及意义

当今社会，以物联网技术[1]为基础的监控设备日渐普及，视频监控应用日益广泛，监控视频数据呈现出爆炸式的增长。根据国际数据公司（IDC）2014 年的报告显示[2]，到 2020 年全球监控视频数据量将达到 5.8ZB，占全球数据量的 44%，堪称“最大的大数据”。众多的摄像头，庞大的监控网络，瞬间就会产生海量的视频数据，如何从这些海量数据中高效地提取出有用的信息，进而按照视频内容和特性去理解这些信息成为智能视频监控技术亟需解决的问题。



图 1-1 视频监控场景示意图

视频监控业务必须针对监控场景中的目标，而行人目标则是监控视频分析中最为感兴趣的对象[3][4][5]。如图 1-1 所示，在大部分的监控场景中，行人都是最为普遍与核心的监测对象，“以人中心”是现代智能监控系统的最大特点之一。行人身份识别[6]作为监控视频分析的核心，广泛应用于楼宇安防、智慧监狱、公共场所管理等各个方面。传统的目标识别技术往往根据对象的颜色、纹理以及形态等外观特征来进行识别。然而，对于行人识别问题，由于人体的非刚性特点，以及外观在不同环境下变化剧烈，仅仅依赖外观特征很难取得较高的识别精度，而步态这种新兴的生物特征却十分有效。

步态[7]，作为一种人类的行为特征，描述了人行走的姿态，包括行走过程中上、下肢关节处呈现出有规律的运动趋势和变化。基于视频分析技术的步态识别，就是通过记录、观察、分析图像序列中行人身体运动方式，然后建立步态模型，并提取稳定的参数特征，再通过计算机去识别的过程，最终可以得到行人的身份信息和属性（性别、年龄、种族等）信息。相比于其他生物特征（人脸、指

纹、掌纹、虹膜、声纹等), 步态是一种极具潜质的生物特征, 在面向智能视频监控的行人识别场景下, 步态特征的优势主要表现在以下三个方面: 1) 可远程获取。步态数据不需要特殊的设备进行采集, 监控人员只需利用普通的监控摄像头, 通过一定距离获取特定目标的步态信息, 采用非接触性的方式进行隐蔽采集。而诸如虹膜、指纹、掌纹等生物特征则需要用户的友好配合来完成采集, 这一点在智能视频监控中十分重要; 2) 鲁棒性。步态特征在低质的监控画面下仍然有较好的识别效果, 这使得步态识别技术既能够在室内场景也能在户外监控场景中维持较高的识别精度。相比而言, 精确的人脸识别和声纹识别则对数据源的质量要求较高; 3) 安全性。步态难以伪装、模仿和隐藏, 在公共场合行人要刻意改变自己的步态, 往往会变得更加可疑反而会引起公众的注意。综上所述, 步态识别技术已经成为计算机视觉与模式识别中一个重要的研究方向, 同时具有巨大的研究价值和市场需求。

从上世纪 60 年代至今, 大量的研究人员在医学[8][9][10]、心理学[11][12][13]和计算机视觉领域的研究都对人体步态特征的可识别性做出了科学性的解释。人体的步态模式最初被应用于医学领域, Murray[8][9]等人关于步态的研究在临床上被用于对一些疾病产生的异常行走姿态的诊断。这些工作从解剖医学的角度分析了每个人生理结构的差异性, 不同人的腿骨长度、肌肉强度、重心高度、运动神经灵敏度等 24 种成分和特性共同决定了人体步态的唯一性[10]。此后, 步态的良好特性立即引起了其他领域科研工作者的广泛关注。早期的关于步态的心理学研究[11][12]显示, 观测者在黑暗环境下仅仅通过行人关节处的反光片就能识别他们熟悉的朋友, 并且识别率高达 70~80%。不仅如此, 这些步态运动模式还能够精确反映行人的性别、体态(身高和体重)、状态(紧张或放松)和情绪(高兴或失落)等重要特性[13]。这些成功的医学和心理学研究都对人体步态模式的唯一性给出了验证, 同时指出了: 步态是一种可识别的周期性运动模式, 也为接下来基于计算机视觉的步态识别研究奠定了扎实的理论基础。

数十年来, 由于智能视频监控系统的快速发展以及安防监控巨大的市场需求, 基于机器视觉的自动步态识别技术逐渐受到人们的关注, 在论文的第二章中, 我们将会详细阐述目前基于视频分析的步态识别技术研究进展。根据文献调研我们发现, 尽管研究领域已经在步态识别技术上取得了一些进展, 但是现有的研究对于在大规模数据库中精确的步态识别还不够成熟, 识别的准确率较低。基于步态的身份识别技术所面临的困难在于行人的步态会遇到观测视角、衣着、携带物品、行走速度、光照和复杂背景等众多不确定因素的影响, 具体表现为: 1) 不同场景下的相同行人之间的类内差异太大(如图 1-2 所示); 2) 不同行人之间的类间差异太小(如图 1-3 所示)。目前的步态识别技术大体可以分为两类: 基于模型匹配的方法(Model-based)[14]和基于外观匹配的方法(Appearance-based)

[15][16][17]。基于模型匹配的方法从步态序列中抽取人体结构特征进行匹配，这种方法对采集图像的分辨率要求较高同时伴随着复杂的计算消耗。基于外观匹配的方法虽然适合于户外场景的步态识别，对采集图像分辨率要求有所降低，但是通常需要将步态图像序列转化为复杂的步态特征图来进行识别。因此，步态识别问题的挑战在于：1) 如何选择具有区分度的图像特征；2) 如何充分利用步态的周期性特性来进行时序建模。传统的手工设计特征（如 HOG[18]，SIFT[19]，LBP[20]，MSER[21]等）难以提取到精确的步态特征，很难打破目前步态识别技术的瓶颈，人体步态特征的多样性以及背景环境的复杂性都对传统的特征表示和模式识别技术提出了更高更新的挑战和要求。

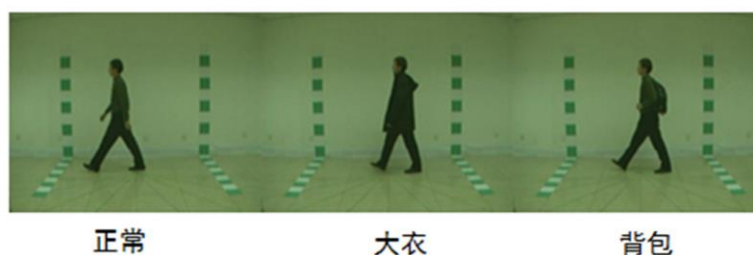


图 1-2 相同行人不同穿戴示意图



图 1-3 不同行人侧面 90°视角特征轮廓图

近年来，机器学习，尤其是深度学习[22][23][24]成为一个越来越热门的研究领域，它在语音、图像、OCR 等方面都有突破性的进展，并展现了强大的学习数据本质和高度抽象化特征的能力。层初始化等训练方法显著提升了深层模型的可学习性。与传统的机器学习模型相比，深层模型经过了若干层非线性变换，带给模型强大的表达能力，从而有条件为更复杂的任务建模。与人工特征相比，自动学习特征，更能挖掘出数据中丰富的内在信息，并具备更强的可扩展性。在大数据的趋势下，有了充足的训练样本，复杂的深层模型可以充分发挥其潜力，从而挖掘出海量数据中蕴含的丰富信息。强有力的基础设施和定制化的并行计算框架，让以往不可想象的训练任务加速完成，为深度学习走向实用奠定了坚实的基础。在学术界和工业界，已有 Caffe¹[25]、Theano²、Keras³、Torch⁴等多个针对不同深度模型的开源实现，Google、Facebook、百度、腾讯等公司也实现了各自的

¹ Caffe, <http://caffe.berkeleyvision.org/>

² Theano, <http://deeplearning.net/software/theano/>

³ Keras, <http://keras.io/>

⁴ Torch, <http://torch.ch/>

并行化框架。深度学习引爆的这场革命，将人工智能带上了一个新的台阶。但是，深度学习技术需要依赖于大量数据来进行模型训练，与之矛盾的是，在步态识别领域的数据库中，行人的身份数据很多，但是同一身份行人的步态数据却很少。不仅如此，传统的卷积神经网络主要应用于大规模的图像分类问题，而步态识别问题与分类问题存在着巨大的领域鸿沟，这就导致了直接应用传统的分类卷积神经网络到步态识别问题往往得不到理想的识别结果。因此，在视频监控领域探索将深度学习技术应用于步态分析虽然前景广阔，但是依然面临着很多挑战，这其中蕴含着巨大的学术意义和很强的实用价值。

综上所述，步态识别是一个热点研究课题，具有巨大的研究价值和市场需求，而现有的研究方法并没有解决其面临的问题。因此，本文选取行人的步态分析作为研究方向，就其尚待解决的热点难点问题进行了深入研究，提出了高效、准确地步态识别算法，旨在解决如下挑战问题：如何通过深度学习技术在步态识别问题上选择具有区分度的步态特征；如何克服目前深度学习方法中的分类模型与步态识别问题的领域鸿沟；如何利用步态的周期性模式对步态序列进行特征表示。如图 1-4 所示，本文针对步态识别技术所涉及到的这些关键问题展开了深入研究，希望在提高步态识别方法精度的同时，完善步态识别技术的理论研究，最终为高精度的步态识别走向应用提供技术支持。

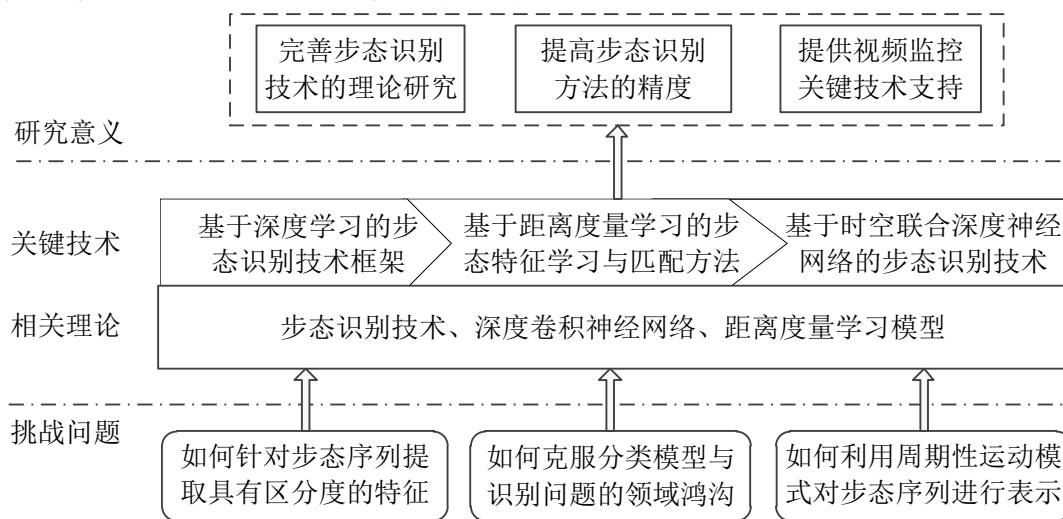


图 1-4 本文研究框架图

1.2 研究内容

本文我们重点研究了基于深度学习的步态识别若干关键技术(如图 1-5 所示)，包括基于传统卷积神经网络的步态识别方法、距离度量学习理论在步态识别中的应用以及基于时空联合深度神经网络的步态识别技术。具体的研究内容包括以下三个方面：

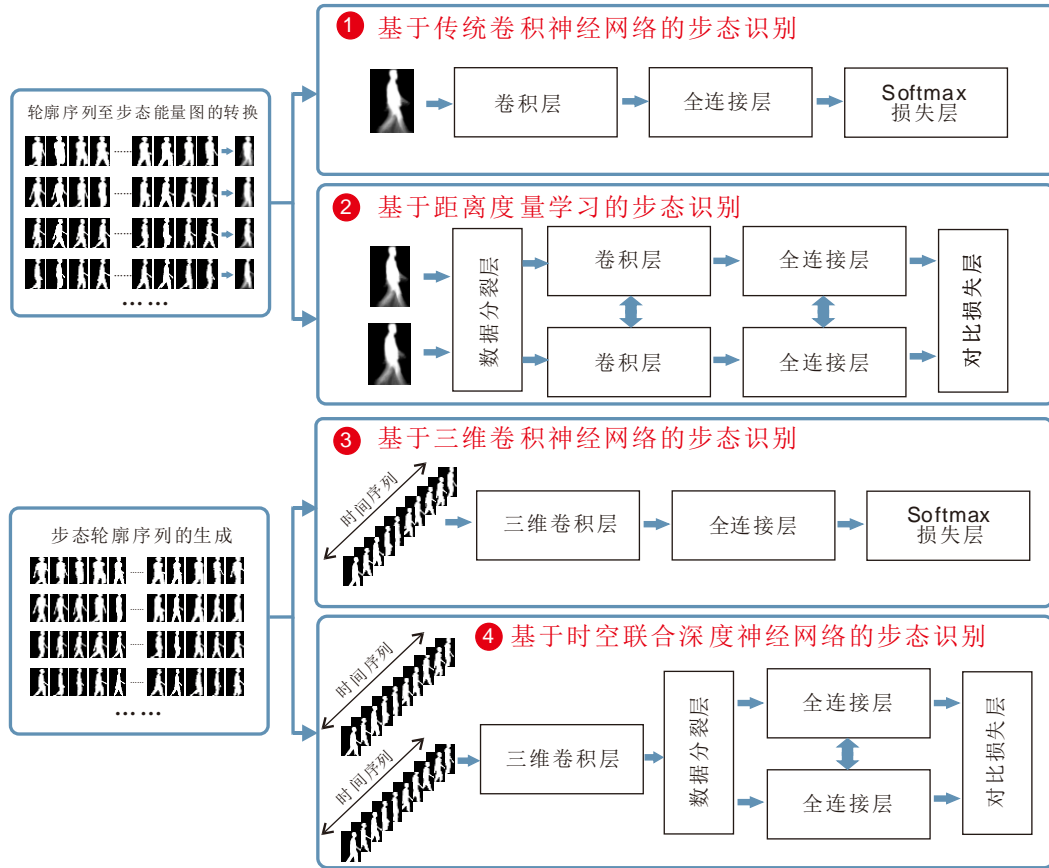


图 1-5 本文整体技术路线

(1) 基于传统卷积神经网络的步态识别技术

步态识别问题的首要挑战在于如何选择具有区分度的图像特征。传统的人工图像特征已经无法满足步态精细识别的需求，将深度学习模型引入步态识别问题有助于解决特征选择问题。目前流行的步态识别技术框架通常是先对连续的步态帧序列进行预处理，再将一个完整周期的轮廓序列采用某种方式转化为一张步态特征图，对行人身份和属性的判别就是基于该步态特征图的视觉特征。步态特征图是各种复杂的因素以非线性的方式结合在一起的，它包含了行人的身份、姿态、年龄、行走速度、环境光线等各种信息，传统的单一手工设计特征或者是特征融合的表达方法都很难处理这种复杂的步态特征图。深度学习的关键就是通过多层非线性映射将这些因素成功地分开，例如在卷积神经网络的最后一个隐含层，不同的神经元代表了不同的因素，如果将这个隐含层当作特征表示，身份识别、年龄估计、性别分类就会变得非常简单，因为各个因素之间变成了简单的线性关系，不再彼此干扰。考虑到深度特征的优势以及现在步态特征的局限性，本文的第一个研究重点就是设计一个基于深度学习的步态识别技术框架来提高识别的精度。为了克服现有步态数据库样本容量小以及深度学习训练速度慢的挑战，我们将原始的视频帧序列进行融合，计算其步态能量图作为深度神经网络的输入。然后，

将 ImageNet⁵上训练得到的模型作为起点，利用步态训练数据集和反向传播对其进行参数微调（Fine-tuning），将模型适应到步态识别任务。接下来，将训练好的模型作为特征提取器，利用其隐含层的输出作为步态特征表达，并使用 K 近邻算法计算测试步态序列与数据库中待识别步态序列的距离，根据相似度最大时匹配库中的优化步态特征对应的已知人物信息，确定所述待识别人的身份信息。实验表明，将深度卷积神经网络直接应用于步态识别问题时能显著提高基于步态的属性分类任务的精度，但是其在身份识别任务的精度方面还有待提高。

（2）基于距离度量学习的特征匹配学习方法

将传统的卷积神经网络直接应用于步态特征的提取，虽然能够获取区分度较强的图像特征，但是仍然面临着很多问题。首先是训练数据量的不足。深度模型是一种多层的复杂神经结构，是一个良好的非线性的生成模型，如果训练的数据量不足容易出现过拟合或者欠拟合的情况。其次是分类问题与识别问题存在领域鸿沟。目前的深度学习模型最开始主要应用于大规模的目标分类问题，然而，从模式分类的角度来看，步态识别问题与分类问题存在巨大的领域鸿沟，这就导致了直接应用传统的分类卷积网络到步态识别问题往往得不到理想的识别结果。针对这些问题，本文将引入距离度量学习模型来解决训练数据量不足以及分类和识别任务的领域鸿沟问题。Siamese Neural Network 最初由 Yann LeCun[84][85]提出并被成功应用于人脸识别与手写签名识别任务，Siamese 结构由两个并行的卷积神经网络构成，网络的每一层之间共享权值，只需要输入相同行人的图像对（正例）和不同行人的图像对（负例）进行训练，其主要思想在于通过学习一个损失函数将输入的数据模式映射到一个隐藏的语义空间，具体到步态识别问题上，对于相同行人的步态图，其相似性距离变小，对于不同行人的步态图，其相似性距离变大。这种并行的卷积结构十分有利与特征识别于匹配任务。因此，本文探索利用并行神经网络结构来克服步态识别的问题中的诸多挑战问题，提出一整套完整的适用于步态识别问题的深度学习解决方案。实验表明，在相同视角和不同视角的步态识别任务中，本文提出的方法与已有最好的方法相比，平均提升了 5% 的正确识别率。

（3）基于时空联合深度神经网络的步态识别技术

二维卷积神经网络虽然具备提取丰富具有区分度特征的能力，但是受限于二维空间的图像输入。将二维卷积神经网络应用与步态识别问题时，需要设计复杂的步态特征图作为输入，特征图的计算过程中往往伴随有步态时序信息的损失，输入信息的不准确可能导致输出结果的不准确，从而不能较有效地提高识别的准确性。步态作为一种运动模式，具有空间和时间两个维度的特征，周期性是步态

