分类号 学号 M201572828

学校代码 10487 密级



**硕士学位论文**

**基于深度特征的行人跟踪研究**

|  |  |
| --- | --- |
| 学位申请人： | 余成跃 |
| 学科专业： | 计算机应用技术 |
| 指导教师： | 凌贺飞 教授 |
| 答辩日期： | 2018年 月 日 |

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**for the Degree of the Master of Engineering**

**The Research on Person Object-Tracking**

**Based on Deep-Feature**

|  |  |
| --- | --- |
| **Candidate :** | **Yu Chengyue** |
| **Major :** | **Computer Applied Technology** |
| **Supervisor:** | **Prof. Ling Hefei** |

**Huazhong University of Science & Technology**

**Wuhan 430074, P. R. China**

**May, 2018**

独创性声明

本人声明所呈交的学位论文是我个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，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：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电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密□， 在 年解密后适用本授权书。

本论文属于

不保密□。

（请在以上方框内打“√”）

学位论文作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

# 摘 要

全国主要城市都在致力于建设“平安城市”，作为其中一项不可或缺的功能，行人再识别因其应用价值与研究价值吸引了大量研究者的兴趣，主要研究如何从不同的摄像头中找到查询行人图像。行人视觉外观会受分辨率、光照、角度、姿势、遮挡等一系列因素影响，传统手工特征因为表达能力有限，无法有效区分行人图像较大的类内差异和较小的类间差异，深度学习方法特征表达能力更好，能在一定程度上解决问题，其特征的区分能力还可以进一步提高。

基于属性信息和深度卷积神经网络，提出三种不同的行人再识别算法：（1）结合对行人身份标签的分类方法和验证方法，提出一种行人再识别联合学习算法。主要创新点是**提出了一种约束对比验证损失函数**，用于约束特征值，将特征距离限制在一定范围内。验证目标和分类目标同时迭代优化，可以既学习更有区分性的特征，又度量行人之间的关系。（2）基于行人多标签属性识别网络，提出一种同时学习多种属性和行人身份的行人再识别算法。主要创新点在于**属性标签是自动标注的**，无需手工进行，节省时间。行人的属性信息对外观变化具有鲁棒性，有利于辅助行人再识别的研究。（3）基于上述两种算法，提出一种多属性融合的解决方案。主要创新点是**在一个端到端的框架中融合行人身份的分类损失、验证损失和属性的分类损失、验证损失**，在反向传播时传播各损失的加权之和。融合多种策略的方法可以使各组成部分优势互补，进一步提高行人再识别率。

通过在Market-1501和PRW两个数据集上进行充分的实验验证，结果表明，CNN分类结合对比验证方法和多属性方法的性能都比较有竞争力，而结合多属性多策略的方法又显著提高了再识别的准确率，在Market-1501数据集上，累积匹配特性第一准确率达到了70.0%，平均准确率均值达到了45.7%，已经超过了许多顶尖的方法。

**关键字：**深度学习，卷积神经网络，多属性融合，行人再识别，行人多属性识别

# **Abstract**

Major cities in China are devoted to building the Safety City system. As an integral part of the Safety City system, person re-identification has drawn great interest in researchers due to its application and research significance. It aims at studying how to figure out a person of interest in different cameras. The visual appearance of a person is affected by a variety of factors such as resolution, illumination, viewpoint, pose and partially occlusion. The conventional hand-crafted based features have limited representation ability for large intra-class and small inter-class variations. The deep learning method have a better representation for the features, it solved the problems to a certain extent, even though, the discriminative ability still have space to improve.

Based on the attribute information and deep Convolution Neural Networks (CNN), three different person re-identification algorithms are proposed. (1)Combining the classification and verification CNN methods with person ID labels, a joint learned person re-identification algorithm is proposed to. The main innovation point is the Constraint Contrast Verification Loss Function designed to impose an additional constraint to the feature values so that the metric distance can be limited in a certain range. Both the verification and classification objectives are iteratively optimized to learn more discriminative pedestrian features and measure the relationship between pedestrians in the meantime. (2)Based on the person multi-labeled attribute identification network, an algorithm that simultaneously learns multi-attributes and person IDs is designed to solve the person re-identification issue. The main innovation point is that the attribute training labels are built from automatic annotations rather than time-consuming manual annotations. The pedestrian attributes are robust to appearance changes, thus are very helpful for person re-identification. (3)Based on the above two algorithms, a multi-attributes fused solution is proposed. The major innovation point is that the proposed end-to-end framework integrates an ID verification loss, an ID classification loss, a number of attribute verification losses, and a number of attribute classification losses, and back-propagates the weighted sum of the individual losses. The combination of multiple strategies can make the various components complement each other and further improve the person re-identification accuracy.

Experiments are conducted on the Market-1501 and PRW datasets. The results show that both the combination CNNs and the multi-attribute method achieve competitive performance and the multi-task method significantly improves the re-identification accuracy. On the Market-1501 dataset, the Cumulative Match Characteristic Rank-1 accuracy is improved to 70.0%, and the mean Average Precision is improved to 45.7%. The proposed method outperforms most of the state-of-the-art methods.

**Keywords:** Deep Learning, Convolution Neural Networks, Multi-Attributes Fusion, Person Re-identification,Pedestrian Attribute Recognition

# 目 录

[摘 要 I](#_Toc481565410)

[ABSTRACT II](#_Toc481565411)

[1 绪论](#_Toc481565412)

[1.1 研究背景和意义 (1](#_Toc481565413))

[1.2 国内外研究现状 (3](#_Toc481565414))

[1.3 主要研究内容 (7](#_Toc481565418))

[1.4 论文组织结构 (9](#_Toc481565422))

[2 基于分类与对比验证的行人再识别](#_Toc481565423)

[2.1 问题分析 (11](#_Toc481565424))

[2.2 基于分类与对比验证策略的行人再识别算法 (13](#_Toc481565425))

[2.3 实验分析 (18](#_Toc481565430))

[2.4 本章小结 (28](#_Toc481565438))

[3 基于属性的行人再识别](#_Toc481565439)

[3.1 问题分析 (30](#_Toc481565440))

[3.2 行人多属性识别器 (31](#_Toc481565441))

[3.3 基于属性的行人再识别算法 (33](#_Toc481565444))

[3.4 实验分析 (37](#_Toc481565449))

[3.5 本章小结 (45](#_Toc481565455))

[4 基于多属性与多策略融合的行人再识别](#_Toc481565456)

[4.1 问题分析 (47](#_Toc481565457))

[4.2 基于多属性与多策略融合的行人再识别算法 (48](#_Toc481565458))

[4.3 实验分析 (53](#_Toc481565462))

[4.4 本章小结 (60](#_Toc481565467))

[5 总结与展望](#_Toc481565468)

[5.1 全文工作总结 (61](#_Toc481565469))

[5.2 未来工作展望 (62](#_Toc481565470))

[**致 谢** (64](#_Toc481565471))

[**参考文献** (65](#_Toc481565472))

[**附录1 攻读学位期间发表的学术论文目录** (72](#_Toc481565473))

# 1 绪论

## 1.1 研究背景和意义

近来，国内外恐怖事件频繁发生，例如2014年的昆明火车站蒙面暴徒砍人事件，2016年的法国尼斯恐怖袭击事件，2017年伦敦货车冲撞行人事件、慕尼黑地铁站发生枪击案等，对全世界人民的生命财产安全、人民生活幸福度都带来了极大的威胁。因此，各国在公共安全维护的投入也越来越大。我国目前也正大力建设“平安城市”，目前已有超过2000万个监控摄像头覆盖了全国各地车站、机场、校园、商场、街道等公共区域。“平安城市”的建设，大量摄像头的搭建，给人们的生命财产安全、幸福高质量的生活带来了保障的同时，也带来了视频的爆发式的增长，给视频分析处理技术带来了极大的挑战。

在视频监控系统中，产生的大量数据，其中绝大多数都是与人相关，而且我们所关心的也大都是人的特征、姿态、动作、行为等等，因此与人相关的信息获取就至关重要。行人目标跟踪，就是确定视频序列中行人的关系，获得一个目标在整个视频中完整的运动轨迹。这能有效的把视频中的快照进行关联，大量减少分析的快照数量；另外，对于行人的运动速度、行人动作的分析都提供了基础。随着深度学习技术的发展，能够更好的提取目标特征，对目标进行多维度、多层次的表达，使得更加准确的对目标进行描述；相关滤波技术的逐渐发展，与深度学习的有机融合，再加上GPU运算速度上的提升，使得在用深度特征进行目标跟踪时，速度不在是一个实际应用中的约束。在智能监控系统中，使用行人目标跟踪技术，能够快速对视频进行分析处理，为“平安城市”建设节省大量人力物力。