⁵ ImageNet, <http://image-net.org/>

识别与传统的人脸识别或者动作识别最为显著的差异,如何充分利用最为原始的步态序列图成为另一个提高步态识别准确率的瓶颈。三维卷积神经网络可以用于动作识别,通过从空间和时间的维度提取特征,然后进行 3D 卷积,以捕捉从多个连续帧得到的运动信息。在这个结构中,卷积层中每一个特征图都会与上一层中多个邻近的连续帧相连,以此捕捉运动信息。在步态序列中使用三维卷积神经网络模型,同时结合并行的 Siamese 神经网络结构,可以在时间和空间维度获取更有区分性的步态特征。因此,本文提出通过计算步态运动的三维卷积特征来增强模型,该模型可以从连续性的周期序列中捕捉时间和空间信息,进一步提高步态识别的准确率和实用性。除此之外,我们提出了将三维神经网络与 Siamese 结构相结合的 3D-Siamese 网络,这种时空联合深度神经网络能够在三维空间进行特征度量学习。实验证明,本文算法能够充分利用二者的优势,从而达到最佳的步态识别效果。

1.3 论文结构

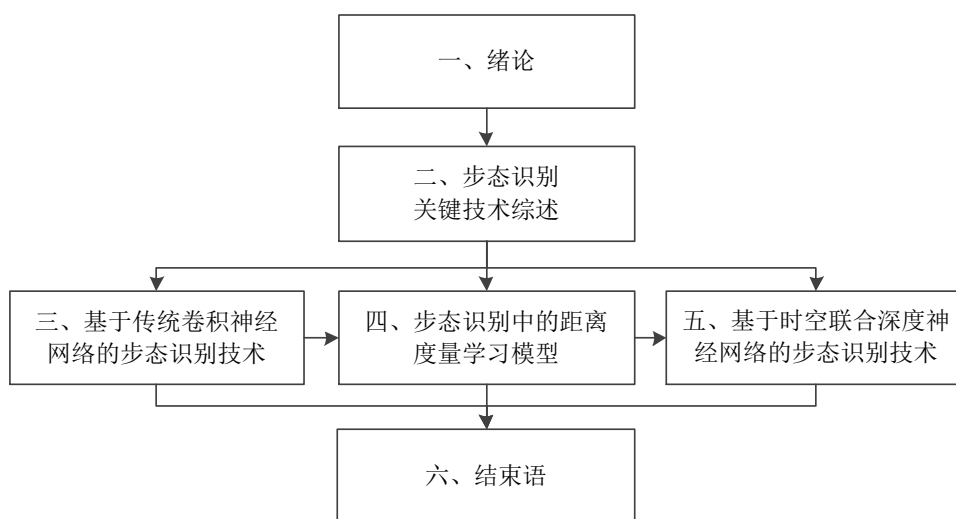


图 1-6 论文章节结构图

综上所述,本文分别对基于传统卷积神经网络的步态识别技术、基于距离度量学习模型的步态识别技术以及基于时空联合深度神经网络的步态识别技术等方面进行深入的研究,由此构成了一个完整的基于深度学习的步态识别技术研究框架。论文分为六章,各章关系如图 1-6 所示,全文的内容组织如下:

第一章,论述论文的研究背景和研究意义,并概述研究内容和研究成果。

第二章,对步态识别技术的研究现状、研究方法及存在的问题进行系统概述和总结,并对深度学习和距离度量学习等基础理论进行简要介绍。

第三章,阐述了基于传统卷积神经网络的步态识别技术。主要研究了基于步态的属性分类和行人身份识别。

第四章，阐述了基于距离度量学习模型的步态识别技术。本章针对传统卷积神经网络不能直接应用于步态识别任务的问题，提出将度量学习模型引入步态特征训练框架，从而克服了传统神卷积神经网络在分类和识别问题中的领域鸿沟以及训练数据量不足的问题。

第五章，阐述了基于时空联合深度神经网络的步态识别技术。具体来说即作为一种基于并行的三维卷积神经网络的步态识别技术。本章依托于第三章和第四章的研究成果，针对步态的周期性特点，从空间和时间两个维度来提取更富表达力的步态特征。

第六章，结束语。本章对论文的研究成果进行总结和讨论，并给出了下一步研究工作的方向。

第二章 步态识别相关技术综述

2.1 引言

步态识别作为一种新兴的生物识别技术具有可远程获取、隐蔽采集、鲁棒性强、难以伪装和模仿等优势,因此受到了越来越多的关注。基于视觉技术的步态识别一直是计算机视觉与模式识别领域重要的研究方向,研究人员始终尝试在视频监控应用中取得突破。因此,在过去的十五年里,很多基于不同技术的步态识别方案如雨后春笋般层出不穷,包括步态算法验证数据库的构建、设计更为高效的步态特征图等等,其目的是进一步提高步态识别方法的性能。但是,受限于传统人工特征的瓶颈,没有任何一种步态识别方案能在不同环境下都取得完美的效果。值得一提的是,无论这些方法在实现细节上有何不同,从算法的角度来看,一个典型的步态识别关键技术点却是相同的,它主要包括步态数据采集、数据预处理、步态周期性检测、步态特征提取和特征分类与匹配五个不同的阶段。这五个技术阶段中,又以如何获取优良的步态特征以及特征匹配两个方面受到最多的关注,需要声明的是步态数据的采集和预处理等步骤并不在本文的讨论范围之内。本章主要回顾步态识别的国内外研究进展,包括整体方案、步态数据库和典型技术等。同时,本章还简要介绍了深度学习和距离度量学习的背景知识和基本理论。下面首先对上文所述的五个步态识别技术阶段进行简要概述。

(1) 步态数据采集

步态数据采集是步态识别的第一步。按照数据源的划分,目前步态识别的研究包括基于特殊传感器数据(足底压力传感器[26][27]、穿戴式传感器[28])和基于视频数据两大类,本文主要研究基于视频数据的步态分析。为了获取步态视频,通常的做法是使用视觉传感器对行人行走的过程进行拍摄,同时记录每帧的步态位置及姿态。从视觉传感器类型来看可以采用普通的 RGB 摄像头、深度摄像头[29]或者夜间红外摄像头[36]。从数据采集形式来看可以是单摄像头采集、多摄像头同步采集或者三维步态信息采集。

(2) 数据预处理

在获取步态数据之后,一般情况下需要对其进行预处理。预处理的目的在于针对上一步采集的步态视频,将人体轮廓和背景环境分割开来,良好的人体轮廓分割效果对于后面的特征提取和识别影响重大。前景分割一直是计算机视觉领域的热点难点问题,由于复杂的背景、人体的非刚性特点,分割出清晰的轮廓是十分具有挑战性的工作。广泛采用的方法包括背景估计法[30]、帧间差分法[31]、图像分割法[32][33]等等。近些年,也有一些工作试图绕过人体轮廓分割的过程,

采用光流法[34]或者基于运动特征的方法直接提取步态运动模式特征进行分析。实验表明,这两种思路都取得了较好的成果。

(3) 步态周期检测

步态是一种周期性的图像信号,一段步态视频往往包含多个周期的步态序列。因此,对一个完整步态周期序列进行特征提取有助于提高步态识别的准确率。考虑到步态轮廓的噪声,信号处理领域经常采用的峰值检测(Peak Detection)和最大自相关系数(Maximizing Autocorrelation)可以直接应用于步态的周期判断。典型的周期检测方法包括步态剪影的宽、高之比,二值轮廓序列中人体前景像素点个数[37]或者下半身像素点占比[38]等。当下,也有一些先进的步态识别算法试图基于准周期、半周期或者不完备周期来进行行人的步态分析[35],在实际的监控场景中,往往难以获取精确的步态周期,所以这个方向依然是步态识别领域一个尚未突破的难点。

(4) 步态特征提取

特征提取步态是步态的核心环节,同时也是研究人员最为关注的内容,大量可行的方案已经被相继提出。根据步态特征提取的方式不同,已有的特征表示方法可以分为基于模型的方法和基于外观的方法。两种方法根据其特性的不同,适用的场景也不同,我们在下面的章节中会对这些方法进行分类介绍。

(5) 特征分类与匹配

在获取步态特征之后,可以直接对两段步态序列计算特征相似度(比如欧氏距离、余弦距离等)来判断两段序列是否是同一身份的行人。最近很多研究着眼于将主成分分析(Principal Component Analysis, PCA)和线性判别分析(Linear Discriminant Analysis, LDA)方法[17]进行结合来对特征进行分类和判别。为了进一步降低特征提取的干扰性,研究人员也采用排序支持向量机(Primal Rank Support Vector Machine) [39]来进行特征分类和匹配。

在本章接下来的内容中,我们对已有步态识别数据库和步态识别技术,尤其是特征建模方法进行了综述,并分析其优势和不足,从而帮助我们研究更适合的步态特征提取和表示方法。同时,围绕深度学习和距离度量学习等关键技术,本章亦对二者的基本原理和背景知识进行了详细阐述。

2.2 步态识别国内外研究进展

目前,来自世界范围内的很多科研机构已经开始对步态识别技术进行研究,其中包括 MIT、Carnegie Mellon University、Georgia Institute of Technology、University of Southampton (UK)、University of South Florida、University of Maryland、Osaka University (Japan)、北京航空航天大学、中国科技大学、西安交通大学、北京邮电大学、中科院自动化所等多家国内外著名高校和研究机构,步

态识别主流研究团队也是从这些机构产生，其相关成果发表于计算机视觉、模式识别、人工智能、多媒体计算和生物识别领域的顶级期刊（IEEE TPAMI、IEEE TIP、IEEE TMM、IJCV）和顶级会议（CVPR、ICCV、ECCV、ACM MM、AAAI、IJCB、FG）。“工欲善其事，必先利其器”，已有的研究工作主要围绕构建步态测评数据库以及提出新的步态识别算法两个方面，本节将从这两个方面对步态识别的国内外研究进展进行综述。

2.2.1 步态识别数据库

作为步态识别研究的第一步，步态数据的采集建库非常重要，好的数据库能够科学合理公正地评价算法好坏，同时为步态识别技术从实验室走向产业应用奠定良好的基础。目前，很多高校和科研机构分别建立了自己的步态数据集并开放给全球各地的研究人员进行研发（如表 2-1 所示），大家在这些公开数据集中测试步态识别算法的性能，相互交流合作，促进学术发展。通常一个完善的步态数据库需要满足很多要求，从本世界初期至今的研究过程可以看出，步态数据库的构建有以下三个趋势：1）样本容量逐步增大。文献中最早公开的步态数据库是 UCSD，该数据库只包含了 6 个人的行走数据。2000 年开始，受到美国国防部高级研究计划局资助的 HID⁶项目影响，众多研究机构开始进行步态识别研究，并且陆续推出了各自的步态数据库。从 Georgia Tech[40]、MIT[41]、CMU[42]、UMD[43]的几十人，逐步发展到 SOTON[44][45]、CASIA[46][15][36]、OU-ISIR[47]的成百上千人的规模，毫无疑问的是，数据库的规模越大对步态算法的挑战越大，同时更接近实际步态识别的应用需求；2）观测视角和环境逐渐复杂。多视角步态识别和复杂环境场景是步态识别的实际需求，CASIA-B[15]数据库采集了相同行人的 11 个视角的步态数据，SOTON[44][45]和 USF[48]数据库则更多考虑除了步态之外的其他影响因素，比如鞋子类型、携带物、行走路面等等。3）评估意图更为明确。不同的数据库往往在评估目的上有所侧重，比如 CASIA-C[36]意在评估室外夜间环境下的步态识别算法，OU-ISIR LP[47]则因为其规模超过 4000 人而特别适合于进行行人身份识别，CASIA-B[15]也常常用于基于步态的性别识别等任务中，近些年提出的 TUM[49]数据库已经开始考虑利用深度信息来进行步态识别。

在本文的步态识别实验我们用到了其中两个步态数据库：CASIA-B 和 OU-ISIR Large。这主要是基于以下三个方面的考虑：1）数据库的样本容量足够大。CASIA-B 和 OU-ISIR Large Population 是目前样本规模最大的两个数据库，对于行人识别任务来说十分具有挑战性；2）数据库的使用范围较广。CASIA-B 数据库考虑了样本容量、性别、穿戴、视角变化等多个因素，因此也是最为广泛

⁶ Human ID Challenge Problem at USF, <http://figment.csee.usf.edu/GaitBaseline/>

使用的步态算法测评标准集； 3) 数据库较新。OU-ISIR Large Population 数据库是目前世界上最大的步态识别数据库，从 2012 发布以来，目前还在持续更新，它给出了标准的查询（Probe）和源数据（Gallery）的目录，非常适合于各种算法进行效果比较。

表 2-1 常用步态识别数据库

数据库名称	发布时间	样本容量 (人数, 序列数)	评估意图
UCSD	1998	6, 42	户外、最早的步态数据库
Georgia [40]	2001	20, 268	户外、多视角、多距离
CMU [42]	2001	25, 600	室内、多视角、步速、抱球
MIT[41]	2001	24, 194	室内、侧面视角
HID-UMD[43]	2001	55, 222	户外、多视角
SOTON Large[45]	2001	115, 2128	户外+室内、多视角
SOTON Small[44]	2001	12, 100+	室内、多视角
USF[48]	2005	122, 1870	户外、鞋子、携带物、阴影遮挡、时间（33 人在半年内采集两次）
CASIA-A[46]	2001	20, 240	户外、行走方向
CASIA-B[15]	2005	124, 1240	室内、多视角、穿着、性别
CASIA-C[36]	2005	153, 1530	夜间红外、步速、背包
OU-ISIR LP[47]	2012	4007, 7842	室内、多视角、穿着、年龄
TUM-GAIT[49]	2014	305, 3370	户外、穿戴

2.2.3 步态识别方法分类

表 2-2 步态识别方法分类

	模型特征	外观特征
空间建模	基于腿部二维结构分析 基于全身二维结构分析 基于三维结构分析	基于步态轮廓图的分析 基于步态轮廓线的分析
时间建模	基于时序建模的分析 基于人体参数的分析	基于光流的分析 基于周期性的分析 基于动态匹配的分析 基于时间切片的分析

步态分析的两个关键步骤是：1) 步态特征的提取和 2) 特征匹配和运动模式分析，目前研究人员针对这两个方面提出了大量切实可行的实施方案，从特征

和表达的角度来看，步态识别技术大体上可以划分为两类[50]：基于模型的方法（Model-based）和基于外观的方法（Appearance-based），其中基于外观的方法又称为非模型分析方法（Model-free）。从步态的空间特征和时间特征两个方面来看[51]，步态识别技术又可以划分为基于空间特征分析的方法和基于时序建模的方法。因此，很难将现有的某种方法严格地划分到某一类别之下，本节试图结合以上这两个角度对目前的步态特征建模方法进行回顾。表 2-2 中对相关方法进行了大类别的归类，下面依次对其中每类中的典型方法做一个简要的概述。

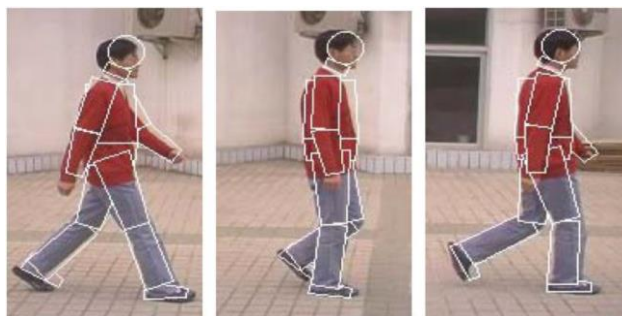


图 2-1 人体全身结构模型提取结果[54]

基于模型特征的方法来源于人的运动动力学知识，这些先验知识可以转化为步态的形状和参数特征，研究人员从中提取出便于识别的符合人体步态特征的静态参数来构建物理模型。从空间维度来看，人在运动过程中全身或者腿部的运动可以作为分析对象，具体而言这类方法可以分为基于腿部二维结构、全身二维结构和三维结构这几类方法。从时间维度来看，可以分为基于时序建模的分析和基于人体参数的分析。Yam[52]等人利用行走过程中腿部的摆动来构建了两个模型：双边对称（Bilateral symmetric）和振荡模型（Oscillator model），然后通过傅里叶变换使用相位和幅值的乘积来表征膝盖和大腿的运动模式。除了腿部之外，很多研究人员也提出利用身体其他部位的结构特征来辅助步态的识别，基于全身二维结构的方法获得了很多关注。比较典型的是以下几种：David K Wagg[53]等人通过提取人体周围边缘信息，通过椭圆或者多边形拟合的方法提取轮廓参数，并用于最终的步态识别。Wang[54]等人将行人的静态躯干模型和运动模式相结合，使用了一种特征融合的方式来提高步态识别的准确率。Frederic Jean[55]将行人头部和脚步的三条运动轨迹进行识别，结合光流法对腿部的关键点进行跟踪。进一步地，随着计算机处理复杂数据的能力逐步增强，研究人员开始研究基于三维结构的步态识别方法。Zhao[56]等人利用多帧图像来构建人体三维模型，并通过一个局部最优化的算法来处理运动模式，用三维结构来处理步态分析问题的优势是能够处理跨角度步态识别问题，但是同时需要构建比较大规模的步态数据库，同时计算复杂度也比较大。

以上基于模型的方法优势在于能够比较精确地对人体进行建模同时获取步

态运动参数,并且能够克服一些目标遮挡的情况,但是缺点也很明显。基于模型的方法往往对采集图像的分辨率要求较高同时伴随着复杂的计算消耗,在实际场景中,监控摄像头的分辨率通常难以达到要求。与之相对应的是基于外观特征分析的方法,在这类方法中,研究者通过前景背景分割,从步态的序列中直接抽取视觉特征,然后计算视觉特征之间的相似性来进行步态识别。这类方法主要包括基于步态轮廓图和步态轮廓线两类,这两类方法的大体思路都是通过背景减除、帧间差分、图像分割等技术先提取人体形状或者轮廓,然后再在提取后的特征图上提取空域视觉特征,比如LBP[20]特征、Gabor[57][58][59][60]特征等。为了充分利用步态的时域特征,研究者陆续开始结合一些时间维度的信息来进行时间建模,比如基于光流的分析[61][62][34]、基于周期性的分析[17][63][62][64][65]、基于动态匹配的分析[66][38][67][68]和基于时间切片的分析[69][70][71]等。基于外观匹配的方法虽然适合于户外场景的步态识别,对采集图像分辨率要求有所降低,但是通常需要将步态图像序列转化为复杂的步态特征图来进行识别,如何选择具有区分度的图像特征成为一大难点,如何充分利用步态的周期性特性来进行时序建模也是亟待解决的问题。