到目前为止，已经有很多目标跟踪的算法，但是运用到实际场景中的并不多。主要是因为跟踪中存在的巨大挑战。而且，对于行人目标跟踪任务，又有其特殊的难点。主要挑战如下：

（1）在跟踪过程中，目标发生外观形变、光照变化、尺度变化、平面外旋转、平面内旋转等。

（2）目标出现遮挡、出视野、快速运动和运动模糊的问题。

在行人目标跟踪任务中，由于行人自身的特殊性，走路的姿势经常变化，自身携带物品与行人之间相互遮挡，使得行人目标任务的外观形变和遮挡问题尤为突出。图1.1是以OTB100[]与MOT2016[]数据集为例说明常见的追踪过程中的问题。其中第一行为运动模糊，第二行为姿态变化，第三行为遮挡问题。设计出一种能应对上述的难点，又不失效率的行人目标跟踪方法，是十分具有挑战性的研究任务。



图1.1 行人跟踪任务中的挑战

近年来，深度学习在视觉领域取得了巨大的成功，例如目标的分类、检测、分割、检索等。目标跟踪作为视觉领域的传统问题，使用深度学习的方法，也一定能取得很好的效果。本文将围绕行人目标跟踪面临的挑战展开研究，使用基于深度学习和相关滤波的方法，设计实现针对行人的目标追踪算法，并对该算法的准确性，鲁棒性等进行充分的实验验证。

## 1.2 国内外研究现状

目标跟踪是指，在视频序列的第一帧指定一个矩形框，这个矩形框在数据集中是人工标注的，但在实际应用场景中大多是检测算法的检测结果，然后需要跟踪算法在后续的视频序列中跟住这个矩形框。目前的跟踪算法基本都是广义上的跟踪算法，即并不关心跟踪的目标是什么，这种广义上的目标跟踪算法，提取的目标特征都是基本上是统计或梯度特征，如灰度特征[]、hog特征[]、颜色直方图[]等，并没有针对性，对于后续的处理有很大的不足。

### 1.2.1 传统的目标跟踪算法

跟踪算法一般来说可以分为生成模型方法和判别模型方法。生成类方法的大概流程是，根据当前帧的目标区域建立模型，在下一帧中寻找与模型最相似的区域，该区域即为预测的目标位置。经典的生成类跟踪算法有卡尔曼滤波[]、粒子滤波[]、mean-shift[]、朴素贝叶斯方法[]、monte carlo[]等。这一类方法，在面对跟踪过程中的复杂变化时，其鲁棒性和准确度方面的表现不尽人意。在VOT2016[]中，ASMS[]和DAT[]也只有32和31名，其效果可见一斑。在当前相关滤波和深度学习如此火热的环境下，基于生成类的目标跟踪算法基本已经淹没在历史的洪流中了。

判别类方法，又称tracking-by-detection，与经典的计算机视觉处理问题的套路一致，即图像特征+机器学习。大体流程是，以目标区域为正样本，背景区域为负样本，使用机器学习的方法训练一个分类器，在之后视频帧中，使用这个分类器进行预测，查找最优区域，并根据该预测区域更新分类器参数。经典的算法，如Struck[]、TLD[]算法等。

### 1.2.2 基于相关滤波的目标跟踪

相关滤波类方法（correlation filter，CF）最早是信号领域的方法，用来描述信号之间的相关性。最早用在跟踪领域的相关滤波算法可以追溯到2010年的MOSSE[]算法，其使用的是单通道灰度特征，速度高达615fps。CSK[]、KCF\DCF[]、CN[]等都是在MOSSE的基础上进行改进。CSK优化了样本的选取过程，设计出密集采样方法，引入了循环矩阵和核的概念，解决了传统算法中采稀疏采样导致的样本冗余问题，从此之后循环矩阵和核技巧在相关滤波的目标跟踪领域大放异彩。KCF在CSK的基础上，使用了多通道梯度的HOG特征，论文中有岭回归、循环矩阵、核技巧、快速检测等方法的理论基础和完整的公式推导。与KCF对应的CN，同样是在CSK的基础上提出，CN使用了颜色特征（color names）。

自从KCF大方异彩后，后续很多相关滤波的工作都是在KCF基础上进行改进，SAMF[]，就是在KCF基础上，联合CN特征和HOG特征，加入尺度估计，解决了KCF中的尺度问题，并且由于CN特征与HOG特征的结合，SAMF对遮挡问题也具有一定抵抗能力。SRDCF[]是在DCF基础上的改进，通过加入空间惩罚项来解决KCF/DCF中的边界效应问题，然而正是由于加入了惩罚项，导致破坏了岭回归的闭式解，只能使用高斯-赛德尔迭代法求解，这使得SRDCF的速度非常慢。

由上面介绍可以看出，相关滤波方法在深度学习方法大热之前，使用的大都是简单的特征如HOG特征、灰度特征、颜色特征等，主要的发展或者说改进方向是训练样本的采样算法，已经采集算法带来的问题，如边界效应等。其中的在闭式解的推导问题上，取得了巨大的突破，使得相关滤波相关的追踪算法效率非常高。然而，追踪算法效果的决定性因素还是在特征表达上，简单特征对目标的表达毕竟是很有限的，虽然也有很多相关的研究来尝试使用多种特征融合、多尺度处理等算法来解决问题，如Staple[]等，但效果上也没有什么巨大的突破。

### 1.2.3 基于深度学习的目标跟踪

深度学习方法目前已经在很多计算机视觉领域的问题中取得了巨大的突破，图像任务中最常用也是表现最好的是卷积神经网络（Convolutional Neural Network, CNN）。最早的CNN模型可以追溯到1998年LeCun等人的LeNet[]，它主要是用来解决手写体字识别。然而，当时计算机运算速度有限，CNN一般层次较深，在反向传播（Back Propagation, BP）中计算速度太慢，人们更加关注支持向量机（SVM）等速度快的机器学习算法，CNN的研究一度停滞不前。2006年Hinton等人[]在Science上发表了一篇CNN降低数据维度的文章，使得CNN中参数大量降低，使训练、过拟合（over-fitting）等问题的得以解决，并且随着大规模数据集的出现，如ImageNet[]，GPU大量并行运算使得计算机运算速度的提升，CNN才再次出现在人们的视野中。2012年的ImageNet大规模视觉识别比赛中，Krizhevsky等人[]的AlexNet以遥遥领先的结果获得了第一名，由此，各种深度学习的模型如雨后春笋般涌出。在分类领域GoogleNet[]、VGGNet[]、ResNet[]，网络模型层次越来越深，效果也越来越好。在检测领域，RCNN[]、SPP-net[]、Fast-RCNN[]、Faster-RCNN[]、Mask-RCNN[]等一系列模型的提出，检测算法的准确率与效率都越来越高。另外，深度学习框架也一个接一个出现，Caffe[]、Tensorflow、Torch等都使得深度学习网络的搭建、训练、应用到实际项目中都越来越简便，极大的推动深度学习的发展。

CNN在视觉领域其他领域大热的时候，在目标跟踪任务上的应用却很迟缓，直到2015年才逐渐出现了FCNT[]、DLT、ROLO等使用CNN的目标跟踪算法。Nam等人的MDNet[]更是在VOT2015 Challenge上获得冠军。另外，使用深度学习与相关滤波相结合的DeepSRDCF[]是在SRDCF的基础上，使用卷积特征进行跟踪，效果还不错，但效率较低。后续的c-cot[]、eco[]等都是结合深度特征与相关滤波，不断改进，准确率和效率都越来越高。SiamFC[]、CFNet都是属于Simaese[]类型的网络，有两条支路，分别表示参考目标和测试样本，其中CFNet更是在网络中加入了Corrlation Filter，尝试将相关滤波作为网络中的一个层加入到CNN中，实现end-to-end的学习方式，充分发挥深度学习的优势。

## 1.3 主要研究内容

本文将围绕着行人目标跟踪所面对的挑战性难题展开研究，使用基于深度学习的方法获得行人特征，与相关滤波的方法相结合，尝试设计实现出一种准确率高、鲁棒性强、效率能够满足实时应用场景需求的行人目标跟踪算法。本文重点将放在行人目标特征提取，跟踪过程中训练与测试样本选取，模型更新策略等几个方面。主要研究点如下：

（1）针对行人目标跟踪这一问题，设计基于行人属性的特征提取算法。在目标跟踪中，特征是目标的唯一表达，特征的好坏直接决定了跟踪算法的准确性。CNN提取的特征，具有表达能力强、鲁棒性好等特点，这些特性已经在其他视觉领域经过验证，因此在目标跟踪算法中使用CNN的深度特征，将能够达到很好的效果。并且本文针对跟踪目标为行人这一特性，设计出基于多属性融合的深度特征，对目标具有极强的表征能力。