总体而言,不同的步态识别方法各有优劣,应该根据具体环境来选择不同的方法。比如,在图像分辨率较高的情况下,一旦行人运动过程往往没有规律同时伴随着一定遮挡和穿着的变化,基于模型特征的方法往往更有效;在视角变化比较剧烈并且步态数据库充足的情况下,基于三维模型的步态识别方法效果显著;在图像分辨率较低的情况下,基于外观匹配的方法往往鲁棒性更强。但是总体而言,通过文献调研来看,近几年步态识别方法呈现出以下几个趋势:1) 基于外观特征的研究更为热门。一方面是因为模型匹配的方法往往更依赖于高质的步态数据源,并不能很好的地适应实际的应用场景。另一方面是因为随着计算机视觉领域特征表示方向的不断发展,视觉特征建模逐步走向成熟,尤其是深度学习的发展逐步打破了以往人工特征表示的瓶颈;2) 结合步态时序特征的是提高识别准确率的潜在方向。目前已经有很多研究者意识到时序特征对于步态识别的重要性,然而这些方法往往计算比较复杂,比如光流法、动态时间匹配等方法。如何将步态的时域空域特征进行融合成为一个难点。

2.3 深度学习

深度学习是机器学习技术的一个重要分支[23][24]。众多深度模型在过去的十年中陆续被提出,包括:深度置信网(Deep Belief Network, DBN)[72]、深度神经网络(Deep Neural Network, DNN)[76]、玻尔兹曼机(Boltzmann Machines, BM)[73]、限制性玻尔兹曼机(Restricted Boltzmann Machines, RBM)[74]、深

度玻尔兹曼机（Deep Boltzmann Machine, DBM）[75]等。本文主要研究基于卷积神经网络[77]的步态识别技术。因此在本节中，我们将简要介绍神经网络，尤其是卷积神经网络相关的背景和理论知识。

2.3.1 BP 神经网络基本原理

深度神经网络是传统 BP 神经网络的进化和发展。以监督学习为例，假设我们有训练样本结合 (x^i, y^i) ，那么神经网络算法能够提供一种复杂的非线性映射模型 $h_{W,b}(x)$ ，它具有参数 W 和 b ，可以用这些参数来拟合我们的数据 y^i 。

首先介绍最简单的神经网络，图 2-2 为神经网络中每一个单元的构成，我们称之为一个“神经元”，该单元也称之为 Logistic 回归模型，其输出可以表示为：

$$h_{W,b}(x) = f(W^T x) = f\left(\sum_{i=1}^3 W_i x_i + b\right) \quad (2-1)$$

其中“神经元”是以 x_1, x_2, x_3 及偏置“+1”作为输入的单元，函数 $f(\cdot)$ 为“激活函数”，通常使用 Sigmoid 或者 tanh 函数。以 Sigmoid 为例，Sigmoid 具有单调递增以及反函数单调递增等性质，常被用作神经网络的阈值函数，将变量映射到 $[0,1]$ 之间。

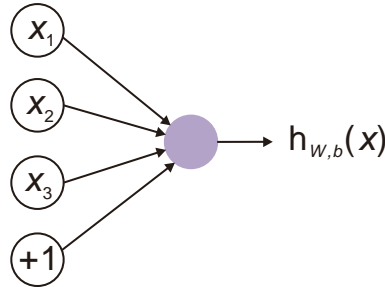


图 2-2 神经元示意图

所谓神经网络就是将许多“神经元”联接在一起，这样，一个“神经元”的输出就可以是另一个“神经元”的输入，就像大脑结构分层次启发，神经网络可以模仿人脑那样高效准确地表示和处理信息。如图 2-3 所示，我们用圆圈来表示神经网络的节点，标上“+1”的圆圈被称为偏置节点，最左边的一层称为输入层，中间所有的节点组成的一层称为隐藏层，最右边的一层称为输出层。

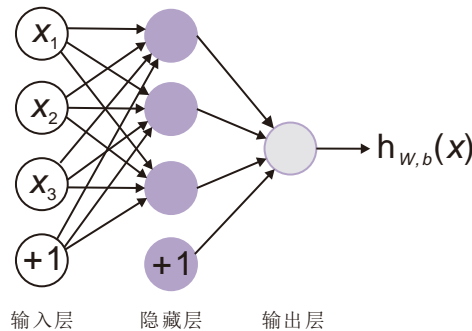


图 2-3 传统 BP 神经网络示意图

BP 神经网络是神经网络发展的里程碑式的成绩,它可拓展性强,计算简单,并且能够完成多种模式识别任务。然而在实际应用过程中,它也有一些不足,比如收敛速度慢,容易过拟合,优化过程容易陷入局部最优。面对复杂的模式识别任务是,简单的 BP 神经网络已经不能满足研究者们的需要。

2.3.2 多层神经网络

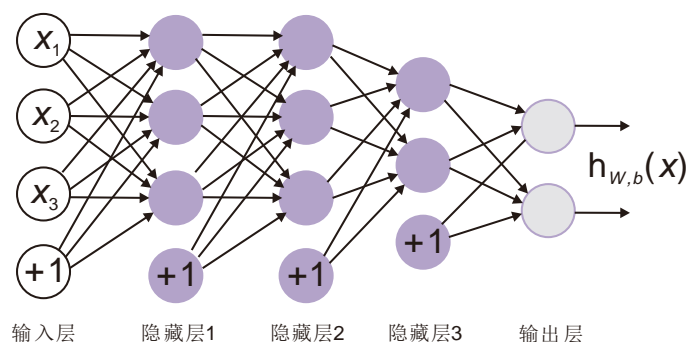


图 2-4 多层神经网络示意图

前文所述通过构建输入层、隐藏层和输出层生成了一个三层网络,虽然该网络在面对简单问题是十分有效,但是它还仅仅是一个非常“浅”的网络,无法应对复杂高维问题。受到大脑结构分层次启发,研究人员一直致力于多层神经网络的研究,多层神经网络的构建可基于浅层网络。如图 2-4 所示为包含多个隐藏层的神经网络,最左侧为输入层,最右侧为输出层,中间的每个层之间紧密相联。这种模式下,要计算神经网络的输出结果,我们可以按照之前描述的等式,按部就班,进行前向传播,逐一计算第一个隐藏层的所有激活值,然后是第二个隐藏层的激活值,以此类推,直到第三个隐藏层。这是一个前馈神经网络的例子,因为这种联接图没有闭环或回路。要求解这样的神经网络,需要样本集 (x^i, y^i) ,其中 $y^i \in \mathbb{R}^2$ 。如果需要预测的输出是多个的,这种神经网络很适用。综上所述,多层神经网络的优化过程其本质上是 BP 神经网络优化过程的进一步拓展,因此不可避免存在收敛速度慢,过拟合以及容易陷入局部最优等不足。

2.3.3 深度卷积神经网络

卷积神经网络是深度学习的一种典型架构,它特别适合对多维图像进行处理,原因在于为其网络结构与图像的拓扑结构相匹配,使得图像的原始像素点可以直接作为模型的输入,降低了数据的预处理要求,避免了传统模式识别算法中复杂的特征提取和数据重建过程。卷积网络通过图像的一小部分(局部感受区域)作为层级结构的最低层的输入,将信息依次传输到不同的层,每层通过一个卷积滤波器去获得输入数据的最显著的特征,例如一些边缘信息或者角点特征,所以这

个方法对于平移、比例缩放、倾斜或者其他形式的形变具有高度不变性。

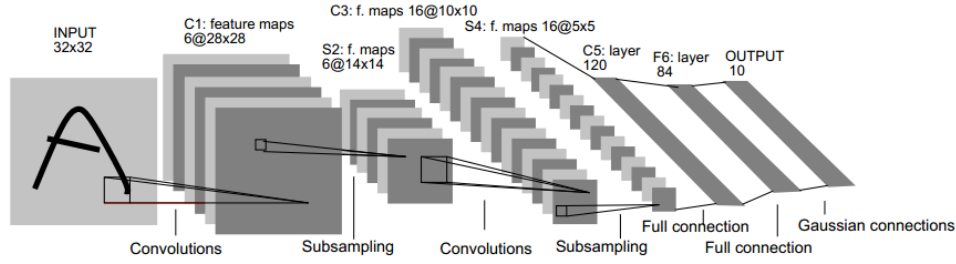


图 2-5 深度卷积网络示意图

如图 2-5 所示为深度卷积网络 LeNet 的示意图，它通常包括卷积层、降采样层（也称池化层）、全连接层和输出层，以实现逐层提取图像的特征。卷积层的作用是针对给定的输入图像矩阵，选取若干个卷积核（网络参数）对输入做卷积滤波，得到的输出可视为输入的一种特征表示，即通过卷积映射到下一层的特征图层。一般地，当前层的神经元与前一层的局部感受区域相连接，并提取该局部的特征，记录了前一层网络的空间位置信息。对于降采样层而言，通过池化变换可由下层的表示得到上一层表示的映射，降采样层保证了网络对于输入矩阵的平移和旋转不变形，对输入样本具有较高的抗畸变能力，同时起到了对数据进行降维的作用，减少了整个卷积神经网络需要学习的参数个数。对于全连接层而言，首先将上一个降采样层的输出由三维矩阵形式拉直为一维向量，然后，在每相邻两层之间构建一个非线性变换，通过所述非线性变换可由下一层的表示得到上一层表示的映射。多个全连接层可以看做上一节所述的多层神经网络。对于整个卷积神经网络而言数据的原始信息通过多层的卷积层、池化层和全连接层之后，卷积网络得到图像数据的深度特征，根据训练的任务的不同，通过最终的损失函数层对图像的特征进行分类判定或者回归分析等等。

2.4 距离度量学习

在计算机视觉领域，算法处理的对象往往是图像或者视频序列这种非结构化的数据，非结构化意味着它们并不能够直接向量化，而需要通过特征变化转化为自然的向量表达式，因而很多传统的基于向量空间输入的机器学习算法并能直接适用于图像或视频处理。具体来说，基于向量空间的距离函数度量方法本质上是把每一幅图像通过某种映射转换到欧式空间的一个点，并用欧式空间的良好性质在其中进行学习和训练，例如：对于 n 维空间内的两个特征向量 $x = (x_1, x_2, \dots, x_n)^T$ 和 $y = (y_1, y_2, \dots, y_n)^T$ ，可以用欧式距离表示它们之间的距离

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \sqrt{(x - y)^T (x - y)} \quad (2-2)$$

但是定义图像特征之间的欧式距离并不能良好地反应数据之间的语义相似性，距离度量的好坏直接会影响学习器的效果。例如在图 2-6 中，如果利用欧式空间作为度量空间，图中黑色（A）样本和红色（B）样本之间的距离要小于黑色（A）样本和绿色（C）样本之间的距离，然后从语义空间角度来看，绿色（C）样本位于以黑色样本（A）为中心的椭圆空间边沿，而红色样本（B）却位于椭圆空间之外，显然，如果可以根据训练样本定义一种能够合理描述当前样本的距离度量方法，能够大大提高进一步的数据分析的效果。

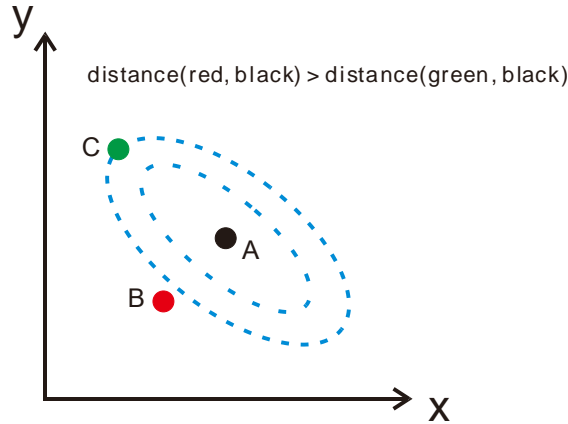


图 2-6 度量学习原理示意图

距离度量学习（Distance Metric Learning, DML），或者简称度量学习，是通过已经标定或者未标定的样本，计算出一个能够在已经数据样本下最能刻画样本相似度度量的矩阵或者距离函数，通过学习出来的度量矩阵或者距离函数，输入特征能够映射到一个隐含的语义空间当中，再在这个隐含的语义空间中计算不同样本之间的相似度。从形式化的角度来看，距离度量可以表示为一个 Mahalanobis 距离[82][83]，Mahalanobis 距离是对欧式距离的扩展，同样地，对于 n 维空间内的两个特征向量 $x = (x_1, x_2, \dots, x_n)^T$ 和 $y = (y_1, y_2, \dots, y_n)^T$ ，二者的 Mahalanobis 距离表示

$$d_A(x, y) = \|x - y\|_A = \sqrt{(x - y)^T A (x - y)} \quad (2-3)$$

其中， $A = V^T D V$ ， V 表示 A 的特征向量矩阵， D 是对角阵并且对角线为 A 的特征值。这里需要注意的是 A 必须为半正定矩阵，这样才能保证 $d_A(x, y)$ 的度量性。Mahalanobis 距离本质上是对原始的样本点进行线性变换和伸缩的操作，不同的学习算法得到的距离参数都会不同，因此能够适应不同的应用，大幅度提升学习算法的性能。在步态识别问题中，由于步态特征图包含了行人的身份、年龄、性别、体态等多种信息，因此需要根据问题的不同，有针对性地设计特征距离度量方式。度量学习已经被广泛应用于机器学习、计算机视觉等任务[78][79][80][81][83][84][85]。从训练数据的格式角度来说，目前 DML 的方法可以分为两类：输入为二元组和三元组。其中二元组是基于成对约束的度量学习方

法，每一对输入图像伴随着表示该对图像是否相似的二值标签。三元组则包含每一个输入样例及其相似图像与非相似图像的样例。目前也有一些方法并不局限于输入数据格式，而是在机器学习的框架之下直接利用数据标签自动地学习不同类别直接的距离度量方式，比较典型的是 Large-Margin Nearest Neighbours (LMNN)[78]。

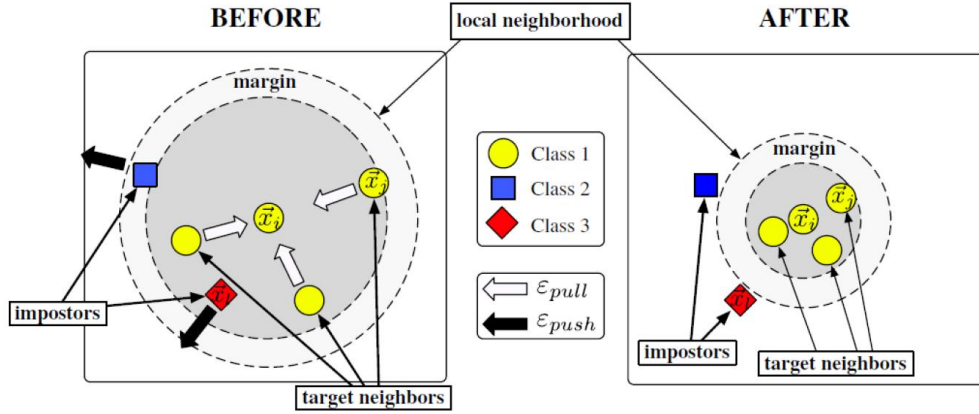


图 2-7 LMNN 算法示意图[78]

深度神经网络进一步拓展了度量学习的应用范围。一种典型的基于神经网络的度量学习框架便是 Siamese 结构[84][85]，Siamese 最初被 Yann LeCun 提出用于手写字符识别[85]和人脸识别[84]。其主要思想在于通过学习一个损失函数将输入的数据模式映射到一个隐藏的语义空间，具体到步态识别问题上，对于相同行人的步态图，其相似性距离变小，对于不同行人的步态图，其相似性距离变大。这种并行的卷积结构十分有利用特征识别与匹配任务。

2.5 本章小结

本章结合国内外研究现状，首先对步态识别基本算法框架进行了概述，然后重点介绍了目前公开的步态数据库和相关的步态识别算法，最后针对本文的研究支撑技术深度学习和距离度量学习的背景知识进行介绍。针对以上对国内外研究现状以及相关技术的调查，我们有以下结论：

(1) 将深度学习技术应用于步态识别领域是我们的研究重点。现阶段，随着计算机视觉领域特征表示方向的不断发展，视觉特征建模逐步走向成熟，尤其是深度学习的发展逐步打破了以往人工特征表示的瓶颈，因此很有必要探索如何构建一套基于深度学习的步态识别技术方案；(2) 在获得深度步态特征的同时如何克服深度学习面临的一系列挑战也是需要研究的主要问题。深度模型的训练往往需要大量的训练数据，训练过程比较耗时，不仅如此，常用的深度模型大多针对图像分类领域，而步态识别问题和分类问题存在巨大的领域鸿沟问题，如何解决以上这些难点，也是本文需要研究的主要问题；(3) 结合步态时序的特征是提

高识别准确率的潜在方向。目前已经有很多研究者意识到时序特征对于步态识别的重要性,然而这些方法往往计算比较复杂,比如光流法、动态时间匹配等方法。如何将步态的时域空域特征进行融合,并且充分利用深度模型的优势,也是本文需要研究的重要问题。

第三章 基于传统卷积神经网络的步态识别技术

3.1 概述

如上所述,目前流行的步态识别流程通常是先对连续的步态帧序列进行预处理,再将一个完整周期的轮廓序列采用某种方式转化为一张步态特征图,对行人身份和属性的判别就是基于该步态特征图的视觉特征,常用的视觉特征包括 SIFT[19]、HOG[18]、LBP[20]、MSER[21]及其它们针对具体问题的拓展版本等等。步态特征图是各种复杂的因素以非线性的方式结合在一起的,它包含了行人的身份、姿态、年龄、行走速度、环境光线等各种信息,传统的单一人工特征或者是特征融合表示方法都很难处理这种复杂的步态特征图。深度学习的关键就是通过多层非线性映射将这些因素成功的分开,例如在卷积神经网络的最后一个隐含层,不同的神经元代表了不同的因素。如果将这个隐含层当作特征表示,身份识别、年龄估计、性别分类就会变得非常简单,因为各个因素之间变成了简单的线性关系,不再彼此干扰。考虑到深度特征的优势以及现在步态特征的局限性,非常有必要在步态识别领域运用目前先进的深度卷积神经网络模型。然而,对步态数据提取深度特征的过程中面临着三个挑战:1)步态库数据样本量较小。深度学习需要依赖于庞大的训练数据,小样本数据应用深度学习模型是目前机器学习领域的研究热点。即使是目前最庞大的 OU-ISIR LP[47]数据库也只有 4000 人的规模,每个行人只有 4 段步态序列。如何将现有数据最大化利用,是第一个挑战;2)模型训练速度慢。深度卷积神经网络在对目标训练集进行反向传播训练时很耗时,如何利用现有的模型并将其应用到步态识别任务是第二个挑战;3)如何设计步态特征提取和匹配算法框架。现有的深度模型主要针对大规模的图像分类任务,如何将其迁移到步态识别任务,并设计一个完整的算法框架也是一个难点。

为了解决以上问题,我们设计了一种针对步态识别问题的深度学习解决方案。如图 3-1 所示为本文提出的步态识别技术流程图。首先,为了解决数据量不足的问题,我们将原始的视频帧序列进行融合,计算其步态能量图(Gait Energy Image, GEI) [17]作为深度神经网络的输入。然后,将 ImageNet 上训练得到的模型作为起点,利用步态训练数据集和反向传播对其进行 fine-tuning 训练,将模型适应到步态识别任务。接下来,将训练好的模型作为特征提取器,将其隐含层的输出作为步态特征表达,代替常用的人工特征。最后,利用 K 近邻[86]算法计算测试步态序列与数据库中待识别步态序列的距离,根据相似度最大时匹配库中的优化步

态特征对应的已经人物信息，确定所述待识别人的身份信息。

数据采集

预处理

周期性检测

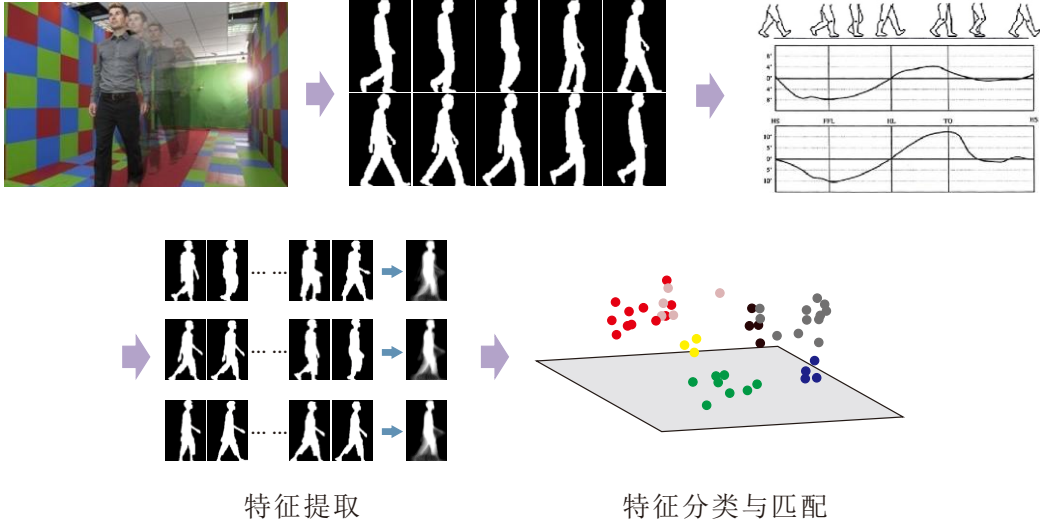


图 3-1 本文提出的步态识别技术流程图

3.2 步态能量图的生成

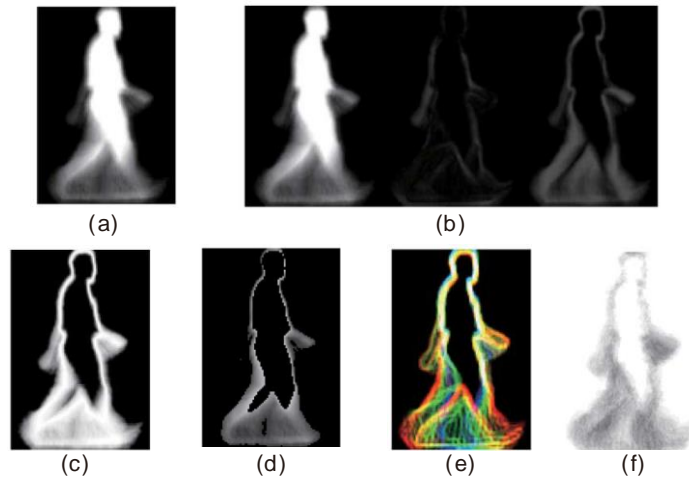


图 3-2 几种不同的步态特征图合成结果：(a) 步态能量图，(b) 步态频繁域图，(c) 步态熵图，(d) 基于步态熵图的步态掩模，(e) 步态时序图，(f) 步态流图。

目前步态分析的特征表示方式主要依赖于人体外观与周期性运动模式，比如步态平均轮廓图，也称之为步态能量图[17]。其他广泛使用的步态表示方法包括：步态熵图（Gait Entropy Image, GEnI）[87]，步态频繁域图（Frequency-Domain Feature, FDF）[88]，基于步态熵图的步态能量掩模（Masked GEI based on GEnI, MGEI）[89]，步态时序图（Chrono-Gait Image, CGI）[90]以及步态流图（Gait Flow Image, GFI）[34]等等。

本文中，我们采用步态能量图作为深度神经网络的训练输入，基于以下几点原因：首先，步态能量图的计算十分简单，只需要将一个步态周期内的序列进行

平均；其次，步态能量图能够将行人的运动序列用一幅特征图进行表示，同时最大化保留了运动的时序信息；最为重要的是步态能量图对于单帧的轮廓分割不佳带来的噪声具有很强的鲁棒性。因此，在基于传统卷积神经网络的步态识别方法中，步态能量图非常适合作为卷积神经网络的训练输入。

具体来说，我们利用目前最好的图像分割方法对原始的静态帧序列进行前景分割得到人体步态轮廓图，然后通过缩放归一化到统一的尺度，方便接下来的神经网络训练。对于一个给定的已经进行长宽归一化后的步态轮廓图 $I(x, y, t)$ ，该步态序列的二值灰度步态轮廓图 $G(x, y)$ 可以定义为：

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N I(x, y, t) \quad (3-1)$$

其中 N 为当前步态周期的总帧数， t 为帧号， x 和 y 为步态轮廓图的二维像素点的坐标。图 3-3 为部分步态能量图实例。



图 3-3 步态序列分割结果（每隔四帧）及其合成的步态能量图

3.3 基于深度卷积神经网络的特征训练与提取

3.3.1 模型训练

深度卷积神经网络能够学习针对步态能量图的分层特征表达。最底层可以从原始像素学习滤波器，刻画图像局部的边缘和纹理特征，通过对各种边缘滤波器进行组合，中层滤波器可以描述不同形态步态能量图的局部轮廓，比如腿部、手

臂、髋部和头部等，最高层描述的是整体步态能量图的全局特征。如果是对于基于步态的性别分类任务，我们只需要直接利用现有的卷积神经网络模型对步态能量图进行二分类训练即可。但是对于行人的身份的识别，我们需要训练通用基础识别模型，然后利用其作为特征提取器，把最高的隐含层的输出作为特征表达，代替常用的手工设计的特征。

如图 3-4 所示，我们采用参数微调的方式对现有的卷积神经网络模型进行重新训练，然后将其作为特征提取器来抽取特征待识别的步态序列以及数据库中的步态序列的特征。这样做有两个优势：1) 卷积神经网络模型能够从简单的抽象输入图像中自动提取出具有区分性的特征而不需要了解领域知识。2) 参数微调能够在现有的已经训练好的模型上进行重新参数微调能够克服数据量不足以及随机参数训练可能导致的新模型不收敛问题，同时加快训练速度。

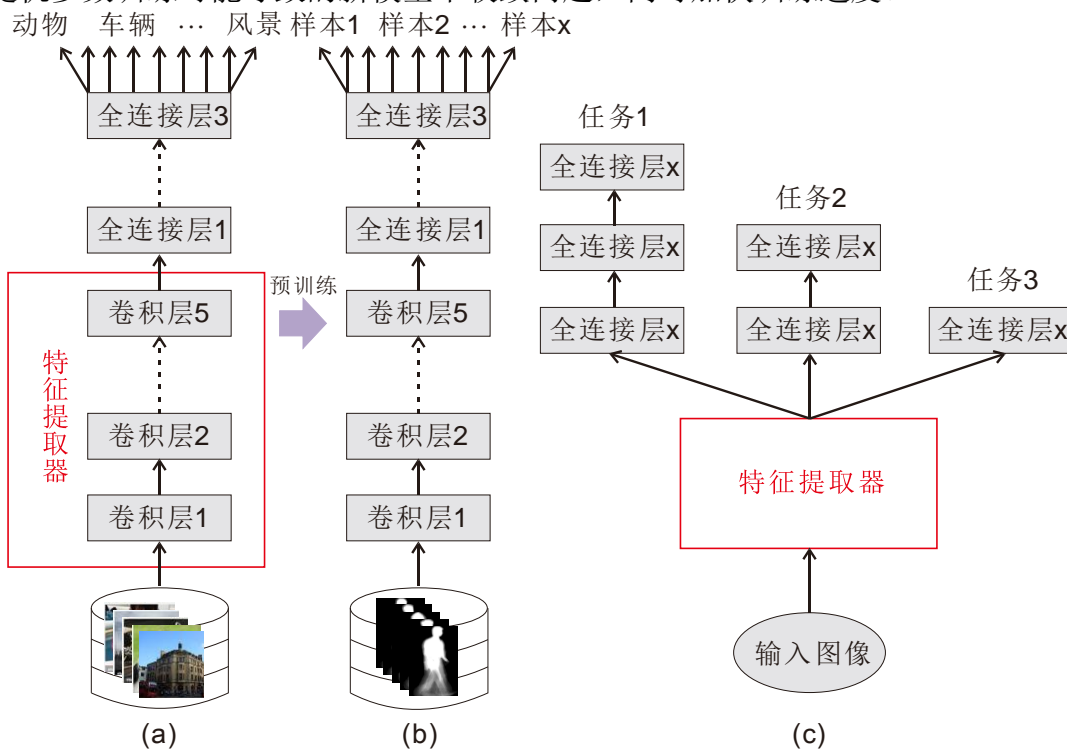


图 3-4 基于深度学习的模型训练与层次化特征提取框架。(a) 预训练的通用基础大类识别模型，(b) 经过参数微调后领域专家模型，(c) 特征提取模型。

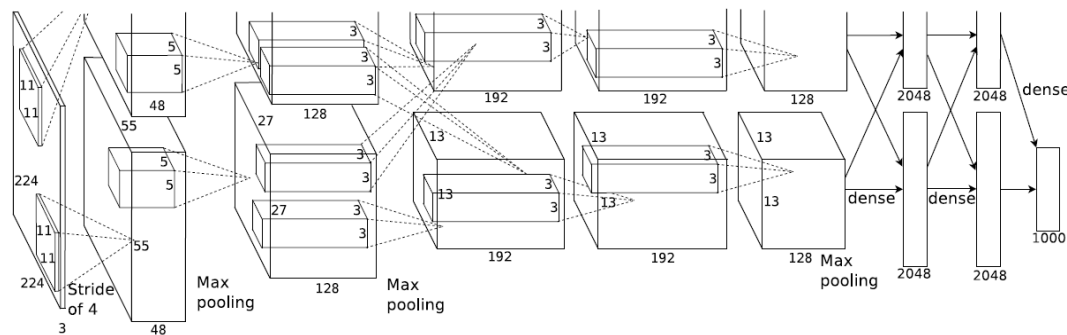


图 3-5 AlexNet[91]网络结果及参数示意图

在本文中,我们使用由多伦多大学的 Alex Krizhevsky 设计的 AlexNet 网络[91] 进行训练,实现采用了美国加州大学伯克利分校提出的 Caffe 框架[25]。该网络模型是在“ILSVRC-2012”数据集上训练得到的,并且获得了 1000 类分类问题的第一名。原始的 CNN 网络包含两个部分:1) 输入层、网络卷积层 (Convolution Layers) 和最大池化层 (Max-pooling Layers); 2) 两个全连接层 (Fully Connection Layers) “FC1”、“FC2”, 和最终输出图像在 1000 个类别上概率分布的输出层 (Output Layers)。这里,为了将步态特征图的信息映射到语义空间,在性别分类任务中,我们将输出层的神经元个数设为 2,在行人身份识别任务中,我们将输出层的 1000 个类别替换为与数据集中行人数量相关联的行人身份数量。在模型训练之前,我们首先使用文献[91]中训练好的神经网络参数对本模型进行初始化。在训练过程中,由于现有的步态数据集相对 ImageNet 数据库来说规模较小,小数据训练时卷积层的网络参数是最难更新的,又因为从 ImageNet 学习得到的底层滤波器往往描述了各种不同图像的局部边缘和纹理信息,这些滤波器对一般的图像有较好的普适性。所以,我们的方法是固定低层的卷积层参数,优调全连接层参数,只对上层网络进行参数更新,然后使用随机梯度下降法对整个 CNN 模型进行训练,当模型逐渐稳定后,再将损失函数的值后验传播到其他所有网络层,对整个网络模型进行微调,直至模型收敛。

3.3.2 特征与匹配

在获得已经训练好的模型参数之后,对于每一张步态能量图,都可以输入网络进行前向传播,抽取最后三层全连接层 (FC1,FC2,FC3) 的输出作为特征表示,实验表明 AlexNet 网络中倒数第二个全连接层的特征表达最佳[92],因此在下面的实验中,我们选择倒数第二层的 4096 维特征作为最终的步态特征。

最后,我们利用 K 近邻算法[86]计算测试步态序列与数据库中待识别步态序列的欧式距离,取出匹配库中与测试样本距离最近的步态特征对应的已知人物信息,确定所述待识别人的身份信息。

3.4 实验结果与分析

实验中,我们在性别分类和身份识别两个任务上验证了基于传统卷积神经网络的步态识别方法。性别分类的目的是利用数据库中已知性别的样本特征估计出待测样本的性别标签。如表 3-1 所示,属性分类和身份识别两个问题从研究目标和问题建模方法都有所不同。性别分类是一个二分类问题,传统的卷积神经网络模型已经证明可以较好地解决分类问题,而身份识别则是一个识别问题,需要进

行特征的验证与匹配。

表 3-1 属性识别与身份识别比较

	属性识别	身份识别
研究目标	粗分类，例如性别、年龄、种族、体型等。	精确识别，中远距离、人数较少等场景中对特定目标的查找和验证。
问题模型	二分类或者类别较少	识别与验证问题
数据获取	数据广泛、统计建模	特定身份行人的多视角数据
解决方案	运动检测，特征提取，运动模式分析等	
挑战	类内差异大，类间差异小	

3.4.1 评测数据集

首先介绍本节实验中使用的数据库。CASIA-B[15]数据库是一个大规模的、多视角的步态库，采集于 2005 年 1 月，也是目前最为常用的步态识别数据库，大量算法在该数据库上进行评估。该数据库采集了 93 名男性和 31 名女生的步态数据，每个人有 11 个视角（0°，18°，36°，……，180°），在三种行走条件下采集，其中普通条件步态序列 6 段，穿大衣步态序列 2 段，携带包裹步态序列 2 段。CASIA-B 是目前公开的最大规模带有样本性别标签的步态数据库，研究领域大多数步态属性分类算法都在该数据库上进行评估。因此，我们在此数据集上进行了性别分类实验。

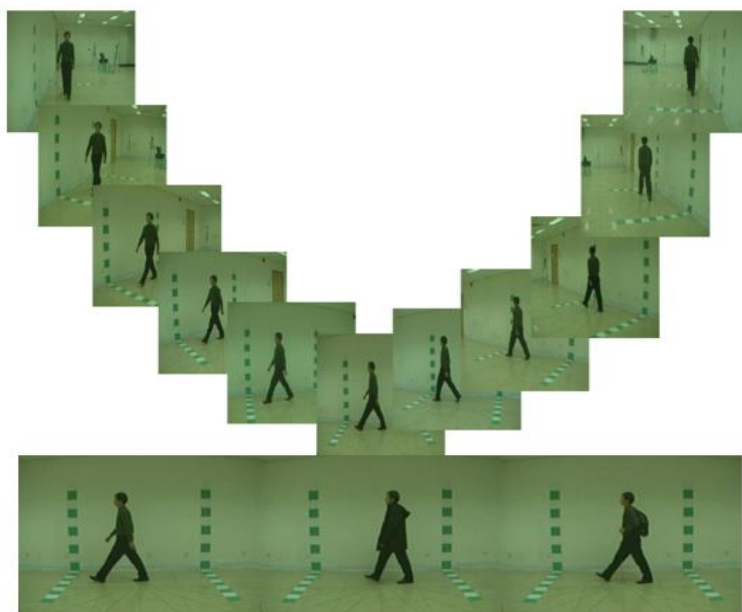


图 3-6 CASIA-B 数据库[15]



图 3-7 OU-ISIR Large Population 数据库[47]

接下来,我们选择在 OU-ISIR LP 中的 OULP-C1V1-A[47]数据库上的进行了身份识别实验。OU-ISIR LP 是目前世界上最大的步态识别数据库,是目前应用最广的步态数据库,因此对于身份识别任务也最具挑战性。OU-ISIR LP 共采集了 2032 名男性和 1803 名女性的步态数据,一共分为 4 个观测视角 (55° , 65° , 75° , 85°) 以及全局的观测视角 All,与此同时,数据库给出了标准的查询(Probe)和源数据(Gallery)目录,非常适合于各种算法进行效果比较。

3.4.2 基于步态特征的性别分类

考虑到性别分类是一个二分类问题,因此应对同等数量的男性和女性进行性别分类实验,目前大部分论文提出的方法都是选择行人 90° 视角下的普通行走状态步态数据进行性别分类实验。为了方便公平公正地评估本章提出的算法和其他算法的性能,我们严格按照[93]中的实验设置开始评测。在[93]中,作者在 CASIA-B 中随机选取了的 31 名男性和所有的 31 名女性在侧面 90° 视角下普通行走状态步态数据,并用交叉校验的结果作为正确分类率。每次交叉校验的具体过程是,随机选取未测试过的一段男性步态序列和一段女性步态序列作为测试数据,其他 60 段样本用来训练各自提出的模型。如此重复 31 次后,所有样本都被测试一次。最后,用这 62 段样本的正确分类率对分类算法进行评价。