（2）相比于其他视觉领域的任务，目标跟踪对效率的要求更加迫切，在实际应用中能够实时进行目标跟踪往往是基本要求了。然而使用CNN模型，即使使用GPU加速运算，速度上的问题也一直都是存在的，如何在目标跟踪任务中使用深度特征，又能保证实时性能是在实际应用中的巨大挑战。本文将引用ROI-Pooling的池化方法，将能够做一次卷积，同时提取多个目标框的特征，以减少运算时间。另外，通过结合预选框（proposal box）策略代替滑动窗口的方法，大量减少筛选框的个数，以加快算法速度。不仅如此，并且通过预选框策略还能够实现帧图像全局搜索，以期解决目标高速运动或离开视野域的问题。

（3）遮挡问题，在目标跟踪任务中是一直存在的，而针对行人目标跟踪这一特殊任务，由于跟踪目标是行人，而行人所处的场景往往是更加复杂的，人作为一种群居动物，遮挡问题更是常见。因此，对于行人目标跟踪问题，能否有效的解决遮挡问题是算法好坏的决定性因素。本文针对遮挡时所带来的脏数据，污染训练样本的问题，设计出模型更新策略，有效避免训练样本被污染，同时精简而有效的模型更新，也使得速度上有所提升。

## 1.4 论文组织结构

论文共有五章，图1.2是各章内容的组织结构图，结合图更便于阐述论文的各章主要内容，理解论文的行文思路，具体结构描述如下。



图 1.2

第一章首先介绍了行人目标跟踪的研究背景和意义，并通过大量的文献介绍目标跟踪的国内外研究现状，根据现有方法中存在的问题或局限性制定了研究内容，最后总结了论文的组织结构。

第二章对现有目标跟踪算法中提取特征方法进行分析，提出所存在的问题，针对这些问题，提出行人多属性融合的特征提取方法，构建CNN网络，针对性的提出多属性融合的损失函数，并通过实验确定参数，最后用实验进行对比验证方法的有效性。

第三章中针对行人目标跟踪在实际应用中的效率问题，结合目标检测中的技术，在网络中添加roi-pooling层，并结合使用预选框（proposal box）策略，实现由粗到细的目标筛选，提高特征提取效率，并且用这种方法构建训练样本，训练更加强壮的网络。最后，用实验进行大量验证。

第四章，针对行人目标跟踪任务重遮挡严重这一问题，设计制定跟踪模型跟踪算法，对特征进行相关性判断，有效进行模型更新。给出模型根系算法流程图，并通过实验进行效果对比验证。

第五章对整个行人目标跟踪研究进行归纳和总结，提出研究中的问题与不足，并提出后续的改进工作。

# 2 基于行人多属性融合的深度特征提取

在目标跟踪任务中，使用的特征大都是简单的特征如灰度特征、HOG特征、颜色特征等，使用这些简单的手工设计特征，主要是由于效率方面的需求，并且也没有足够大的数据集进行深度特征的训练。即使使用深度特征，由于大规模数据集的缺失，并不能针对性的进行训练，而是简单实用分类训练得到的特征。本章将以CNN为基础设计网络结构，提取行人多属性融合特征，并将其应用到行人目标跟踪任务中。

## 2.1 问题分析

目前的目标跟踪算法中使用的特征除了传统的手工特征，还有就是用CNN网络训练得到的深度特征。然而追踪任务不同于分类、检测任务，目前的数据集规模太小，不足以训练一个鲁棒性强的卷积神经网络，所以目前在追踪领域使用的基本都是用ImageNet数据集训练得到的网络模型。将这种没有经过微调的深度特征应用到目标跟踪中，无法充分发挥深度特征的优势。

不仅如此，由于ImageNet数据集是分类任务，其目标是对物体的种类进行区分，而在目标跟踪任务中，是需要判断是否是同一个目标，因此使用这样的CNN提取的目标特征将不具有很强的区分性。例如，在行人目标跟踪任务中，一般场景中会存在非常多的行人，我们需要将这些非目标行人当作背景处理，然而对于分类问题，这些行人都会标记为同一个类别，即“行人”，因此在目标跟踪领域中使用这种CNN进行特征的提取，将不能有效的进行目标与背景的区分。

另外，目前在跟踪领域，大家并不关心跟踪的目标是什么，对所有的跟踪目标使用的特征提取方法是相同的，这主要是沿用了手工特征时使用的方法，因为在跟踪领域使用的手工特征大都是与图像的色彩、纹理或梯度相关，如HOG、灰度特征等，而使用CNN大家更关注的往往是更高层次的语义信息，这就使得用一个大而广的模型去对所有的目标进行特征提取并不合适，这样做会使得目标与背景间的区分度不够。

有上述可以看出，在目标跟踪领域对深度特征的挖掘是远远不够的，基本只停留在用解决其他问题的方法来进行特征的提取，并没有一个针对性强、鲁棒性好的深度特征提取方法。本章将把针对行人目标跟踪这一问题，设计出有针对性的深度特征提取算法。

## 2.2 基于多属性融合的行人特征提取算法

### 2.2.1 网络框架结构设计

在网络框架结构设计上，为了方便训练与横向比较，将CaffeNet[]作为基本结构。训练过程中，输入的标签中包含行人ID和行人属性，再结合分类损失和验证损失，两两组合构成了如图右侧所示的损失函数部分。行人ID和行人属性的分类损失与验证损失的组合，将克服使用分类方法时图像间辨识能力不够，使用验证损失时标签信息利用效率低的问题。另外，由于验证损失需要计算一对图像特征间的距离，训练时如何均匀的生成正负样本对是一个难点，在我们的设计中无需预先对图像进行分组处理，只需要对一个批次的图像进行正向传播，在验证损失层对一个批次中的图像进行两两组合计算特征间距离。

在目标跟踪过程中，该网络结构作为特征提取模块，是预先训练好的离线模型，通过实验，我们确定选取第五个卷积层的深度特征，实验详见2.3章，用该特征进行具体的目标跟踪与跟踪模型更新等，具体细节见第4章。

图2.2中的网络主体结构基本与CaffeNet一样，有五个卷积层，其中第一、第二和第五个卷积层后都有一个最大池化层。不同的是，去掉了第七个全连接层，通过实验进行验证，我们将选取第五个卷积层特征进行行人目标跟踪模型的输入，更希望第五个卷积层能够靠后分布，更加接近语义特征。不就如此，由于是多种损失函数的组合，为了能够方便不同损失函数的计算，我们在每一个损失层前都添加了其对应的一个全连接层，用该全连接层对数据维度进行调整。损失层中包含多种损失函数，总体损失是多个损失的线性叠加，具体的叠加权重将由实验确定。



### 2.2.2 分类损失函数

分类损失使用了softmax的损失函数，如公式（）所示，为简化表述，行人ID和行人属性的分类损失都使用该公式表达，其中ATTR代表行人属性，I代表行人ID。假设有J种属性，则共有J+1种标签，第0个标签表示的是行人ID，第1到第J个标签分别对应J种属性，其中每种标签有个类别。在一个包含*N*张图像的训练批次中，标签*j*的平均损失函数表达如公式(4.1)所示。是图像对于属性j的预测值属于每个类别k的概率。多个属性的总体分类损失函数如公式（）所示。总体分类损失函数如公式（）所示，其中的权重α又实验确定，具体详见2.3章。

在训练过程中需要使用梯度下降的方法进行反向传播，需要对损失函数进行求导，由于分类损失函数是各个标签损失的线性叠加，而且各个标签的损失函数相同，这里将以公式（）为例给出梯度求导公式如公式（）所示，总体损失梯度也将是各个标签损失梯度的线性叠加。其中表示损失层前一层的特征。

### 2.2.3 验证损失函数

验证损失函数是最先应用在人脸识别验证领域，我们在deepid2[]的验证损失函数的基础上进行改进，其目的是为了使得特征空间中是同一个行人目标对应的特征距离较小，不同行人目标间的特征距离较大。在进行验证损失的计算之前，将会通过一个全连接层实现的特征进行降维，使其收敛的速度更快。由于验证损失使用feature map的特征进行对比验证，并没有行人ID、行人属性的区分，因此这里将进行统一描述。对于某个批次中第n个图像、第j个标签的特征向量对，的验证损失如公式(2.1)所示。

其中，代表该特征对所对应的行人对的第j个标签是否一致，表示欧式距离，即L2范数，M为一个阈值，用来防止离群图像特征对间距离过大带来的损失过大的影响，表示向量每一项的绝对值，**1**表示单位向量，表示第一范数。公式()中第一项和第二项分别对应于图像特征对是否是相同的行人标签，第三项是为了使得图像特征向量每一项的绝对值都尽量接近于1的约束，这样能够更方便的确定M的取值。其中M和β的值将在实验中确定，详见2.3章。