表 3-2 在 CASIA-B 数据库上的 90° 度视角下性别分类方法的准确率

方法	准确率
Li[94]	93.28%
Yu[93]	95.97%
Hu[95]	96.77%
[96]	98.39%
GEI+CNN	100.00%

表 3-2 为提出的方法与其他方法的正确分类率结果。可以发现,基于深度卷积神经网络的方法获得 100% 的性别分类准确率,高于现有的其他方法正确分类率。这是由于 90° 视角是最能识别一个人步态特征的角度,同时,深度网络在图

像分类任务上具有显著的优势。

为了进一步验证提出方法的性能,我们在其他视角的步态序列上做了同样设置的实验,图 3-8 为在 CASIA-B 数据全部 11 个视角下性别分类的正确率,对比方法我们选择了目前性别分类效果最佳的基于混合条件随机场模型的性别分类方法[96]。可以看出,相比于文献[96]的方法,基于 GEI+CNN 的方法在全部 11 个视角下都取得了较高的正确分类率,其在各个视角下的平均分类精度达到 90%,值得一提的是,在 0°和 180°视角这两个较难的视角下,我们提出的方法依然具有很好的效果。

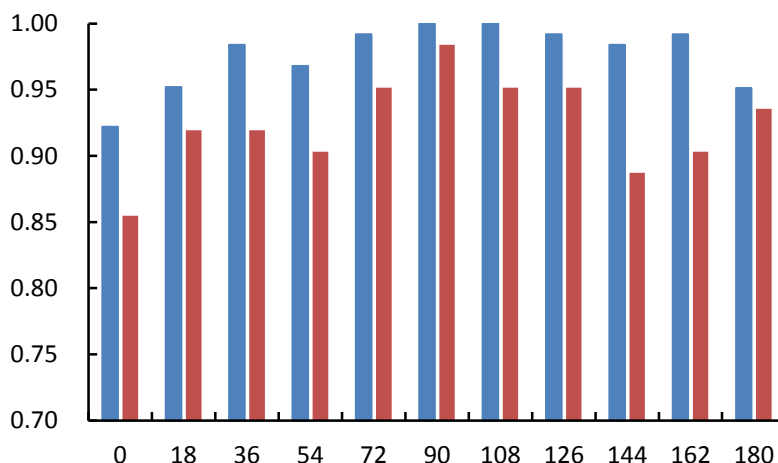


图 3-8 在 CASIA-B 数据库上不同视角下性别分类准确率, 每一角度下左侧柱为提出的方法, 右侧柱为文献[94]的方法

3.4.3 基于步态特征的身份识别

接下来,我们在 OU-ISIR LP 数据库中进行了身份识别的实验,表 3-3 为该数据库在各个视角下的样本数量统计。在实验中,我们只使用 Gallery 库进行训练,因此没有 Probe 数据会用于训练。在接下来的实验中,我们使用与[47]中相同的测试条件,在相同视角和不同视角下评估了我们算法的效果。在这里,我们使用 Rank-1 和 Rank-5 的识别准确率作为评价指标。

表 3-3 OU-ISIR LP 数据库样本数量统计

视角	55°	65°	75°	86°	All	总计
行人数量	3,706	3,770	3,751	3,249	3,141	3,835
男性/女性	1,977/1,729	2,007/1,763	1,995/1,756	1,688/1,561	1,644/1,497	2,032/1,803

在表 3-4 和表 3-5 分别给出了本文提出的方法在 OU-ISIR LP 数据库中 Rank-1 和 Rank-5 的识别准确率,其中行头显示的是查询视角,列头显示的是匹配库的

视角，对角线中的数值表示在同一视角下的身份识别准确率，非对角线部分为跨视角下的身份识别准确率。对角线上括号中的数值显示了目前在数据库上效果最好的 GEI 匹配方法[47]的识别准确率，注意，GEI 匹配方法的方法只在相同视角下评估了识别准确率。通过观察我们发现本章提出的方法相比目前最优的方法在识别率上有 3%~10% 的差距，不仅如此，在跨视角的步态识别情况下，识别的准确率不尽人意，这说明跨角度识别问题本来具备一定难度，不同视角下的行人在行走姿态、外形轮廓上差异较大，从实验结果来看，角度差别越大，识别难度也就越大。

表 3-4 在 OU-ISIR LP 数据库上身份识别 Rank-1 正确率

视角/数据	Gallery			
	55°	65°	75°	85°
Probe	55°	73.96 (84.7)	68.94	57.42
	65°	67.19	76.72 (86.63)	72.65
	75°	53.82	69.79	77.87 (86.91)
	85°	40.07	56.76	69.13
				78.82 (85.72)

表 3-5 在 OU-ISIR LP 数据库上身份识别 Rank-5 正确率

视角/数据	Gallery			
	55	65	75	85
Probe	55	86.64 (92.39)	85.35	79.09
	65	83.26	88.67 (92.84)	86.87
	75	75.47	85.55	89.39 (92.78)
	85	64.20	77.62	85.41
				90.09 (93.01)

经过前文的理论分析和本节的实验观察，我们分析实验结果不佳的原因在于以下两点：1) 步态库训练样本太小。本节提出的技术框架并为完全解决识别任务面临的挑战，深度学习需要依赖于庞大的训练数据，目前用于对模型训练的数据量远远小于 ImageNet 分类任务的图片量，并且步态图像与 ImageNet 图像库存在领域鸿沟，二者共同导致了新的模型并不能完全适应步态识别任务；2) 分类模型和识别问题存在领域鸿沟。大部分的可以被用于预训练的深度模型都是来源于分类任务，而识别任务与分类任务的具有本质的区别。为了解决以上问题，我们将在下一章中介绍如何使用度量学习模型进行特征的相似度训练和匹配。

3.5 本章小结

在本章中，针对如何将深度卷积神经网络应用在步态识别任务的问题，我们研究设计了一套针对步态识别问题的深度学习解决方案。根据步态序列的特性，我们首先将原始的视频帧序列进行融合，计算其步态能量图作为深度神经网络的

输入。然后以已经训练调优的深度模型作为起点，对其进行参数微调来适用于步态识别问题，将其作为特征提取器。在特征提取过程中，将模型的隐含层的输出作为步态特征表达，代替常用的手工设计特征。最后，利用 K 近邻算法计算测试步态序列与数据库中待识别步态序列的距离，根据相似度最大时匹配库中的优化步态特征对应的已知人物信息，确定所述待识别人的身份信息。

实验表明，深度神经网络强大的特征表征能力能够很好地解决步态识别中的性别分类问题，但是由于步态数据库规模较小以及分类问题和识别问题之间的领域鸿沟，本章提出的计算框架并没有很好地解决身份识别问题。下一章中我们将介绍如何使用度量学习模型来进行解决基于步态的身份识别问题。

第四章 步态识别中的距离度量学习模型

4.1 概述

从上一章的实验中，我们可以发现虽然将深度卷积神经网络在性别分类任务中取得了目前最佳的分类准确率，但是直接将其应用于身份识别问题时还是面临很多挑战。针对这些挑战，我们引入距离度量学习模型来解决训练数据量不足的问题，同时克服了分类和识别的领域鸿沟。如图 4-1 所示为我们提出的基于 Siamese 神经网络的步态识别整体框架图。Siamese 结构由两个相同的并行的卷积神经网络构成，网络的每一层之间共享权值，只需要输入相同行人的图像对（正例）和不同行人的图像对（负例）进行训练，其主要思想在于通过学习一个损失函数将输入的数据模式映射到一个隐藏的语义空间，具体到步态识别问题上，对于相同行人的步态图，其相似性距离变小，对于不同行人的步态图，其相似性距离变大。这种并行的卷积结构十分有利用特征识别与匹配任务。

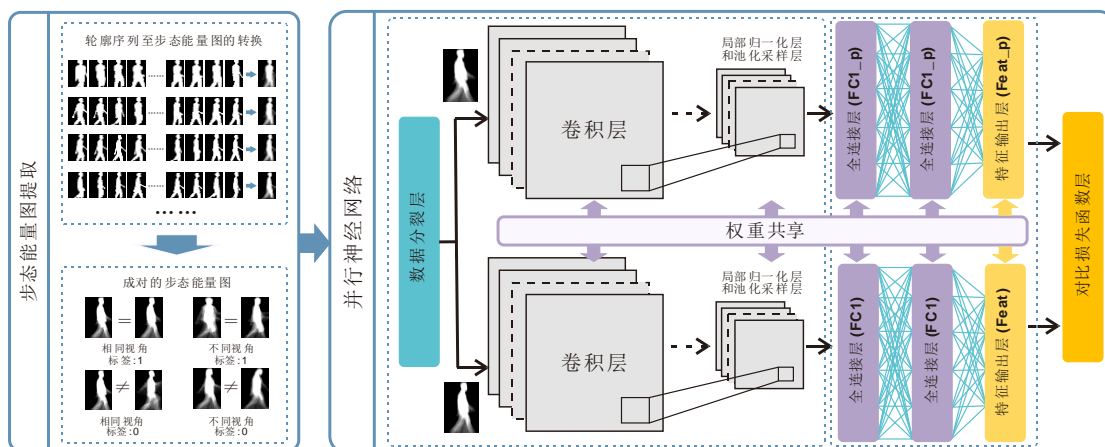


图 4-1 基于 Siamese 神经网络的步态识别技术整体框架图

具体地，整个流程分为以下几个步骤：首先，将原始的步态序列合成为步态能量图（GEI）作为神经网络的输入，虽然这在一定程度上损失了步态的时序和周期性信息，但是可以很好地解决步态图数据量不足的问题。其次，在训练阶段，我们将两张步态能量图同时输入 Siamese 网络，经过数据切分，两张图像分别进入 Siamese 网络的两个分支，这两个分支参数权值共享，每一个分支由相同的卷积神经网络构成，它包含两个部分：1) 两个网格卷积层（Convolution Layers）和最大池化层（Max-pooling Layers）；2) 两个全连接层（Fully Connection Layers）“FC1”、“FC2”，和最终的维度为 2 的特征输出层（Output Layers）。然后，两条 CNN 的输出汇聚到对比损失层（Contrastive Loss Layer），通过对比损失函数来

计算目标函数的对比损失值。最后，我们利用反向回溯算法对整体模型的参数进行训练。在测试阶段，对于每一个步态帧序列，我们首先合成其对应的 GEI 图。然后只使用一个 CNN 分支来抽取步态特征图的特征“SiaNet.FC”。最终使用 K 近邻[48]的方法在整个步态数据库中查询相同身份的行人。

4.2 基于 Siamese 神经网络的步态识别技术

4.2.1 网络结构

目前大多数深度学习模型主要应用于大规模的目标分类问题，物体分类任务要求回答一张图像中是否包含某种物体，然而，从模式分类的角度来看，步态识别问题与分类问题存在巨大的领域鸿沟，这就导致了直接应用传统的分类卷积网络到步态识别问题往往得不到理想的识别结果。Siamese 结构最初由 Yann LeCun 提出用于手写字符识别[85]和人脸识别[84]。Siamese 结构是并行的神经网络结构，它由两个完全相同的并行卷积神经网络构成，网络的每一层之间共享权值，只需要输入相同行人的图像序列（正样本）和不同行人的图像序列（负样本）进行训练，其主要思想在于通过学习一个损失函数将输入的数据模式映射到一个隐藏的语义空间，具体到步态识别问题上，对于相同行人的步态图，其相似性距离变小，对于不同行人的步态图，其相似性距离变大。因此，这种并行的卷积结构十分有利于特征的识别与匹配任务。

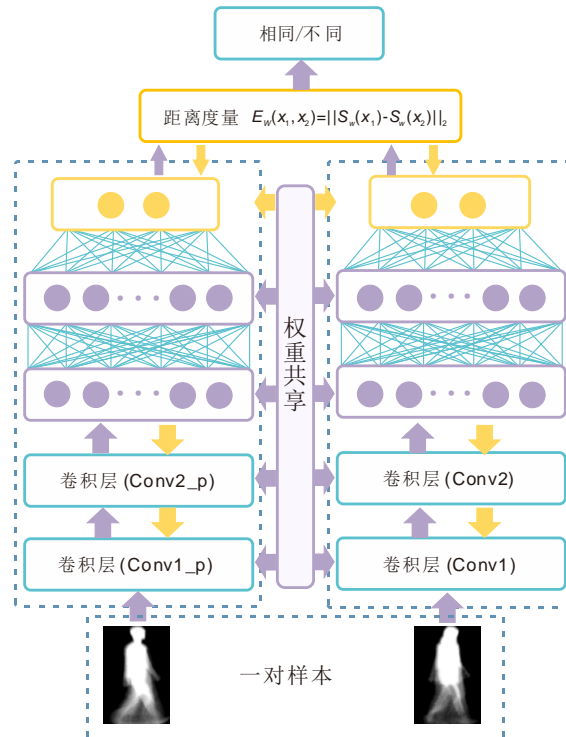


图 4-2 本文设计的 Siamese 网络结构示意图

本文构造的 Siamese 结构如图 4-2 所示，每一个分支由两部分构成 1) 两个网格卷积层和最大池化层；2) 两个全连接层和最终的维度为 2 的特征输出层。每一条网络分支的具体参数设置如下，扫描步长指卷积核每次在图像中运动的像素大小。

表 4-1 模型参数表

网络名称	卷积核尺度	输出维度	扫描步长
卷积层 1	5×5	20	1
最大池化层 1	2×2	—	2
卷积层 2	5×5	50	1
最大池化层 2	2×2	—	2
全连接层 1	—	500	—
全连接层 2	—	10	—
特征输出层	—	2	—

4.2.2 训练和测试策略

在实验中我们选择了 OU-ISIR LP 数据库进行算法评测，因此我们只使用 OU-ISIR Large Population 的 Gallery 库进行 Siamese 网络训练。根据 Siamese 网络的基本原理，需要构建满足训练条件的正负样训练结合。首先将 Gallery 中每个样本预先计算好的步态能量图归一化为 128×171 像素，以方便作为神经网络的输入。然后，从步态数据库中选择正负样本构建神经网络训练数据集，其中身份相同的两段步态序列为正样本对，同时设置标签为 1；身份不同的两段步态序列为负样本对，同时设置标签为 0。具体地，在本次实验中，训练集来源于 3835 人的步态轮廓序列图，由于行人样本数量远远大于同一身份行人步态视频序列数，训练集中正样本数量远远少于负样本数量。因此，为保证神经网络训练过程中样本平衡，我们抽取所有的 20000 对正样本，并且随机抽取了等量的负样本作为最终的训练集。

训练过程由以下四步构成：

(1) 将训练数据对同时输入时空联合深度神经网络，经过数据切分层，两张步态能量图分别进入网络的两个分支，这两个分支参数权值共享，每一个分支由相同的卷积神经网络构成。

(2) 每个分支中的输入图像进行前向操作，两条卷积神经网的特征输出层汇聚到对比损失层，通过对比损失函数来计算目标函数的对比损失值。对比损失函数 $\mathcal{L}(W)$ 如下：

$$\mathcal{L}(W) = \sum_{i=1}^P L(W, (y, x_1, x_2)^i) \quad (4-1)$$

$$L(W, (y, x_1, x_2)^i) = (1 - y) \cdot \max(m - E_W(x_1, x_2)^i, 0) + y \cdot E_W(x_1, x_2)^i \quad (4-2)$$

$$E_W(x_1, x_2) = \|S_W(x_1) - S_W(x_2)\|_2^2 \quad (4-3)$$

其中, x_1 和 x_2 表示输入的一对样本对, $(y, x_1, x_2)^i$ 表示训练数据库中的第 i 个样本, $L(W, (y, x_1, x_2)^i)$ 表示每对样本的对比损失值。当第 i 个样本对属于正样本, 即表示相同身份的行人时, y 为 1; 当第 i 个样本对属于负样本, 即表示不同身份的行人时, y 为 0。 $E_W(x_1, x_2)$ 为该样本对的距离, W 为模型参数。我们可以通过 W 将样本 x_1 和 x_2 映射至 $S_W(x_1)$ 和 $S_W(x_2)$ 。 P 为训练集中的样本对的个数, m 为设定的大于零的边界阈值。

(3) 利用反向回溯算法对整体模型的参数进行训练。

(4) 重复 (2) 和 (3) 步骤, 直至所有训练数据上模型收敛。

测试识别过程中, 首先利用训练好的模型对测试步态能量图和匹配库中所有步态能量图提取特征, 然后使用 K 近邻的方法逐一计算测试样本与匹配样本的特征相似度, 输出相似度最高的匹配序列, 从而识别相同身份的行人。具体步骤如下:

(1) 对测试步态序列计算步态能量图, 并归一化到相同尺寸。

(2) 将 (1) 中得到的步态能量图输入至 Siamese 网络的任一分支, 采用前向传播算法提取特征。具体地, 将其输入至图 4-2 所示的卷积神经网络, 网络的初始化参数采用已经训练好的模型参数, 提取第二个全连接层的输出作为步态能量图的特征, 特征维度为 500 维。因此, 每一段步态视频均可以由一段 500 维的特征向量表示。

(3) 对匹配库中所有步态视频提取步态轮廓序列图, 计算步态能量图, 并归一化到相同尺寸, 依据 (2) 中的方法, 将其输入 Siamese 神经网络的一个分支, 提取所有匹配库中的步态特征。

(4) 使用 K 近邻的方法计算测试序列与匹配序列的相似度, 输出相似度最高的结果, 从而识别相同身份的行人。

4.3 实验结果与分析

本章中, 我们在 OU-ISIR LP 数据库进行了身份识别的实验, 如上文所述, 在实验中, 我们只使用 Gallery 库进行训练, 因此没有 Probe 数据会用于训练。该训练集的包含了每个行人样本四个观测视角 (55°, 65°, 75°, 85°) 以及全局的观测视角 All 的步态序列。我们在相同视角和不同视角下评估了提出算法的

效果。同样地，我们使用 Rank-1 和 Rank-5 的正确识别率作为评价指标。

4.3.1 相同视角下的身份识别

首先，我们比较了本文提出的方法在相同视角的步态识别准确率。下表 4-2 和 4-3 比较了我们的方法与目前最优方法的对比结果，其他对比方法包括：GEI[47] 和 FDF[47]模型匹配方法，HWLD[98]，以及本文第三章提出的基于传统 CNN 的方法。请注意，在文献[98]中，HWLD 方法仅仅测试了在 85° 和 ALL 两个数据集上的准确率。正如表中所示，我们提出的 SiaNet.FC 方法在几乎所有的数据集上都取得了最高的识别准确率。与 GEI, FDF 以及 HWLD 这些基于人工特征的方法相比，基于深度学习的方法可以自动地从 GEI 中提取具有区分性且表现丰富的步态特征，比如在人的头部、髋部等语义模糊的局部区域，卷积特征依然具有很好的区分性。相比于传统的 CNN 方法，本文提出的基于 Siamese 结构的方法能够利用距离度量学习很好地解决分类问题与识别问题的领域鸿沟，因此可以得到更高的步态识别准确率。

表 4-2 在 OU-ISIR LP 数据库上 Rank-1 识别准确率

	55°	65°	75°	85°	All
HWLD[98]	—	—	—	87.70	95.50
GEI[47]	84.70	86.63	86.91	85.72	94.24
FDF[47]	83.89	85.49	86.59	85.90	94.17
CNN.FC2	73.96	76.71	77.87	78.82	86.09
SiaNet.FC	90.12	91.14	91.18	90.43	96.02

表 4-3 在 OU-ISIR LP 数据库上 Rank-5 识别准确率

	55°	65°	75°	85°	All
HWLD[98]	—	—	—	94.70	98.50
GEI[47]	92.39	92.84	92.78	93.01	97.13
FDF[47]	91.53	92.81	92.88	92.83	97.10
CNN.FC2	86.64	88.67	89.39	90.09	93.56
SiaNet.FC	94.98	95.90	95.92	95.97	98.31

在图 4-3 中，我们还展示了部分识别结果，其中第一列为待识别的步态特征图，其后依次为 Rank 5 的识别结果，红色实现方框为正确识别结果。从中我们可以观察到，不同行人之间的从外观轮廓来看极其相似，肉眼难以分辨，尤其在头部、躯干、腿部等区域更加容易混淆。同时，在步态识别中类间差异太小会导致识别困难加大，比如在 55-55 组中，第一匹配样本与待测步态图在躯干部分更

加相似，而真实匹配的行人由于头部姿势的轻微改变导致 Top 1 结果识别错误。在真实的行走序列之中，每个人的步速、躯干和手部的摆幅、步子大小都有所不同，这是由于步态除了空间特征之外，还具有时序特征，周期性是步态识别问题相比于一般动作识别问题最大的不同，因此我们有待于改进仅仅依赖步态能量图这种特征的匹配方式。

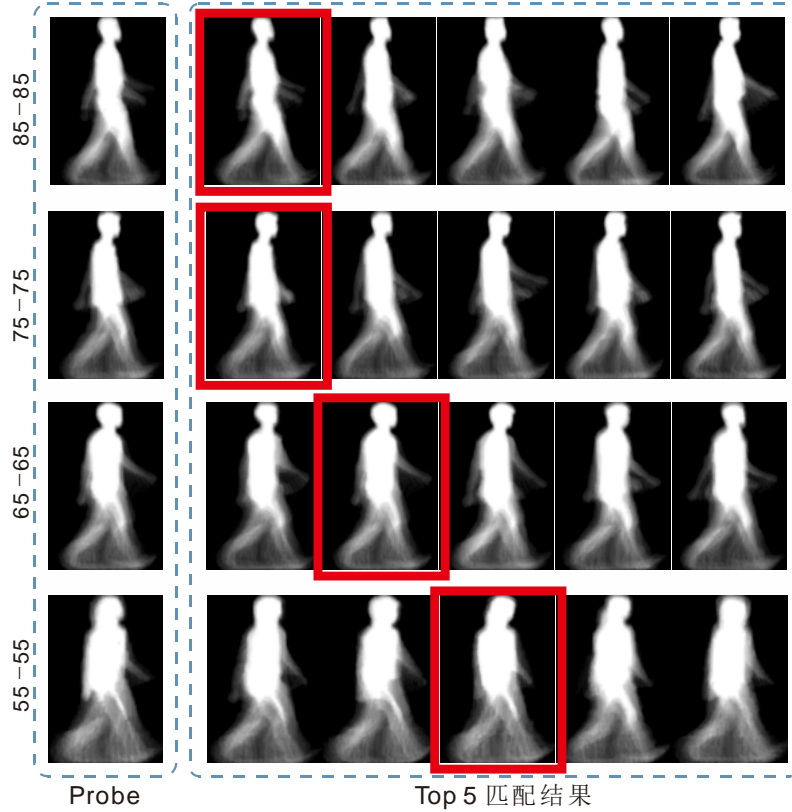


图 4-3 基于 Siamese 网络的模型训示意图

4.3.2 不同视角下的身份识别

然后，我们同时评估了本文提出的方法在跨视角步态识别问题上的效果。对比方法选择了目前最优的跨视角步态识别方法：AVTM_PdVS[99]，AVTM[99]，woVTM[99]和 RankSVM[100]。其中，前三种方法是专门为跨视角的步态识别问题设计的，它们通过利用一个额外的大规模 3D 步态数据库来训练视角转移模型，来克服不同视角步态特征的差异。RankSVM 也是一个基于距离度量学习的方法，针对诸如多视角、携带物、步速等变量的步态识别问题。需要注意的是，这四种方法都只是随机选择了整个 OU-ISIR Large Population 中的 1912 个行人进行了测试，而我们提出的方法在全部 3835 个行人上都进行了测试，因此实验的难度更大。如图 4-4 所示，在四组跨视角实验中，我们的方法相比于其他方法在 Rank-1 正确识别率方面取得了稳定的提升。这表明基于 Siamese 结构的方法对于跨视角

的步态识别问题也有很强的鲁棒性。根本原因是，在训练阶段，我们同时输入了相同视角和不同视角的步态图对进行距离度量训练，这种策略使得我们的方法能够很好地适用于跨视角的步态识别问题。

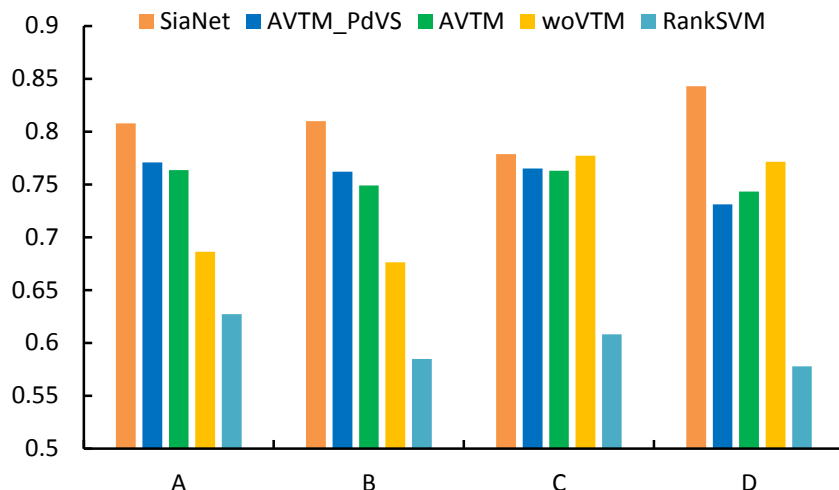


图 4-4 不同视角下步态识别准确率。其中 A~D 组实验分别代表(65°,75°)、(75°,65°)、(75°,85°)和(85°,75°)实验设置

4.4 本章小结

在本章中，我们提出了一种基于 Siamese 神经网络的步态识别方法，其优势在于能够克服分类问题与识别问题的领域鸿沟，同时弥补了步态数据库规模不够的缺憾。整个网络由两条并行的卷积神经网络构成，网络之间参数权值共享，并通过对比损失函数来优调模型参数，最后采用 K 近邻算法计算步态特征相似度。本方法能够克服目前深度学习应用于步态识别问题两大挑战，因此，能够提高步态识别的准确率。

实验表明，无论是在相同视角下的步态识别还是跨视角的步态识别问题上，我们提出的方法在相同的实验环境下，取得比现有方法更好的识别率。然而，从一些匹配失败的例子中我们也可以发现，仅仅依赖于二维的步态特征图进行匹配并没有充分发挥步态周期性的优势，因此在接下来的研究中很有必要利用步态的三维时序特征，从中提取出更具区分度的步态特征来进一步提高识别的准确率。

第五章 基于时空联合深度神经网络的步态识别技术

5.1 概述

步态识别问题从本质上来说是对一段连续的行人行走序列进行特征提取与模式分类的过程。从相关文献调研来看,传统的步态识别方法的思路通常是将一个周期的步态序列合成一张步态特征图,再基于该特征图的视觉特征进行属性分类或者身份识别的工作,然而特征图的计算过程中往往伴随有步态时序信息的损失,输入信息的不准确可能导致输出结果的不准确,从而不能较有效地提高识别的准确性。步态作为一种运动模式,具有空间和时间两个维度的特征,周期性是步态识别与传统的人脸识别或者动作识别最为显著的差异,如何充分利用最为原始的步态序列图成为另一个提高步态识别准确率的瓶颈。三维卷积神经网络[104]

(3D-CNN)最初提出用于动作识别,通过从空间和时间的维度提取特征,然后进行 3D 卷积[103] (C3D),以捕捉从多个连续帧得到的运动信息。在步态序列中使用 C3D 模型,同时结合并行神经网络结构,可以在时间和空间维度获取更有区分性的步态特征。因此,本章提出通过计算步态运动的三维卷积特征来增强模型,该模型可以从连续性的周期序列中捕捉时间和空间信息,进一步提高步态识别的准确率和实用性。除此之外,我们提出了将 C3D 网络与 Siamese 结构相结合的 3D-Siamese 网络,这种时空联合深度神经网络能够在三维空间进行特征度量学习。

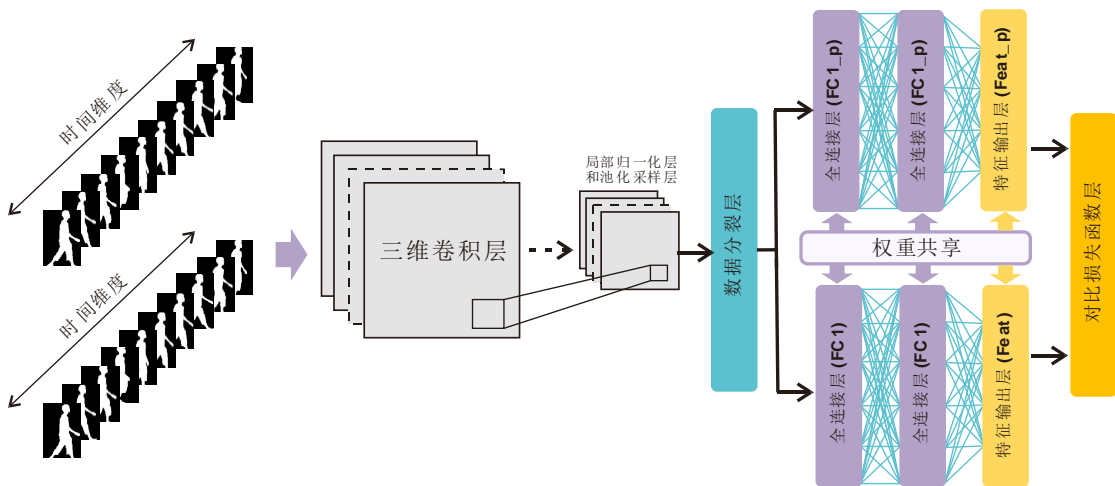


图 5-1 基于 3D-Siamese 网络的模型训示意图

如图 5-1 所示为基于 3D-Siamese 网络的步态识别技术框架图。我们首先在步态数据库的训练视频中提取每一行人样本的步态轮廓序列图,通过构建满足

Siamese 网络的正负样本库来对整个网络进行训练，两条网络中相同的层之间权重共享，损失函数采用对比损失函数。然后，在测试阶段，每一段待识别的样本序列直接输入 Siamese 网络的任一分支，提取该序列的时空深度特征。最后，使用 K 近邻方法逐一计算测试序列与匹配序列之间的相似度，返回的相似度最高的匹配序列即为身份相同的行人样本。

5.2 三维卷积神经网络

5.2.1 基于 CNN 的视频处理方法

最近几年，CNN 在图像任务上取得了很好的效果，研究人员开始尝试在一系列视频处理任务中使用 CNN。最近的在视频任务中取得较好成果的工作包括 [101][102][104]。在文献[101]中，Andrej Karpathy 等人实验了不同的卷积神经网络在视频分类中的结果，从而在 CNN 中表示出视频的时序信息。在一段视频序列中应用 CNN 的一个简单的方法就是对每一帧图像运用 CNN 来识别，但是这种方法并没有考虑到连续帧间的信息，这个问题在步态识别中表现的尤为突出，因为周期性是步态区别于一般的动作识别的最大特征。在文献[102]中，Karen Simonyan 等人用两个独立的卷积神经网络分别处理时间和空间两个维度特征，再将两个模型的结果进行融合。其中空间维度的网络是一条普通的 CNN，而时间维度的网络是将连续几帧的光流叠加起来作为 CNN 的输入。最为重要的是，这个工作第一次引入了多任务学习机制来克服训练数据量不足的问题，具体的做法是将 CNN 的最后一层连接到多个 Softmax 层上，对应不同的数据集，这样就可以在多个数据集上进行多任务学习。

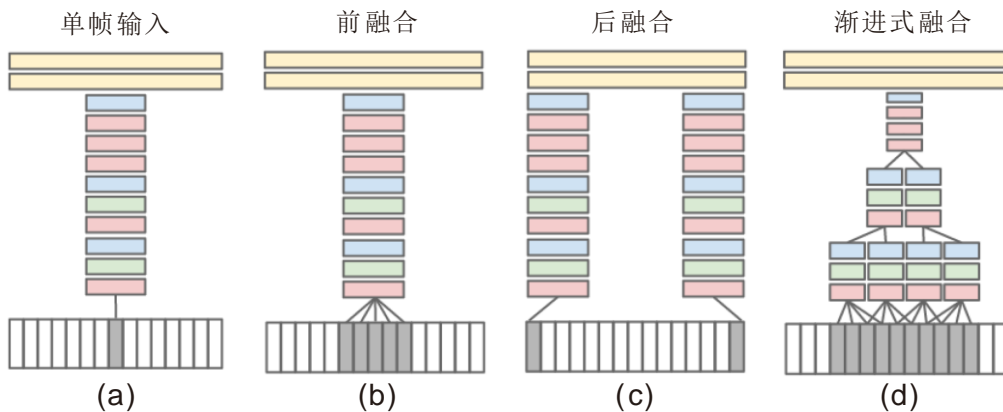


图 5-2 CNN 处理视频的四种不同结构[101]

三维卷积神经网络最初被用与对人的行为分析是在[103][104]中，Shuiwang Ji[104]等人使用三维卷积神经网络来进行动作识别，他们提出的方法首先需要

人体进行检测，通过对头部的跟踪对人体目标进行分割，进而在分割出来的区域进行运动建模，该方法能够同时在时间和空间两个维度进行卷积操作，从而能够捕捉到视频流的运动信息。在文献[103]中，Du Tran 等人提出的 C3D 方法直接对整段视频进行三维卷积操作，而不需要进行目标分割，因此更加具有普适性，他们在目前公开的多个视频数据集上有监督地训练了一个强大的深度模型，能够适应于各种不同的视频分析任务。相比于二维卷积操作或者是在多帧视频上进行二维卷积的操作，C3D 在卷积层和池化层都进行时空性的操作，如图 5-3 所示，二维卷积操作得到的是单张图像，而在多帧图像上进行二维卷积操作得到的仍旧是单张特征图像，因此，只有三维卷积操作才能直接作用与帧立方体上并且得到时空性的特征表达。为了有效地利用步态的运动信息，进一步地提高步态识别的准确率，在本章中使用一种三维卷积的方法，通过将视频帧立方体直接送入 CNNs 的卷积层和池化层进行操作，以捕捉在时间和空间维度都具有区分性的特征。

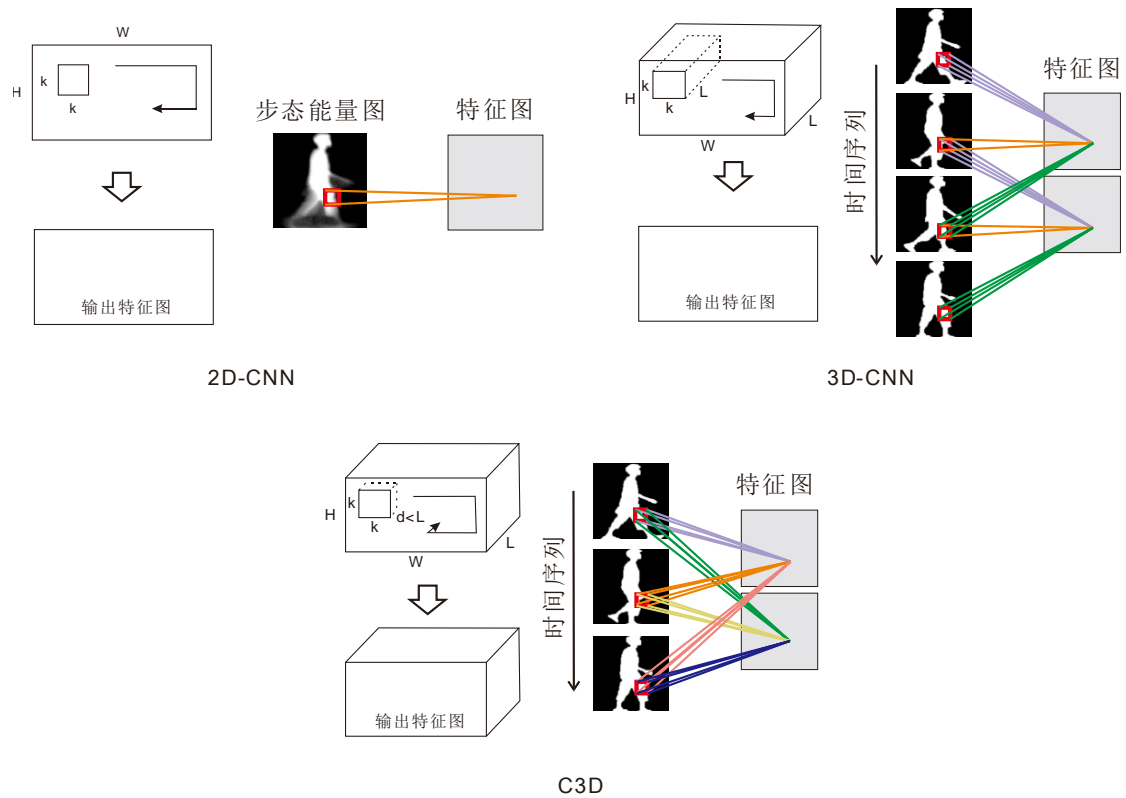


图 5-3 二维卷积和三维卷积特征提取过程示意图

5.2.2 基于三维卷积神经网络的步态序列特征提取

在传统的二维卷积神经网络中，当每一次卷积操作作用于图像中的一个像素点时，都是对其所在特征图的邻域周围其他像素点作求和操作，然后对结果加上

一个偏置后再通过激活函数在神经网络中传导。具体来说，对于二维神经网络中的某一特征图 (x, y) 位置处的特征值 $v(x, y)$ 可以由下式求得：

$$v(x, y) = f(b + \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} C \cdot v(x + w, y + h)) \quad (5-1)$$