对于一个训练批次中的N个图像特征对，J个属性，则共有J+1个验证损失，j=0时代表行人ID，j=1，2…J代表属性，平均损失函数如公式()所示。

同样的，验证损失函数的求导梯度如公式（）所示，其中的三项分别对应着公式()中的三项，公式（）的验证损失即为这三项的叠加。由于max于绝对值操作都是不可微的，所以在这些点将采用次梯度的方式，将其次梯度定义为1。

由于多个标签的验证损失函数是线性叠加的关系，对于总体验证损失函数，也只需要将各个标签的验证损失进行先行叠加即可。

### 2.2.4 总体损失函数

结合分类损失函数与验证损失函数的总体损失函数如公式（）所示，其中μ是两种损失间的相对重要性权重，通过实验确定，详见2.3章，最后一项是损失函数中的正则化项，为了约束参数W，防止权重参数W过大，γ不易过大，一般设置为接近于0的值。总体损失梯度也是各部分损失梯度的线性叠加，训练时将总体梯度通过链式法则向前进行传播，逐渐更新各层权重。

### 2.2.5 算法描述

算法可以分为两个部分，训练过程和特征提取过程，特征提取是正向传播到第5个卷积层，然后使用第5个卷积层的特征进行后续的目标跟踪任务。训练过程是用已有的带标签的数据集，对整个初始化后的网络模型进行参数的训练更新，它分为正向传播和反向传播，正向传播是图像数据和标签通过网络的传播计算损失，反向传播是使用梯度下降的方法把损失梯度通过链式法则一层层进行反向传播，在传播的过程中更新模型参数。由于在训练过程中包含特征提取过程中的正向传播，因此，这里给出训练过程的算法描述，详见算法（）

|  |
| --- |
| **算法2.1 深度特征提取算法** |
| **输入：**初始化的网络，N个训练样本和相对应的标签  **输出：**训练好的网络参数  1: 从N个训练样本中随机选取一个批次的样本进行预处理；  2: 将该批次的样本和标签进行正向传播，计算网络各层的值，最终根据公式（）计算得到分类损失，根据公式（）计算得到验证损失；  3: 若预定总迭代次数达到或者损失小于一定阈值，则结束；否则，继续第4~6步；  4: 根据公式()计算得到分类损失梯度，根据公式()计算得到验证损失梯度；  5: 进行反向传播，用链式法则将损失梯度一层层传播下去，每一层利用各自的损失梯度，使用梯度下降算法对网络模型参数进行更新。  6: 重复第1~3步。 |

## 2.3 实验分析

### 2.3.1 实验环境

本章主要内容是如何提取更加有效的特征，与跟踪相关的内容这里将不进行描述，详见第4章。这里实验将以C-COT为基础，使用本章中训练得到的网络模型进行特征提取，然后进行目标跟踪，将结果进行对比验证。特征提取模型的相关的内容都是在Caffe框架下进行的，C-COT是在matlab中实现的。整个实验，使用了GPU运算，GPU相比于CPU更擅长高强度与高并发量的计算。主要软硬件环境如表（）所示。

|  |  |
| --- | --- |
| 表2.1 软硬件环境 | |
| **项目名称** | **配置说明** |
| CPU | Intel Xeon E5-2650 v3 |
| 内存 | 64G DDR4 2133MHz |
| 硬盘 | 4TB 7200转/分 |
| GPU | GeForce GTX TITAN X |
| 显存 | 12G |
| 操作系统 | Ubuntu 16.04 64位 |
| 实验平台 | Caffe, Matlab R2014b,MatconvNet |

### 2.3.2 数据集说明

实验用到了PETA[]，PRW[]和MOT[]这三个数据集，各个数据集的详细情况和使用原因进行如下说明。

（1）PETA

PETA是由10个公开的小规模数据集集合起来再加以标注属性而形成，整个数据集中共有19000张图像，每张图像又有61个二值属性和4个多分类属性。在实验中，将整个数据集按比例随机划分为训练集、验证集和测试集，各部分分别有9500、1900、7600张图像。使用该数据集主要是因为需要该数据集中对行人属性的标注，用该数据集训练出一个用来对行人多属性进行分类的模型，用该模型对其他数据集进行多属性标签的标注，以用来进行特征提取模型的训练。然而数据集中的属性分布极不均匀，每种属性对行人区分的权重也不同，还有些属性存在被遮挡而难以判断的问题，因此，在经过筛选后，选取了6种属性进行实验。具体属性和取值如表（）所示。

|  |  |
| --- | --- |
| 表3.1 属性及相应取值 | |
| **属性名** | **取值** |
| 性别 | 男性 女性 不确定 |
| 头发长短 | 长发 短发 不确定 |
| 上衣样式 | T恤 衬衫 外套 羽绒服 西服 其他 不确定 |
| 上衣颜色 | 黑 白 红 黄 蓝 绿 紫 棕 灰 橙 多色 不确定 |
| 裤子样式 | 长裤 短裤 长裙 短裙 其他 不确定 |
| 裤子颜色 | 黑 白 红 黄 蓝 绿 紫 棕 灰 橙 多色 不确定 |

(2) PRW

PRW数据集是用6台相机，在校园进行拍摄得到，从视频中获取到的11816帧图像，使用DPM[]检测获得34304个包围框，然后进行标注，共包含932个行人，该数据集是用在行人再识别领域的，本文中使用该数据集是为了扩大训练集。由于在目标跟踪领域中，数据集规模一般不大，直接使用目标跟踪相关的数据集进行深度特征提取模型的训练，很容易出现过拟合的现象。不仅如此，目前并没有专门针对行人的目标跟踪研究，与之对应的相关数据集也基本找不到，而一般的目标跟踪数据集中的行人目标也只占了很小一部分，因此也无法直接拿过来进行特征提取模型的训练。

(3) MOT-16

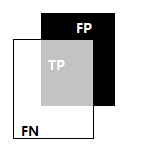
MOT数据集是用来进行多目标跟踪任务的标准数据集，该数据集中包含14个视频序列，其中7个是带有标注信息的训练集，另外7个位测试集。该数据集中主要标注目标为移动的行人和车辆，行人占了比较大的比例，因此较适合本文对行人目标跟踪任务。本文选取了MOT16-02、MOT16-09、MOT16-10、MOT16-11这4个序列作为训练集，MOT16-01、MOT16-07、MOT16-08、MOT16-12这4个序列作为测试集。使用该数据集是为了更加方便的对与目标跟踪相关的指标进行测试验证，方便与其他算法进行比较。

### 2.3.3 评价标准

为了验证本章提取特征的效果，使用目标跟踪领域广泛应用的平均重叠精度（mean Overlap precision，mOP）和自定义的特征余弦相似度分布图为标准进行比较。下面为两种评价标准的详细描述。

(1) mOP

在描述mOP前，先介绍区域重叠度（Region Overlap，RO）的概念，区域重叠度是指预测目标框与真实目标框之间的重叠程度。如图（）所示，黑色代表真实目标框，白色代表预测目标框，灰色代表重叠区域，区域重叠度即为灰色区域除总面积，如公式（）所示，其中代表真实目标框区域，代表预测目标框区域。平均区域重叠即为所有有效帧重叠精度的平均值，如公式（）所示。



假设对于某一个阈值θ，当时，则认为跟踪正确，则mOP就是跟踪正确的准确率。mOP广泛应用于目标跟踪领域，是对跟踪效果评估非常直观有效的手段。

(2) 特征余弦相似度分布图

为了更加直观观察所得提取特征的区分度，将采用特征余弦相似度分布图来对目标特征之间的关系进行刻画。对于一个特征对，其余弦相似度计算方法如公式（）所示。特征余弦相似度分布图将对同一目标和非同一目标的特征余弦相似度分布进行展示，能够非常直观的看出两者之间的分布区别，有利于阈值的选取，区分度的判断等。

### 2.3.4 实验细节

（1）行人属性标签标注

PRW与MOT-16两个数据集都没有行人属性标签，为了减少人工标注成本，在PETA数据集上训练一个行人属性分类模型。为了尽可能提高分类准确度，使用网络层次较深的VGG网络，该网络是在ImageNet[]预训练的基础上，在PETA数据集上进行微调。由于PRW与MOT-16数据集中每一个目标都有多张图像，将使用投票法的方式对每一个目标行人属性标签进行确定，这样能更加充分利用已知信息，以提高行人属性标签的准确性。

（2）数据预处理

为了是使训练的网络更好，在训练前一般需要对数据进行预处理，实验中，将使用水平镜像变换的方式将图像缩放到227×227的尺寸大小。通过将图像减去均值的方式去除光照强度对效果的影响。训练中，将以64为一个图像批次进行训练。

（3）训练过程

为了能够快速收敛，提高模型鲁棒性，使用在ImageNet数据集上与训练的模型进行初始化，在PRW与MOT-16数据集上进行微调。ImageNet是一个超大规模的数据集，图像种类丰富、场景多样，并且其中包含很多行人图像，用该数据集预训练的模型具有很好的泛化能力，对很多任务都有很好的迁移效果。