其中， $f(\cdot)$ 是激活函数，常用的激活函数包括 Sigmoid 函数和 tanh 函数， W 和 H 分别为二维卷积核的宽度和高度， b 为该层神经网络的偏置项， C 为该层网络参数。在接下来的池化层中，再通过下采样操作将特征图进行降维，整个 CNN 的框架就是通过不断叠加这样的“卷积层+池化层”的结构来进行特征提取。

表 5-1 模型参数表

网络层	卷积核 ($d \times k \times k$)	输出维度	扫描步长
C1	$3 \times 3 \times 3$	64	1
C2	$3 \times 3 \times 3$	128	1
C3	$3 \times 3 \times 3$	256	1
C4	$3 \times 3 \times 3$	256	1
C5	$3 \times 3 \times 3$	512	1
C6	$3 \times 3 \times 3$	512	1
C7	$3 \times 3 \times 3$	512	1
C8	$3 \times 3 \times 3$	512	1
P1	$1 \times 2 \times 2$	—	2
P2	$2 \times 2 \times 2$	—	2
P3	$2 \times 2 \times 2$	—	2
P4	$2 \times 2 \times 2$	—	2
P5	$2 \times 2 \times 2$	—	2
FC1	—	4096	—
FC1	—	4096	—
Feat	—	2	—

类似地，我们采用三维卷积核来对卷积层的特征进行操作，在这个结构中，卷积层中每一个特征图都会与上一层中多个邻近的连续帧相连，以此捕捉运动信息。需要注意的是：三维卷积核只能从帧立方体中提取一种类型的特征，因为在整个立方体中卷积核的权值 C 都是一样的，也就是共享权值，都是同一个卷积核，因此，我们可以采用多种卷积核，以提取多种特征。对于卷积神经网络来说，有一个通用的设计规则就是，在后面的层（离输出层近的层）特征图的个数应该增加，这样就可以从低级的特征图组合产生更多类型的特征。形式化地说，对于某一层特征图 (x, y, z) 位置处的特征值 $v(x, y, z)$ 可以由下式计算，其中 (W, H, D) 为卷积核的大小， W ， H 和 D 分别表示卷积核的宽度、高度和时间序列的深度。

$$v(x, y, z) = f(b + \sum_{d=0}^D \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} C \cdot v(x+w, y+h, z+d)) \quad (5-2)$$

在图 5-1 所示的结构中，三维卷积神经网络的配置包含两个部分，如图 5-4 所示：1) 8 个卷积层 (C1, C2, C3, C4, C5, C6, C7, C8) 和 5 个最大池化层 (P1, P2, P3, P4, P5)；2) 2 个全连接层 (FC1, FC2) 和特征输出层 (Feat)。具体参数如下，其中卷积核尺寸为 $d \times k \times k$ ，其中 d 为时间核的深度， k 为空间核的尺寸，扫描步长指卷积核在每次在图像中运动的像素大小。

在本章中，我们以 Du Tran 等人已经训练好的深度模型 Sport1M[103]为起点，在步态数据库 OU-ISIR LP 中进行参数微调训练，实现依然采用了 Caffe 框架[25]。在行人身份识别任务中，我们将输出层的类别替换为与数据集中行人数量相关联的行人身份数量。在模型训练之前，我们首先使用文献[103]中训练好的神经网络参数对本模型进行初始化。在训练过程中，同样地，我们固定低层的卷积层参数，微调全连接层参数，只对上层网络进行参数更新，然后使用随机梯度下降法对整个 CNN 模型进行训练，当模型逐渐稳定后，再将损失函数的值后验传播到其他所有网络层，对整个网络模型进行微调，直至模型收敛。

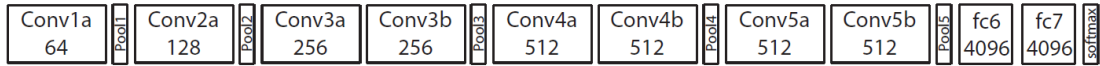


图 5-4 三维卷积网络结构图

5.3 基于 3D-Siamese 网络的步态识别技术

在步态数据库中，我们首先针对每一个行人的步态视频序列运动图像分割方法提取运动轮廓图。然后通过缩放归一化到统一尺度，方便接下来的神经网络训练。具体地，对于 OU-ISIR LP 中的 3895 人的步态视频数据库，每段步态视频序列均已进行标注。该数据库应包含同一行人在不同视角下的步态视频序列，例如 55、65、75、85 度共四个视角，并且上述所有步态视频序列已经标注了行人的身份，使用 Graph Cut 图像分割方法对每一段视频序列提取前景轮廓图，这里所有的轮廓图归一化后尺寸为 128×171 像素。

5.3.1 特征训练

在获得了三维卷积特征之后，我们按照上一章中的方法从步态数据库中选择正负样本构建神经网络训练数据集，其中身份相同的两段步态序列为正样本对，同时设置标签为 1，身份不同的两段步态序列为负样本对，同时设置标签为 0。具体地，在实验中，训练集来源于 OU-ISIR LP 中的 Gallery 库中的人的步态轮廓

序列图，由于行人身份远远大于同一身份行人步态视频序列树，训练集中正样本数量远远少于负样本数量。因此，为保证神经网络训练过程中样本平衡，我们抽取所有的正样本，并且随机抽取了等量的负样本作为最终的训练集合。

然后，将训练数据对同时输入并行的深度神经网络，经过数据切分层，两段特征分别进入网络的两个分支，这两个分支参数权值共享，每一个分支由相同的三个全连接层构成。

最后，两条三维卷积神经网的特征输出层汇聚到对比损失层，通过对比损失函数来计算目标函数的对比损失值。我们利用反向回溯算法对整体模型的参数进行训练。直至在所有训练数据上整体时空联合深度神经网络模型收敛。

5.3.2 特征匹配与识别

在测试识别过程中，首先利用上一节中训练好的模型对测试步态轮廓序列图和匹配库中所有步态轮廓序列图提取特征，然后使用 K 近邻的方法逐一计算测试序列与匹配序列的特征相似度，输出相似度最高的匹配序列，从而识别相同身份的行人。具体步骤如下：

首先，对测试步态视频提取步态轮廓序列图，并归一化到相同尺寸。

然后，得到的步态轮廓序列图输入 $S13$ 中并行卷积神经网络的任一分支，采用前向传播算法提取，步态轮廓序列图的特征。具体地，将其输入如图 2 所示的三维卷积神经网络，网络初始化的参数采用 $S1$ 中训练好的模型参数。在具体操作过程中，我们提取 $FC2$ 层的输出为步态轮廓序列图的特征，特征维度为 4096 维。因此，每一段的步态视频序列均可以由一段 4094 维的特征向量表示。

最后，使用 K 近邻的方法计算测试序列与匹配序列的相似度，输出相似度最高的结果，从而识别相同身份的行人。

5.4 实验结果与分析

在实验中，我们首先使用了和上一章实验中相同的实验设置。下表 5-2 和 5-3 比较了我们的方法与目前最优方法的对比结果，对比方法我们同样选择了 $GEI[47]$ 和 $FDF[47]$ 匹配方法， $HWLD[98]$ 以及上文提出的 $CNN.FC2$ 以及 $SiaNet.FC$ 。从表中我们可以看出，我们提出的基于 3D-Siamese Neural Network 的方法显著超越了目前最优的其他方法大约 1%~7%，值得一提的是，基于三维卷积神经网络的方法在 75° 和 85° 数据集上均取得了最佳的识别效果。然而，我们也可以发现在 55° 和 65° 视角下准确率不及本文上一章中提出的 $SiaNet.FC$ 方法，主要原因是以下三个方面：1) 本文使用的预训练模型来自于 Sport1M 数据集，

相比于 ImageNet 数据集, Sport1M 从数据的丰富性和数据量两方面来说都有不少差距, 因此从特征的表现力来说, 并不如上一章中训练的模型提取出的特征具有区分度; 2) 55°和 65°相比侧面视角来说, 更难以识别, 这是因为从正面视角来看, 行人的躯干、手部和腿部运动并不明显, 而侧面识别往往是最容易区分的; 3) 从轮廓图的角度来说, 正面视角相比侧面视角往往难以捕捉到帧与帧之间的运动信息, 因此这个任务对于三维卷积神经网络来讲是十分具有挑战性的。综上所述, 虽然本章提出的方法并没有在所有视角下都取得了最佳的识别率, 但是在更容易捕捉运动信息的侧面视角上仍然取得显著的识别效果, 进一步提升了步态识别的精度。

表 5-2 在 OU-ISIR LP 数据库上 Rank-1 识别准确率

	55	65	75	85	All
HWLD[98]	—	—	—	87.70	95.50
GEI[47]	84.70	86.63	86.91	85.72	94.24
FDF[47]	83.89	85.49	86.59	85.90	94.17
CNN.FC2	73.96	76.71	77.87	78.82	86.09
SiaNet.FC[97]	90.12	91.14	91.18	90.43	96.02
3D-SiaNet	68.23	70.53	92.05	90.55	93.45

表 5-3 在 OU-ISIR LP 数据库上 Rank-5 识别准确率

	55	65	75	85	All
HWLD[98]	—	—	—	94.70	98.50
GEI[47]	92.39	92.84	92.78	93.01	97.13
FDF[47]	91.53	92.81	92.88	92.83	97.10
CNN.FC2	86.64	88.67	89.39	90.09	93.56
SiaNet.FC[97]	94.98	95.90	95.92	95.97	98.31
3D-SiaNet	72.01	73.80	96.32	95.36	96.18

5.5 本章小结

本章提出了一种基于时空联合深度神经网络的步态识别方法, 利用步态的周期性模式, 将三维卷积神经网络的特征学习过程并行化, 通过对比损失函数来优调模型参数, 最后采用 K 近邻算法计算步态特征相似度。实验表明, 该方法能够克服目前深度学习方法中的分类模型与步态识别问题的领域鸿沟, 利用步态的周期性模式对步态序列进行特征表示, 因此, 能够提高步态识别的准确率。

第六章 结束语

6.1 总结

步态是一种极具潜质的生物特征，通过对步态的分析，人们可以得到行人的身份信息和属性信息。相比于其他行人识别技术，步态识别技术具有可远程获取、鲁棒性强和安全性强的优势，在“以人中心”的现代智能监控系统中，步态识别技术已经获得了越来越多的关注，因此，步态识别问题具有巨大的研究价值和市场需求。

本文中，我们重点研究了基于深度学习的步态识别若干关键技术，包括基于传统卷积神经网络的步态识别技术、距离度量学习理论在步态识别问题中的应用以及基于时空联合深度神经网络的步态识别技术。本文首先对步态识别相关技术进行了较为全面的综述，剖析了步态识别的五个技术阶段，并从多个维度对现有的步态特征提取与表示方法进行归纳和总结，并分析了传统步态识别方法的不足。接下来，我们针对深度学习理论应用于步态识别问题中出现的特有挑战，围绕卷积神经网络和距离度量学习理论进行了较为深入的研究。通过研究基于深度学习的步态识别关键技术，本文通过选择具有区分度的步态特征，克服目前深度学习方法中的分类模型与步态识别问题的领域鸿沟，最后利用步态的周期性模式对步态序列进行特征表示。希望在提高步态识别方法精度的同时，完善步态识别技术的理论研究，最终为高精度的步态识别走向应用提供技术支持。本文的主要创新点与贡献在于：

(1) 设计了一种基于深度学习的步态识别技术框架。

本文的第一个研究重点就是如何设计一个基于深度学习的步态识别技术框架来提高识别的精度。为了克服现有步态数据库样本容量小以及深度学习训练速度慢的挑战，我们首先研究设计将原始的视频帧序列进行融合，计算其步态能量图作为深度神经网络的输入。然后，以预训练得到的模型作为起点，利用步态训练数据集和反向传播对其进行 Fine-tuning 训练，将模型适应到步态识别任务。接下来，将训练好的模型作为特征提取器，将其隐含层的输出作为步态特征表达，并利用 K 近邻算法计算测试步态序列与数据库中待识别步态序列的距离，根据相似度最大时匹配库中的优化步态特征对应的已经人物信息，确定所述待识别人的身份信息。实验表明，将深度卷积神经网络直接应用于步态识别问题时能显著提高基于步态的属性分类任务的精度，但在身份识别任务的识别精度还有待进一步提高。

(2) 提出了一种基于 Siamese Neural Network 的步态识别方法。

将传统的卷积神经网络直接应用于步态特征的提取,虽然能够获取区分度较强的图像特征,但是仍然面临着训练数据量的不足以及分类与识别问题存在领域鸿沟的挑战。针对这些问题,本文引入了距离度量学习模型来解决训练数据量不足以及分类和识别问题的领域鸿沟问题。其主要思想在于通过学习一个损失函数将输入的数据模式映射到一个隐藏的语义空间,具体到步态识别问题上,对于相同行人的步态图,其相似性距离变小,对于不同行人的步态图,其相似性距离变大。这种并行的卷积结构十分有利与特征识别与匹配任务。因此,本文利用并行神经网络结构来克服步态识别的问题中的诸多挑战问题,提出了一整套完整的适用于步态识别问题的深度学习解决方案。实验表明,在目前世界上最大的步态数据库上,本文提出的方法与已有最好的相比,在相同视角和不同视角的步态识别任务中,平均提升了 5% 的正确识别率。

(3) 提出了一种基于时空联合深度神经网络的步态识别方法

二维卷积神经网络虽然具备提取丰富具有区分度的特征的能力,但是受限于二维空间的图像输入。步态作为一种运动模式,具有空间和时间两个维度的特征,周期性是步态识别与传统的人脸识别或者动作识别最为显著的差异,如何充分利用最为原始的步态序列图成为另一个提高步态识别准确率的瓶颈。因此,我们提出了通过计算步态运动的三维卷积特征来增强模型,该模型可以从连续性的周期序列中捕捉时间和空间信息,进一步提高步态识别的准确率和实用性。除此之外,我们提出了将三维神经网络与 Siamese 结构相结合的 3D-Siamese 网络,这种时空联合深度神经网络能够在三维空间进行特征度量学习。实验表明,本文算法能够充分利用二者的优势,从而达到最佳的步态识别效果。

6.2 展望

本文对基于深度学习的步态识别的若干关键问题做了深入的研究,并且取得了一些研究成果,所提出的方法达到了设计目标,但所涉及了主要工作依然存在很大的提升空间,还有可以改进的地方,具体在以下几个方面:

(1) 大规模真实场景步态数据库

步态识别研究的第一步便是采集数据建库,好的数据库能够科学合理公正地评价算法好坏,同时为步态识别技术从实验室走向产业应用奠定良好的基础。目前公开的数据虽然已经逐步发展到千人规模并且部分数据库考虑了实际监控场景的需求,然后还没有任何数据库能够兼顾样本规模和采集场景的真实性,因此,构建一个千人规模的真实场景下的步态数据库是未来研究的第一个重点方向。

(2) 端到端的步态识别算法框架

步态识别是一个应用驱动的科学问题，在本文中，我们较为详细地分析步态识别技术的五个关键阶段，并且在深度学习的理论框架下显著提高了步态特征提取效率和匹配的准确度。然而，一个端到端的步态识别算法框架亟待研究，深度学习具备处理原始图像和视频数据能力，因此如何从原始的步态视频中直接进行身份识别而不必进行繁琐的数据预处理过程是下一的工作难点和方向。

(3) 半周期或准周期步态识别技术

目前常用的步态识别方法需要尽量获取一个完整的步态周期，再从中提取时空特征进行分析，然而在实际场景中，由于目标运动的不规律性、拍摄质量的下降以及拍摄时长的限制，往往难以得到完整的步态序列，如何在半周期或是不完全周期的情况下进行精准的步态识别，是一个非常值得研究的问题，并且具有重要的研究价值。

(5) 多模态步态识别技术

步态是一种生物特征，其他生物识别技术包括：人脸识别、指纹识别、声纹识别等等。单一模态的生物识别技术虽然在特征场景下各有优势，但是多模态的生物识别技术能够综合利用人的多种固有属性进行身份识别，从而在各种场景下都具有普遍性。比如，在监控场景下人脸和步态特征的联合应用往往克服了单模态的诸多不足，因此将步态特征与人脸或是其他特征相融合进行行人识别都是非常有意义的研究工作。

参考文献

- [1] Ma H D. Internet of things: Objectives and scientific challenges[J]. Journal of Computer science and Technology, 2011, 26(6): 919-924.
- [2] 黄铁军, 郑锦, 李波, 等. 多媒体技术研究: 2013——面向智能视频监控的视觉感知与处理[J]. 中国图象图形学报, 2014, 19(011): 1539-1562.
- [3] 傅慧源. 面向人群监管的智能视频场景理解技术研究[D]. 博士学位论文, 北京邮电大学, 2014.
- [4] Ma H D, Zeng C, Ling C X. A reliable people counting system via multiple cameras[J]. ACM Trans. on Intelligent Systems and Technology, 2012, 3(2): 31.
- [5] Fu H, Ma H D. Real-time crowd detection based on gradient magnitude entropy model[C]. In Proc. of the ACM International Conference on Multimedia, 2014: 885-888.
- [6] 黄凯奇, 陈晓棠, 康运锋, 谭铁牛. 智能视频监控技术综述[J]. 计算机学报, 2014, Vol.37: No.49.
- [7] Gait, Wikipedia. [EB/OL] <https://en.wikipedia.org/wiki/Gait>
- [8] Murray M P, Drought A B, Kory R C. Walking patterns of normal men[J]. The Journal of Bone & Joint Surgery, 1964, 46(2): 335-360.
- [9] Murray M P. Gait as a Total Pattern of Movement: Including a Bibliography on Gait[J]. American Journal of Physical Medicine & Rehabilitation, 1967, 46(1): 290-333.
- [10] Winter D A. Biomechanics and motor control of human movement[M]. John Wiley & Sons, 2009.
- [11] Cutting J E, Kozlowski L T. Recognizing friends by their walk: Gait perception without familiarity cues[J]. Bulletin of the Psychonomic Society, 1977, 9(5): 353-356.
- [12] Kozlowski L T, Cutting J E. Recognizing the sex of a walker from a dynamic point-light display[J]. Perception & Psychophysics, 1977, 21(6): 575-580.
- [13] Troje N F. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns[J]. Journal of vision, 2002, 2(5): 2.
- [14] Cunado D, Nixon M S, Carter J N. Automatic extraction and description of human gait models for recognition purposes[J]. Computer Vision and Image

- Understanding, 2003, 90(1): 1-41.
- [15] Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition[C]. In Proc. of IEEE International Conference on Pattern Recognition, 2006, 4: 441-444.
 - [16] Sarkar S, Phillips P J, Liu Z, et al. The humanid gait challenge problem: Data sets, performance, and analysis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(2): 162-177.
 - [17] Han J, Bhanu B. Individual recognition using gait energy image[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006, 28(2): 316-322.
 - [18] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2005, 1: 886-893.
 - [19] Ng P C, Henikoff S. SIFT: Predicting amino acid changes that affect protein function[J]. Nucleic acids research, 2003, 31(13): 3812-3814.
 - [20] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
 - [21] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and Vision Computing, 2004, 22(10): 761-767.
 - [22] Bengio Y, Courville A C, Vincent P. Unsupervised feature learning and deep learning: A review and new perspectives[J]. CoRR, abs/1206.5538, 2012, 1.
 - [23] Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning[J]. APSIPA Trans. on Signal and Information Processing, 2014, 3: e2.
 - [24] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
 - [25] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. In Proc. of ACM International Conference on Multimedia, 2014: 675-678.
 - [26] Van Doornik J, Sinkjær T. Robotic platform for human gait analysis[J]. IEEE Trans. on Biomedical Engineering, 2007, 54(9): 1696-1702.
 - [27] Zheng S, Huang K, Tan T. Evaluation framework on translation-invariant representation for cumulative foot pressure image[C]. In Proc. of IEEE International Conference on Image Processing, 2011: 201-204.