具体实验训练中，沿用了用经典模型的参数设置，微调的初始学习率设为0.001，权重衰减参数为0.0005，梯度下降加速的动量为0.9。实验中，为了是收敛速度更快，每5000次学习率衰减一次，衰减率为0.6，一共迭代20000次，每迭代1000次进行一次测试，每次将迭代100个批次。

### 2.3.5 各参数对结果的影响

实验中存在多个参数的设置问题，参数的设置将直接影响最终的效果，本文将通过实验，对每个参数具体取值进行实验，以确定最优取值。在实验中，将采用控制变量法，先固定其他值保持不变，只有一个变量发生变化，其他保持不变。由于训练数据量较大，在参数确定过程中，所有实验均进行10000次迭代。

（1）分类损失函数行人ID项权重

初始将M设置为100，β设置为0.001，μ设置为10，对α从0~1以0.2为间隔进行取值，将重叠精度准确率的阈值设置为0.5，则α与mOP变化关系如图（）所示。由图中观察得知，α取值在0.4时效果最好。

（2）验证约束项权重

由于在上面的实验中已经确定α的取值，这里将M设置为100，μ设置为10，α设置为0.4，β在0～0.1之间，以10倍为间隔取值，进行实验，结果结果如图（）所示。当值太小时，约束项不能充分发挥作用，而当值太大时，又会过于弱化其他损失约束项，取0.01时是比较合适的。

（3）验证损失函数阈值

同样的，将μ设置为10，α设置为0.4，β设置为0.01，M在50～300范围内以50为间隔进行取值，实验结果如图（）所示。实际上，M的取值与特征维度是息息相关的，这里确定M的值，对于验证损失前维度调整的值也将确定下来，降维维度为M值的一半。如图中所示，随着M的增大，mOP也缓慢上升，这是由于不同目标间的最大约束值在增加，将能够更加有效将不同目标分离开。虽然看上去M的值越大效果越好，但M值过大将使得不同目标特征相似时验证损失过大，为不影响整体结果，权衡之下将M值取为200，则降维维度为100。

（4）分类与验证损失相对权重

同上，将α设置为0.4，β设置为0.01，M设置为200，μ在2～22范围内以4为间隔进行取值，则实验结果如图（）所示。当μ非常小时，相当于验证网络，当μ非常大时，相当于分类网络，两个极端都不能有效的利用标签信息，观察结果图选择将μ设置为14。

（5）特征层选取

在特征层的选取上，本文通过对五个卷积层都进行实验进行测试。实验选取原图（layer 0）作为输入进行对比，分别对五个卷积层特征进行目标跟踪任务，结果如图（）所示。观察可看出，conv1层和conv5层效果较好，这主要是因为CNN中浅层次更多的表征图像纹理、细节特征，这相当于HOG等手工梯度特征；而更深的层次更能表征语义特征，本文中使用基于行人属性的深度特征提取，因此语义特征对于行人目标的区分度很强。中间的层次是一个有图像梯度、纹理特征到图像语义特征的转换过程，因此效果较差。这里的实验数据与deepSRDCF论文中的结果有比较大的区别，主要是在conv5层，由于deepSRDCF使用的是用ImageNet训练的AlexNet，因此其conv5层的语义特征实际上代表作分类属性，对于目标的标准能力不强，因此表现出来的conv5层的效果一般。

### 2.3.6 实验结果比较

为了验证本文中特征提取算法的有效性，用未经行人属性训练的CaffeNet（C-CNN）和经过行人属性训练的CaffeNet（ATTR-CNN）进行对比试验，对两种算法的mOP与特征余弦相似度分布进行比较。如图（）所示，是RO阈值和OP的关系曲线，随着阈值的增加，OP逐渐下降，但是用行人属性训练的特征提取模型的准确率更高，这主要是因为本文中的特征是针对行人属性进行训练得到，特征更加注重对行人属性的表达，在目标跟踪过程中，即使目标形状、姿态等发生变化，其行人的属性并不会改变，因此该特征对目标变化的抵抗力很高。

如图（）所示，是两种网络的余弦相似度分布，其中颜色标记不同的网络模型，蓝色表示本文设计的网络模型ATTR-CNN，黄色表示对比网络模型C-CNN，线条虚实标识样本集，虚线表示负样本，实现表示正样本。由图可以清晰的看出负样本由于其随机性，基本满足正态分布，且余弦相似度较小，而正样本余弦相似度一般较大。对于正负样本的区分度，ATTR-CNN对于特征的区分性远远高于C-CNN，这主要也是因为加入了行人属性这一有效标识，使得对于行人目标与背景的区分更加简单。

## 2.4 本章小结

本章首先提出目前将深度特征应用到目标跟踪任务中存在的问题，然后针对行人目标跟踪这一任务，设计出基于行人属性的特征提取算法。在CaffeNet的基础上，通过调整网络结构和使用分类加验证的损失函数，训练得到基于行人属性的特征提取模型。对其中的验证损失函数进行修正，并推导出损失梯度函数，详细给出求解过程，并且根据实验对模型中参数的最佳取值进行确定。最后，通过实验与C-CNN进行对比，基于行人属性的目标跟踪算法的准确性得到大幅度的上涨，并且验证了基于行人属性的特征具有更强的区分度，对于行人目标跟踪任务，有很好的效果。

本章的主要贡献如下：

(1) 提出一种基于行人属性的特征提取方法，并将该特征应用到行人目标跟踪任务中，以解决深度特征在行人目标跟踪任务中效果不好的问题；

(2) 对验证损失函数进行修正，添加了特征范围的约束项，使得特征每一项的值都趋近于1，以解决因特征间距离过大难以收敛的问题；

(3) 在较大规模的数据集上，对提出算法有效性进行验证，并通过实验探索参数取值对结果的影响，确定最佳取值。对实验结果进行详细的比较与分析。

# 3 基于属性的行人再识别

性别、头发长短、衣服样式与颜色等属性信息是行人的固有特征，行人短时间内在多个摄像头下穿梭，这些属性信息不会发生变化，将行人属性用于辅助行人再识别的研究，可有效减少外观变化给识别结果带来的不确定性。本章首先实现一种快速有效的行人多属性识别器，为行人再识别数据集添加属性信息，再结合这些属性标签和行人ID的标签，提出一种基于属性的行人再识别算法。

## 3.1 问题分析

计算机视觉任务中常用的局部特征属于低层视觉特征，对外观变化不够鲁棒，而且不同数据集有不同的视觉分布，度量学习对低层视觉特征的泛化能力差，可能在某个数据集上表现很好的方法，迁移到另一个数据集上就不再适用。相比起低层特征，行人属性表达的是行人的中层语义信息。图3.1是PETA数据集[42]中来自不同小数据集的行人图像与部分属性。观察可知，尽管光照强度、拍摄角度、背景环境、行人姿势都各有不同，但行人属性这种固有信息在不同风格的数据集上可以通用。

|  |
| --- |
|  |
| 图3.1 PETA数据集的行人及部分属性 |

很多研究将属性作为图像特征用于辅助其他视觉任务，特别是以行人为研究目标的任务，属性信息非常有利于行人再识别。首先，属性特征携带的大量语义信息对视觉变化鲁棒，在一定程度上化解了行人再识别中由行人自身姿势、遮挡、角度等变化带来的识别难度，比利用视觉特征更具有可区分性。其次，属性识别研究的是行人的局部结构，而行人再识别基于CNN，主要研究全局的特征描述，行人属性可以帮助全局特征获取更细致的局部信息。在当前大规模的数据集中结合属性的主要有Su等人[48]的三步法，结合语义属性与潜在属性共同学习，更侧重于属性预测和过程中的优化工作，以及Lin等人[50]利用多个属性与类标签的相关性联合建模，同时进行属性预测和行人再识别，但是他们采用手工标注行人再识别数据集的方式，低产出的工作量太大，影响算法的整体效率。

本章将行人再识别作为研究重心，属性作为辅助信息，主要目标在于借助属性对外观变化鲁棒这项优点来提高提取行人特征的鲁棒性，从而提高行人再识别算法的准确率，在实现的过程中借助行人多属性识别方法自动地为再识别数据集添加属性信息，提高行人再识别算法整个过程的效率。

## 3.2 行人多属性识别器

### 3.2.1 采用多标签分类的网络结构

行人多属性识别的任务是输入一张行人图像，对它的一组属性进行预测，例如这人是男是女？头发是长是短？衣服什么样式？什么颜色？等等。属性识别作为行人再识别的辅助工具，目标是使用越简单越好，准确率越高越好。将行人多属性识别作为多标签分类的问题，提出一种多标签属性识别网络（Multi-Label Attribute Identification Network, MLAIN）框架，直接对所有属性联合学习CNN网络，将学到的网络作为行人多属性识别器，无需用到部件、姿势、上下文信息，既能减轻训练工作量，又能取得不错的识别效果。

根据经验可知，属性之间存在一定的联系，例如绝大部分男性是短发，喜欢穿黑色、蓝色、灰色等较暗的颜色，不会穿裙子，而女性大部分是长发，服装颜色相对艳丽，有相当一部分会穿裙子。因此，应该将图像和对应的多个属性标签同时作为输入，采用层次较深的网络结构和一个整体的损失函数同时学习多个属性特征。在网络前期，特征之间共享权重，每个属性都可以有效迁移或重用其他属性的知识，识别效果更好。同时，也要考虑属性之间的差异，为了让每个属性学习自己的代表性特征，网络后期开始分裂，每个属性都有自己的高层特征fc7层、分类层和损失层。行人多属性识别网络的总体结构如图3.2所示。

|  |
| --- |
|  |
| 图3.2 行人多属性识别网络总体结构 |

作为辅助工具，对行人多属性识别的要求是准确率越高越好，因此将网络层次更深的VGGNet[23]作为基础结构。VGGNet很好地继承与发展了AlexNet的主要框架，它有五组卷积、两个全连接层和一个分类层，与AlexNet不同的是每组卷积中卷积层的数量可以根据需要进行配置，文献[23]中提供了五种不同配置，卷积层数依次递增，目标识别的性能也依次提升，但是这种提升也不是无限的，卷积层数加到16时就会到达性能瓶颈。

图3.2中的网络采用了卷积层数最多的VGG-19，fc6之前的网络结构都与原始VGG-19一样，让多个属性共享大部分参数，从fc7开始为每个属性分裂出一个全连接层，每个属性全连接层都与fc6连接，且都对应一个fc8分类层和分类损失层，分类层的结点各自对应该属性的一个类别，分类损失层用于计算分类的预测值与实际标签相同的概率，网络的总体损失是多个分类损失之和。

### 3.2.2 联合多种属性的损失函数

与大部分行人多属性识别方法不同的是，这里没有将属性作为二分类问题，而是为每个属性都配置了一个多分类的Softmax损失。假设一共有*J*个属性，每个属性有个类别，与第二章类似，属性*j*经过Softmax回归后的损失函数表示为公式(3.1)。

 (3.1)

其中，是在给定特征输入的情况下，对于属性j的预测值属于每个类别的概率。若要使该损失函数的值最小，则预测值与实际类别k相等的概率要最大。所有属性的总体损失函数表达为公式(3.2)，这里所有属性的分类损失值权重相等。在反向传播时，各分类损失层对相应的分类层求导，在f6层将各梯度的值汇合后继续向后传播。

 (3.2)

## 3.3 基于属性的行人再识别算法

### 3.3.1 采用“投票法”的属性标注

虽然上一节中的行人多属性识别器的准确率尚未达到极致，但是作为标注工具使用已经足够了。图3.3是以Market-1501数据集为例，用上一节中的多标签属性识别器进行识别标注的部分样本。由于实验数据集中每个行人都对应多张图像，对同一行人的不同图像进行属性识别，结果不一定相同，所以采用“投票法”确定其类别。对于每一个属性，同一行人的每张图像都有一张“选票”，根据该图像的预测结果为属性类别“投票”，最后将“得票数”最多的类别作为该属性最终的标签予以记录。图3.3中将每个属性得票最多的类别用黑色字体标识，得票较少的类别用红色字体标识。

|  |
| --- |
|  |
| 图3.3 采用“投票法”标注Market-1501行人属性 |

使用行人多属性识别器标注好属性标签后，最好不再添加人工干预，一来可以大大减少工作量，研究计算机视觉的意义正在于此，二来自动标注的属性更符合计算机视觉对行人属性的认知，尽管在人类认知上不如手工标注的准确，但是更有利于之后的视觉任务。下面将用该属性识别方法对行人再识别任务中用到的数据集都进行识别，把得到的属性信息作为行人再识别的辅助标签。

### 3.3.2 联合行人ID与属性的分类网络结构

基于属性的方法同样以CaffeNet作为基本结构，将属性标签作为网络的主要学习内容。当然，仅采用属性标签是不够的，即使有再全面的属性信息，也不能完整描述一个有独特身份的人，所以将行人ID标签和属性标签一起用于分类识别，行人ID也可以看作一个特殊的属性标签，这种属性对行人个体来说是唯一的。对所有属性的学习在同一个网络中完成，提出的联合行人ID与属性的分类网络总体结构如图3.4所示。

|  |
| --- |
|  |
| 图3.4 结合ID标签和属性标签的行人再识别网络结构 |

该网络的主体结构与第二章的网络完全一样，有区别的是输入多了一些属性标签，在fc7之后分裂为多条均为全连接层的支路，包括一个行人ID分类层fc8\_0（深绿色）和六个属性分类层fc8\_1~fc8\_6（浅绿色），每个分类层之后都紧随着一个计算分类预测值与实际标签相同的概率的分类损失层，各分类损失之和是总体损失，也是网络优化的目标。为了验证属性标签的有效性和行人ID标签对总体结果起到的作用，后文将对仅采用行人ID标签、仅采用属性标签、两种标签都采用这三种算法都进行验证，并将它们作为对比实验。

### 3.3.3 联合行人ID与属性的损失函数

从图3.4可以看出，针对行人ID和每个属性各有一个分类层和相应的分类损失层，所以分类损失分为两个部分，其中一个部分是行人ID分类损失，和第二章的行人分类很相似，如公式(3.3)所示。

 (3.3)

另一个部分是行人属性分类损失，和行人多属性识别很相似，公式(3.4)是属性j的损失函数，是在给定特征输入的情况下，对于属性j的预测值属于每个类别k的概率。

 (3.4)

所有标签的损失函数加权累加，构成总体损失函数，表示为公式(3.5)。

 (3.5)

其中，用于控制行人ID分类损失所占的权重，因为行人ID标签比较准确，所以从策略上考虑会为其赋予较大的权重，具体如何取值将在实验中详细讨论。该损失函数的求解方式与第二章类似，分别对相应的分类层求导，对公式(3.3)和公式(3.4)求导的结果如公式(3.6)和公式(3.7)所示，其中和分别是fc8\_0和fc8\_1~fc8\_6的值。

 (3.6)

 (3.7)

### 3.3.4 算法描述

算法的训练过程也分为前向传播和反向传播，前向传播对原始数据利用网络各层的参数计算各结点的特征值，将各属性分类的结果概率累加起来获得总体的分类损失值，反向传播将多个属性损失值分别对相应的特征值以求导的方式向后反馈，更新网络参数。网络的优化目标是使得所有属性分类的结果都尽量准确。在反复迭代的过程中，极大地优化特征提取的性能，使得特征具有足够的表达能力。训练的具体流程如算法3.1所描述。

|  |
| --- |
| **算法3.1 结合行人ID标签与属性标签的行人再识别算法（ATTR+C-CNN）** |
| **输入：**预先训练过的网络参数，*n*个训练样本和相应的行人ID标签，以及相应的6个属性标签、、、、、  **输出：**训练好的网络参数  1: 对一个训练批次的训练样本进行预处理；  2: 前向传播计算网络各层的值，根据公式(3.3)和公式(3.4)得到行人ID和各属性的分类损失值；  3: 若预定总体迭代次数未达到，则继续第4~5步；若已达到，则结束；  4: 根据公式(3.6)和公式(3.7)得到行人ID和各属性分类损失反向传播的梯度；  5: 反向逐层采用梯度下降，更新网络参数。重复第1~3步。 |

## 3.4 实验分析

### 3.4.1 行人多属性识别实验结果分析

（1）数据集与属性选择

实验采用大型行人属性数据集PETA[42]，PETA由10个公开的小规模行人再识别、跟踪、姿态估计数据集集合起来加以标注属性而形成，一共有19,000张图像，每张图像标注了61个二值属性和4个多分类属性。该数据集的10个子集风格各有不同，既具有挑战性，又非常实用。与文献[42]的基准实验设置一样，随机地将所有图像划分为三个部分，训练集、验证集、测试集分别有9,500、1,900、7,600张图像。

实际上PETA中有将近一半属性的数据分布严重不平衡，例如推婴儿车的数据极少，大约只占总数的千分之一，还有一些属性容易因为遮挡而难以判断，例如鞋子类型和颜色，所以只保留6种容易判断且具有足够区分度的属性。为了进一步使数据分布尽量均衡，实验将有关上衣的6个二值属性合并，将有关裤子的5个二值属性合并，又考虑到有些属性取值不确定或缺乏标注的问题，所以将每种属性加了一个不确定的类型，各属性及相应取值如表3.1所示。

|  |  |
| --- | --- |
| 表3.1 属性及相应取值 | |
| **属性名** | **取值** |
| 性别 | 男性 女性 不确定 |
| 头发长短 | 长发 短发 不确定 |
| 上衣样式 | T恤 衬衫 外套 羽绒服 西服 其他 不确定 |
| 上衣颜色 | 黑 白 红 黄 蓝 绿 紫 棕 灰 橙 多色 不确定 |
| 裤子样式 | 长裤 短裤 长裙 短裙 其他 不确定 |
| 裤子颜色 | 黑 白 红 黄 蓝 绿 紫 棕 灰 橙 多色 不确定 |