- [28] Lee S W, Mase K, Kogure K. Detection of spatio-temporal gait parameters by using wearable motion sensors[C]. In Proc. of IEEE-EMBS Annual International Conference of Engineering in Medicine and Biology Society, 2006: 6836-6839.
- [29] Lu J, Wang G, Moulin P. Human identity and gender recognition from gait sequences with arbitrary walking directions[J]. IEEE Trans. on Information Forensics and Security, 2014, 9(1): 51-61.
- [30] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[C]. In Proc. of IEEE International Conference on Pattern Recognition, 2004, 2: 28-31.
- [31] Foresti G L. Object recognition and tracking for remote video surveillance[J]. IEEE Trans. on Circuits and Systems for Video Technology, 1999, 9(7): 1045-1062.
- [32] Fu H, Ma H, Ming A. EGMM: An enhanced Gaussian mixture model for detecting moving objects with intermittent stops[C]. In Proc. of IEEE International Conference on Multimedia and Expo, 2011: 1-6.
- [33] Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images[C]. In Proc. of IEEE International Conference on Computer Vision, 2001, 1: 105-112.
- [34] Lam T H W, Cheung K H, Liu J N K. Gait flow image: A silhouette-based gait representation for human identification[J]. Pattern Recognition, 2011, 44(4): 973-987.
- [35] Makihara Y, Aqmar M R, Trung N T, et al. Phase Estimation of a Single Quasi-Periodic Signal[J]. IEEE Trans. on Signal Processing, 2014, 62(8): 2066-2079.
- [36] Tan D, Huang K, Yu S, et al. Efficient night gait recognition based on template matching[C]. In Proc. of IEEE International Conference on Pattern Recognition, 2006, 3: 1000-1003.
- [37] Sarkar S, Phillips P J, Liu Z, et al. The humanid gait challenge problem: Data sets, performance, and analysis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(2): 162-177.
- [38] Wang L, Tan T, Ning H, et al. Silhouette analysis-based gait recognition for human identification[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1505-1518.
- [39] Chapelle O, Keerthi S S. Efficient algorithms for ranking with SVMs[J].

- Information Retrieval, 2010, 13(3): 201-215.
- [40] Tanawongsuwan, R and Bobick, A. Technical Report GIT-GVU-03-01[J], Georgia Institute of Tech, 2003.
- [41] Lee L, Grimson W E L. Gait analysis for recognition and classification[C]. In Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, 2002: 148-155.
- [42] Gross, R. and Shi, J. Tech. report CMU-RI-TR-01-18[J]. Robotics Institute, Carnegie Mellon University, 2001.
- [43] Chalidabhongse, T, Kruger, V, and Chellappa, R. Technical Report[J], University of Maryland, 2001.
- [44] Nixon, M S , Carter, J N, Shutler J D, et al. Technical Report[J], University of Southampton, 2001.
- [45] Shutler J D, Grant M G, Nixon M S, et al. On a large sequence-based human gait database[M]. Applications and Science in Soft Computing. Springer Berlin Heidelberg, 2004: 339-346.
- [46] Wang L, Ning H, Hu W, et al. Gait recognition based on procrustes shape analysis[C]. In Proc. of IEEE International Conference on Image Processing, 2002, 3: III-433-III-436 vol. 3.
- [47] Iwama H, Okumura M, Makihara Y, et al. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition[J]. IEEE Trans. on Information Forensics and Security, 2012, 7(5): 1511-1521.
- [48] Sarkar S, Phillips P J, Liu Z, et al. The humanid gait challenge problem: Data sets, performance, and analysis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(2): 162-177.
- [49] Hofmann M, Geiger J, Bachmann S, et al. The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits[J]. Journal of Visual Communication and Image Representation, 2014, 25(1): 195-206.
- [50] Makihara Y, Matovski D, Nixon M S, et al. Gait Recognition: Databases, Representations, and Applications[J]. Technical Report, 2015.
- [51] Zhang Z, Hu M, Wang Y. A survey of advances in biometric gait recognition[M]. Biometric Recognition. Springer Berlin Heidelberg, 2011: 150-158.
- [52] Yam C Y, Nixon M S, Carter J N. Automated person recognition by walking

- and running via model-based approaches[J]. *Pattern Recognition*, 2004, 37(5): 1057-1072.
- [53] Wagg D K, Nixon M S. On automated model-based extraction and analysis of gait[C]. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, 2004: 11-16.
- [54] Wang L, Ning H, Tan T, et al. Fusion of static and dynamic body biometrics for gait recognition[J]. *IEEE Trans. on Circuits and Systems for Video Technology*, 2004, 14(2): 149-158.
- [55] Jean F, Albu A B, Bergevin R. Towards view-invariant gait modeling: Computing view-normalized body part trajectories[J]. *Pattern Recognition*, 2009, 42(11): 2936-2949.
- [56] Zhao G, Liu G, Li H, et al. 3D gait recognition using multiple cameras[C]. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, 2006: 529-534.
- [57] Tao D, Li X, Wu X, et al. General tensor discriminant analysis and gabor features for gait recognition[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007, 29(10): 1700-1715.
- [58] Yang X, Zhou Y, Zhang T, et al. Gait recognition based on dynamic region analysis[J]. *Signal Processing*, 2008, 88(9): 2350-2356.
- [59] Chen C, Zhang J, Fleischer R. Distance approximating dimension reduction of Riemannian manifolds[J]. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010, 40(1): 208-217.
- [60] Xu D, Huang Y, Zeng Z, et al. Human gait recognition using patch distribution feature and locality-constrained group sparse representation[J]. *IEEE Trans. on Image Processing*, 2012, 21(1): 316-326.
- [61] Little J, Boyd J. Recognizing people by their gait: the shape of motion[J]. *Videre: Journal of Computer Vision Research*, 1998, 1(2): 1-32.
- [62] Bashir K, Xiang T, Gong S, et al. Gait Representation Using Flow Fields[C]. In *Proc. of British Machine Vision Conference*, 2009: 1-11.
- [63] Liu Z, Sarkar S. Improved gait recognition by gait dynamics normalization[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006, 28(6): 863-876.
- [64] Ho M F, Chen K Z, Huang C L. Gait analysis for human walking paths and identities recognition[C]. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2009: 1054-1057.

- [65] Veres G V, Gordon L, Carter J N, et al. What image information is important in silhouette-based gait recognition?[C]. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2004, 2: II-776-II-782 Vol. 2.
- [66] Boulgouris N V, Plataniotis K N, Hatzinakos D. Gait recognition using dynamic time warping[C]. In Proc. of IEEE Workshop on Multimedia Signal Processing, 2004: 263-266.
- [67] Boulgouris N V, Plataniotis K N, Hatzinakos D. Gait recognition using linear time normalization[J]. Pattern Recognition, 2006, 39(5): 969-979.
- [68] Xu D, Huang Y, Zeng Z, et al. Human gait recognition using patch distribution feature and locality-constrained group sparse representation[J]. IEEE Trans. on Image Processing, 2012, 21(1): 316-326.
- [69] Ran Y, Chellappa R, Zheng Q. Finding gait in space and time[C]. In Proc. of IEEE International Conference on Pattern Recognition, 2006, 4: 586-589.
- [70] Ran Y, Zheng Q, Chellappa R, et al. Applications of a simple characterization of human gait in surveillance[J]. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(4): 1009-1020.
- [71] Huang G C, Wang Y H. Human gait recognition based on XT plane energy images[C]. In Proc. of IEEE International Conference on Wavelet Analysis and Pattern Recognition, 2007, 3: 1134-1138.
- [72] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [73] Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for boltzmann machines[J]. Cognitive science, 1985, 9(1): 147-169.
- [74] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering[C]. In Proc. of ACM International Conference on Machine learning, 2007: 791-798.
- [75] Salakhutdinov R, Hinton G E. Deep boltzmann machines[C]. In Proc. of International Conference on Artificial Intelligence and Statistics. 2009: 448-455.
- [76] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [77] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 1995, 3361(10).
- [78] Weinberger K Q, Blitzer J, Saul L K. Distance metric learning for large

- margin nearest neighbor classification[C]. In Proc. of Advances in Neural Information Processing Systems. 2005: 1473-1480.
- [79] Bar-Hillel A, Hertz T, Shental N, et al. Learning distance functions using equivalence relations[C]. In Proc. of International Conference on Machine Learning. 2003, 3: 11-18.
- [80] Domeniconi C, Peng J, Gunopulos D. Locally adaptive metric nearest-neighbor classification[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(9): 1281-1285.
- [81] Chechik G, Sharma V, Shalit U, et al. Large scale online learning of image similarity through ranking[J]. The Journal of Machine Learning Research, 2010, 11: 1109-1135.
- [82] De Maesschalck R, Jouan-Rimbaud D, Massart D L. The mahalanobis distance[J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50(1): 1-18.
- [83] Xing E P, Jordan M I, Russell S, et al. Distance metric learning with application to clustering with side-information[C]. In Proc. of Advances in Neural Information Processing Systems. 2002: 505-512.
- [84] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification[C]. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2005, 1: 539-546.
- [85] Bromley J, Bentz J W, Bottou L, et al. Signature verification using a “Siamese” time delay neural network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(04): 669-688.
- [86] Muja M, Lowe D G. Fast matching of binary features[C]. In Proc. of IEEE Conference Computer and Robot Vision, 2012: 404-410.
- [87] Bashir K, Xiang T, Gong S. Gait recognition using gait entropy image[C]. In Proc. of International Conference on Crime Detection and Prevention, 2009: 1-6.
- [88] Makihara Y, Sagawa R, Mukaigawa Y, et al. Gait recognition using a view transformation model in the frequency domain[C]. In Proc. of European Conference on Computer Vision. Springer Berlin Heidelberg, 2006: 151-163.
- [89] Bashir K, Xiang T, Gong S. Gait recognition without subject cooperation[J]. Pattern Recognition Letters, 2010, 31(13): 2052-2060.
- [90] Wang C, Zhang J, Pu J, et al. Chrono-gait image: a novel temporal template for gait recognition[C]. In Proc. of European Conference on Computer Vision.

- Springer Berlin Heidelberg, 2010: 257-270.
- [91] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. In Proc. of Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [92] Wan J, Wang D, Hoi S C H, et al. Deep learning for content-based image retrieval: A comprehensive study[C]. In Proc. of ACM International Conference on Multimedia, 2014: 157-166.
- [93] Yu S, Tan T, Huang K, et al. A study on gait-based gender classification[J]. IEEE Trans. on Image Processing , 2009, 18(8): 1905-1910.
- [94] Li X, Maybank S J, Yan S, et al. Gait components and their application to gender recognition[J]. IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2008, 38(2): 145-155.
- [95] Hu M, Wang Y, Zhang Z, et al. Combining spatial and temporal information for gait based gender classification[C]. In Proc. of IEEE International Conference on Pattern Recognition, 2010: 3679-3682.
- [96] 胡懋地. 基于图像序列中行人步态特征的属性分类和身份识别研究[D]. 博士学位论文[D]. 北京航空航天大学, 2013.
- [97] Zhang C, Liu W, Ma H D, et al. Siamese Neural Network based Gait Recognition for Human Identification[C]. In Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, 2016.
- [98] Sivapalan S, Chen D, Denman S, et al. Histogram of weighted local directions for gait recognition[C]. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013: 125-130.
- [99] Muramatsu D, Shiraishi A, Makihara Y, et al. Gait-Based Person Recognition Using Arbitrary View Transformation Model[J]. IEEE Trans. on Image Processing, 2015, 24(1): 140-154.
- [100] Martí-Fdez R, Xiang T. Gait recognition by ranking[M]. In Proc. of European Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 328-341.
- [101] Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks[C]. In Proc. of IEEE Conference Computer Vision and Pattern Recognition, 2014: 1725-1732.
- [102] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[C]. In Proc. of Advances in Neural Information Processing Systems. 2014: 568-576.

- [103] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C]. In Proc. of IEEE International Conference on Computer Vision. 2015: 4489-4497.
- [104] Taylor G W, Fergus R, LeCun Y, et al. Convolutional learning of spatio-temporal features[C]. In Proc. of European Conference on Computer Vision. Springer Berlin Heidelberg, 2010: 140-153.
- [105] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.

致 谢

“你为什么选择做科研？”如果去问一千个苦逼的研究生，我想可能会有一千种答案。学术的道路我想可能就是“提升自我，影响他人”，同时也要承受更多的孤独和寂寞，我依然记得实验结果不好时自己的沮丧，以及论文被接收时自己的喜悦。从现在看来，自己的选择虽不敢说有多好，也许并不是那么金光熠熠，但至少我过得很舒服。

感谢我的导师马华东教授。“师者，所以传道授业解惑也。”从当初保研进入北邮到现在即将毕业，马老师是我学术科研道路的引路人，他带我一步步入门，对我的教诲不仅仅局限于学术领域，更多的是对待科研的态度。从他的身上我看到对科研的狂热执着、对治学的严谨认真以及对工作的精益求精，这些都令我印象深刻，一日为师终身为师，这些点点滴滴也必将对我今后的研究和工作产生深远影响。

感谢刘亮老师，他将正确的科研观和热爱生活的情怀传达给我，无论是科学研究还是对待生活的态度，皆为我辈楷模；感谢傅慧源老师，他是老师更是我的兄长，傅师兄在近三年的时间里对我的科研项目进行悉心指导，帮助我不断进步和成长；感谢刘武老师，他在科研方法和论文写作方面给予了我很多指导和帮助，超强的学术能力也是我今后努力的目标。感谢李文生老师、段鹏瑞老师、张海涛老师、赵东老师，我从他们身上学到了很多很多。

此外，我要感谢实验室热心帮助我的师兄弟们。物联网技术中心确实是一个团结和谐的集体，作为一个视觉领域的初学者，难免会走很多弯路，每当我遇到困难的时候，他们总是很友善热心地给予我帮助。实验室的张征、刘鑫辰、李双群、王浩、卢大炜、陈建伟、高一鸿、周沫、唐毅、初颖俊、张丹、秦通、刘全星、王冲、王泽宇、杜鹏程、赵帅、于韶峰、孙云鹏、雷雪蕊、寇月、付广平、杨贤达、张闯、刘天骄、齐恒、程鹏、刘鲲，挂一漏万，感谢你们的帮助和支持。

最后，最重要地，我要感谢我的父母。他们一如既往地理解我并且支持我从事自己热爱的事业。惭愧的是，在同龄的很多孩子为了票子、房子、车子奋斗的时候，我大部分的时间还是想着论文、程序、算法和项目。感谢他们对我的无限包容和不计回报的支持，才让我有动力在自己选择的道路上坚定地走下去。

张诚
二零一六年，春
于北邮教三

参加科研及获奖情况说明

发表和投稿的学术论文：

- [1] Cheng Zhang, Wu Liu, Huadong Ma, Huiyuan Fu, Siamese Neural Network based Gait Recognition for Human Identification[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 2016. (EI)
- [2] Cheng Zhang, Huadong Ma, Huiyuan Fu, Object Tracking in Surveillance Videos using Spatial-temporal Correlation Graph Model [J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 713 - 720. (in Chinese, EI)
- [3] 张诚, 马华东, 傅慧源. 多摄像机监控中基于时空关联图模型的目标跟踪[C]. 第 10 届和谐人机环境联合会议, 北京, 2014. (口头报告)
- [4] Zheng Zhang, Huadong Ma, Huiyuan Fu, Cheng Zhang, Scene-free Multi-class Weather Classification on Single Images[J]. Neurocomputing. (under review, SCI)
- [5] Zheng Zhang, Huadong Ma, Huiyuan Fu, Liang Liu, Cheng Zhang, Outdoor Air Quality Level Inference via Surveillance Cameras[J]. Mobile Information Systems. (under review, SCI)

专利申请情况：

- [1] Huadong Ma, Wu Liu, Huiyuan Fu, Cheng Zhang. Method and Device for Gait Recognition, U. S. patent, 10713-002US1.
- [2] 马华东, 刘武, 傅慧源, 张诚. 一种步态识别方法与装置, 中国, 专利号: 201510958380.4.
- [3] 马华东, 傅慧源, 张诚, 方瑞. 一种多目标跟踪方法与装置, 中国, 专利号: 201510676986.9.
- [4] 马华东, 傅慧源, 刘鑫辰, 张诚. 一种基于车辆轨迹向量确定车辆实际方向的方法及装置. 中国, 公开号: CN105117737A.

参与的科研项目：

- [1] 国家工信部物联网发展专项项目
- [2] 国家自然科学基金资助项目 (61402048)

获奖情况：

- [1] 北京市优秀毕业生
- [2] 国家奖学金
- [3] 威睿 (Vmware) 奖学金
- [4] 北京邮电大学优秀班主任, 北京邮电大学优秀研究生干部