（2）实验细节

数据预处理将训练集图像缩放到256\*256的大小，在左上角、右上角、左下角、右下角、正中心各裁剪227\*227大小的图像块，结合水平镜像变换的方式，将原来的一张图像扩充为十张，同样要在输入网络之前减去图像均值。根据VGGNet的一般设置，训练集与验证集以50个图像为一个批次，动量为0.9，权重衰减参数为0.0005，微调学习率为0.001，每迭代9000次将学习率降为之前的0.1倍，一共迭代50000次。

行人多属性识别是多标签问题，这里每个样本对应6个标签，但是Caffe中无论是lmdb格式还是读文本文件的方式，都只提供了单标签的解决方案。实验将图像和6个属性标签存储在一个lmdb文件中一起作为输入，在写网络结构的prototxt文件中用一个slice层将不同属性的标签分开，分别计算分类损失值。

（3）实验结果与分析

实验采用的评价标准是每个属性在测试集上的识别准确率，由于所选属性的分布比较均匀，无需对各类别添加权重，实验结果如表3.2所示。用来对比的两个方法都根据所用到的属性类别做了调整，ikSVM是文献[42]中提供的基准方法，是一种改进的SVM分类器，而MLCNN[49]是多标签的CNN方法，也是在一个网络中同时学习所有属性，但对图像进行了复杂的分块，而且网络层次较浅。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 表3.2 不同方法在PETA上的行人多属性识别结果 | | | | | | |
| **方法** | **性别** | **头发长短** | **上衣样式** | **上衣颜色** | **裤子样式** | **裤子颜色** |
| ikSVM[42] | 78.45 | 78.45 | 70.72 | 76.14 | 68.43 | 78.32 |
| MLCNN[49] | 84.34 | 87.53 | 88.75 | 90.59 | 85.20 | **84.86** |
| MLAIN | **91.03** | **92.50** | **93.59** | **91.25** | **89.81** | 83.38 |

由表中结果可以看出，MLCNN方法相比ikSVM方法识别率有大幅提升，MLAIN方法又比MLCNN识别率有一定提升，而且对于绝大部分属性皆是如此。ikSVM的浅层特征显然表达能力不足，因此识别率不如基于CNN的方法。MLCNN虽然也用CNN同时学习多个属性，但是没有MLAIN层次更深的网络所提取的特征表达能力好，而且MLCNN将人体分块，同时运行16个CNN进行测试，对GPU的计算能力要求太高，还非常耗时，MLAIN用一个更深的网络即可达到更优的识别效果，而且使用更方便，可以作为其他行人数据的属性识别器使用。

### 3.4.2 参数对结果的影响

公式(3.5)中的用于控制行人ID分类损失占总体损失的权重，为了选择一个合适的值，在Market-1501和PRW上对不同的值进行比较实验，结果如图3.5所示，通过观察，发现值不宜过大也不宜过小，在2~3之间时效果最好，表示行人ID分类损失的权重应该设置为属性分类损失的2~3倍，考虑到行人的ID标签相对于属性标签来说更为明确，所以应该为其赋予较大的权重，实验对的取值为3。注意到当值取0的时候，也就是完全不考虑行人ID标签，而全部依靠属性，结果显示这样比添加行人ID标签的结果都要差，因此，在采用属性的行人再识别算法中，行人ID的标签对提高再识别率是至关重要的。

|  |
| --- |
| 图3 |
| 图3.5 行人ID分类损失占总体损失的权重对结果的影响 |

### 3.4.3 与本文方法的比较

实验同样将行人图像缩放到227\*227大小，只用水平镜像变换方式扩充，以64个图像为一个批次，在第二章中的C-CNN网络基础上微调训练。每迭代5000次学习率降为之前的0.1倍，一共迭代20000次。将属性算法（ATTR-CNN）、属性加ID标签算法（ATTR+C-CNN）与第二章的三个算法进行比较，在一个图中绘出CMC曲线，Market-1501和PRW上的CMC曲线分别如图3.6和图3.7所示。

从CMC@1的比较结果来看，属性算法和属性加ID算法的准确率都不如第二章中的分类加对比验证算法，但从整体趋势来看，在Market-1501上，属性加ID算法的性能和分类加对比验证算法差不多，而在PRW上属性加ID算法的整体性能甚至更好，这是因为在采集数据集的校园内，行人目标各属性相同的概率较大，采用属性方法很容易因为属性相同而产生误判，在返回的图像要求较少时更容易出错。而且加入了属性的算法确实比仅用行人ID的C-CNN算法好，可见，属性信息对提高行人再识别率是有帮助的。属性加ID算法尚有待提升的地方，尤其是结合分类与对比验证的方向，是应该重点研究的策略。此外，属性加ID的算法比仅用属性的算法好，行人ID标签的作用不可忽略。

|  |
| --- |
| 图3 |
| 图3.6 Market-1501的CMC曲线图 |
| 图3 |
| 图3.7 PRW的CMC曲线图 |

### 3.4.4 与现有方法的比较

实验对仅采用属性标签的算法（ATTR-CNN）和结合属性标签与行人ID标签的算法（ATTR+C-CNN）分别进行训练与测试，此外还有仅用ID标签的算法，也就是第二章中提出的C-CNN，将结果与现在报告出来的多个结合属性或其他辅助项的深度学习方法、第二章中提出的算法及其他相关方法进行比较，Market-1501和PRW的结果分别如表3.3和表3.4所示，与第二章中列出的采用分类结合验证方法的Wang[54]、DLCE[55]等方法相比，有一定的优势。

|  |  |  |
| --- | --- | --- |
| 表3.3与其他方法在Market-1501上的实验对比 | | |
| **方法** | **CMC@1** | **mAP** |
| Wang[54] | 58.3 | 34.0 |
| DLCE[55] | 62.1 | 39.6 |
| SSDAL[48] | 39.4 | 19.6 |
| JLSLA[49] | 47.4 | 21.1 |
| APR[50] | 57.5 | 32.9 |
| CAN[33] | 48.2 | 24.4 |
| LSTM-Siamese[32] | 61.6 | 35.3 |
| Gated SCNN[35] | 65.9 | 39.6 |
| C-CNN | 57.9 | 32.6 |
| V-CNN | 57.1 | 33.6 |
| CV-CNN | **67.5** | **42.0** |
| ATTR-CNN | 61.9 | 37.2 |
| ATTR+C-CNN | 64.8 | 40.4 |

SSDAL[48]先从属性数据集中学习深度属性特征，再用行人跟踪数据集的行人ID标签组成属性三元组进行网络微调，最后在行人再识别数据集上测试，JLSLA[49]结合语义属性与潜在属性共同学习，APR[50]利用多个属性与类标签的相关性联合建模，同时进行属性预测和行人再识别。在Market-1501上，属性加ID算法ATTR+C-CNN的CMC@1达到了64.8%，mAP达到了40.4%，优于表中结合属性的方法，因为在深度学习中结合属性的行人再识别研究还处于初级阶段，这些方法要么仅通过微调的方式迁移属性信息，要么手工标注少量的属性标签，所以表现不是特别好。

CAN[33]和LSTM-Siamese[32]均是在验证网络中融合LSTM的方法，Gated SCNN[35]在每个卷积层之后插入控制函数，用于获取一对测试图像之间的微小差异，注意到在CMC@1评价标准上，实验在Market-1501上比Gated SCNN低1.1%，这是因为Gated SCNN的网络结构经过了精心设计与控制，而属性加ID算法ATTR+C-CNN采用的是网络结构简单的CaffeNet，尽管如此，实验结果在mAP评价标准上比Gated SCNN高0.8%，说明属性加ID算法在保证准确率的同时也重视召回率，只有准确率和召回率都足够高的时候，才能说行人再识别的性能足够好。

|  |  |  |
| --- | --- | --- |
| 表3.4与其他方法在PRW上的实验对比 | | |
| **方法** | **CMC@1** | **mAP** |
| Wang[54] | 53.1 | 29.2 |
| DLCE[55] | **55.3** | 29.6 |
| SSDAL[48] | 37.6 | 14.4 |
| APR[50] | 49.9 | 26.2 |
| Gated SCNN[35] | 53.9 | 31.6 |
| Re-ranking[58] | 52.5 | 31.2 |
| C-CNN | 49.8 | 26.0 |
| V-CNN | 50.6 | 26.4 |
| CV-CNN | 55.1 | 30.4 |
| ATTR-CNN | 53.4 | 30.7 |
| ATTR+C-CNN | 54.6 | **31.9** |

对于PRW上的结果，Re-ranking[58]是将行人再识别作为检索问题重点研究重新排名的方法。由多个方法的比较结果可见，与第二章中列出的采用分类结合验证方法、结合属性或其他技巧的方法相比，属性加ID算法ATTR+C-CNN也有一定的性能优势，在PRW上的CMC@1达到了54.6%，mAP达到了31.9%，优于表中结合属性的方法。虽然在CMC@1上比DLCE低0.7%，但是在mAP上要高1.3%，说明属性加ID算法返回的第一个结果准确率不一定高，但是返回多个结果的总体准确率较高，更符合实际使用的规则，而且结合属性与深度学习的行人再识别还有研究空间。

### 3.4.5 用于属性识别的结果

属性加ID算法ATTR+C-CNN利用属性信息辅助行人再识别的性能，但从某种程度上来说，行人的ID信息也辅助了属性识别的性能，二者相辅相成，共同提高。结合属性和ID标签的算法可以对Market-1501和PRW中的候选集进行属性识别，实验中，将行人多属性识别器所识别的结果作为正确标签，将只用属性标签、没有加入行人ID标签的算法ATTR-CNN作为对比方法，结果分别如图3.8和图3.9所示。

|  |
| --- |
| 图3 |
| 图3.8 Market-1501用于属性识别的结果 |
| 图3 |
| 图3.9 PRW用于属性识别的结果 |

由结果观察可知，在这两个数据集上，ATTR+C-CNN在大部分属性上的表现比没有行人ID的ATTR-CNN好，可见，属性加ID行人再识别网络还可以用于识别行人属性，而且行人的ID信息对属性识别也起到了一定的辅助作用，比仅采用属性的方法更好。

## 3.5 本章小结

本章首先阐明了行人多属性识别与行人再识别之间的联系，属性标签对行人再识别中较难处理的行人自身变化问题具有鲁棒性，很适合用于行人再识别的研究，但行人再识别数据集中缺乏属性标签，所以先实现了一种行人属性识别器，为行人再识别数据集添加属性信息，再分别考虑是否需要结合行人ID标签，对多个标签同时采用分类的网络结构设计，最后的实验结果显示属性加ID的算法超过了同类研究的结果，与一部分性能较好的其他方法相比，也有一定的竞争力，但和第二章的算法相比，还有提升的空间。

本章的主要贡献如下：

（1）基于多标签分类的方法与VGGNet结构提出了一种快速有效的行人多属性识别网络，在大型行人属性数据集PETA上进行了实验，并将其作为属性识别器使用，为行人再识别数据集标注属性标签；

（2）基于CaffeNet提出一种多标签的行人再识别框架，在一个网络中同时学习行人ID和多种属性的特征，使得网络达到总体最优，该再识别框架还能应用于行人属性识别；

（3）在大型数据集Market-1501和PRW上进行充分的实验，探索两种标签权重的取值对结果的影响，分别给出仅用属性标签和结合行人ID与属性标签这两种算法的实验结果，并与基于属性或采用了其他策略的方法、第二章算法进行对比。

# 4 基于多属性与多策略融合的行人再识别

第二章以行人身份为学习对象，分类识别方法特征表达能力强，对比验证方法能充分利用行人之间的关系信息，所以结合二者的应用优点并采用约束对比验证损失函数优化一般的验证方法，提出了一种联合学习分类损失与对比验证损失的行人再识别算法，比单独采用一种策略都要好。第三章研究了对各种外观变化具有鲁棒性的属性信息，将属性用于辅助行人再识别的研究，基于分类识别网络，提出了一种结合属性标签与行人ID标签学习的行人再识别算法，可有效减少外观变化带来的不确定性。本章将结合第二章的多种学习策略、第三章中的多种属性标签，提出一种联合学习框架，进一步提高行人再识别的准确率。

## 4.1 问题分析

第二章和第三章从两种角度分别提出了不同的行人再识别算法，都取得了不错的效果，但这两种算法依然有各自的不足之处。这里首先对第二章和第三章用到的多种方法进行梳理，图4.1中列出了本文研究的所有基础方法与它们的组合关系。

|  |
| --- |
|  |
| 图4.1 本文研究的所有方法 |

第二章中提出的结合CNN分类与对比验证的算法，虽然能够很好地挖掘行人自身的特征表示和行人图像之间的关系，但是仅采用行人的ID标签是比较单薄的，不能很好地抵抗行人自身的姿势、角度、遮挡等变化带来的影响，而且对于行人对比验证来说，在一个训练批次的样本数量不多而类别数量很多的时候，对同批次中的样本两两组合，负样本的数量会远远多于正样本的数量，样本的分布不平衡，影响最终结果。

第三章中提出的结合行人属性和ID标签的CNN分类算法，能够在一定程度上抵抗行人自身变化带来的影响，解决第二章算法的问题，但是，添加属性信息也不是万能的，当两个行人有大量属性相同时，特征之间的区分不够明显，就会给识别结果带来负面影响，所以还需要从其他方面进一步改进。

因此，将第二章和第三章中提出的算法结合起来，融合之前研究中用到的所有标签（ID标签、ID关系标签、属性标签、属性关系标签）和所有策略（分类方法、对比验证方法）提出一个多任务的算法，在一个端到端的网络中同时计算行人ID的分类损失、对比验证损失和属性的分类损失、对比验证损失，将其简称为融合算法。融合算法既能抵抗由拍摄条件带来的分辨率、光照强度、背景、缩放程度不同等问题，又能抵抗由行人自身导致的姿势、角度、遮挡变化的影响，而且各属性类别都较少，两两组合的样本较为均衡。

## 4.2 基于多属性与多策略融合的行人再识别算法

### 4.2.1 融合多属性多策略的网络结构

与前两章的算法一样，融合算法也将CaffeNet作为基本结构，将训练图像成批次地输入网络，无需提前进行组合，在验证损失层才组合同批次的图像，学习图像之间的距离关系。属性标签和行人ID标签一样，既用于分类识别，又用于验证关系。ID标签的分类与对比验证、属性标签的分类与对比验证全都在同一个端到端的网络中完成。

采用了多属性、多策略的CNN网络的总体结构如图4.2所示，该网络的主体结构与前两章一样，在fc7之后分为多个全连接层的支路。其中有一半是结点数与类别数相同的分类层（图中绿色方块），包括一个行人ID分类层（深绿色）和6个属性分类层（浅绿色），每个分类层之后都紧随着一个计算分类预测正确概率的分类损失层。另外一半是用于降低fc7特征维度、方便度量特征间距离的降维层（图中紫色方块），包括一个行人ID降维层（深紫色）和6个属性降维层（浅紫色），每个降维层之后都紧随着一个计算特征之间相似关系误差的验证损失层。

|  |
| --- |
|  |
| 图4.2 采用了多属性、多策略的CNN网络结构 |

### 4.2.2 融合损失函数

从图4.2可以看出，该网络结构对行人ID和每个属性各有一个分类层和降维层，以及相应的分类损失层和验证损失层，下面分别介绍所用的分类损失函数和对比验证损失函数。

分类损失部分和上一章基于属性的方法相似，为了简化表达，将行人ID和属性标签一起描述。一共有*J*+1种标签，每种标签有个类别，当*j*=0时表示行人ID标签。在一个包含*N*张图像的训练批次中，标签*j*的平均损失函数表达为公式(4.1)。是给定特征输入对于属性*j*的预测值属于每个类别*k*的概率。

 (4.1)

所有标签的总体分类损失函数如公式(4.2)所示。

 (4.2)

其中，用于控制行人ID分类损失占总体分类损失的权重，第三章对其取值进行了详细的讨论，这里按照讨论结果将其取为3，其他所有属性的权重取相等的值。该损失函数的求解方式也与第二章类似，分别对相应的分类层求导，对公式(4.1)求导的结果如公式(4.3)所示。

 (4.3)

对比验证损失部分的训练目标对行人ID标签来说是使同一行人图像特征距离较小，不同行人图像特征距离较大，对于属性标签来说则是使得有相同属性的图像特征距离较小，不同属性的图像特征距离较大，多个目标要同时达到最优化。基于此目的构造对比损失函数，对于属性标签*j*，某批次中第*m*个图像特征对之间的平均验证损失函数如公式(4.4)所示，一共有*J*+1种标签，当*j*=0时表示这是ID标签。

 (4.4)

其中，是边界阈值参数，采用的度量方式是欧氏距离。每种标签的类别只有两种取值，是否属于同一行人或是否具有同样的属性，1表示是，0表示否。和的取值在第二章有详细分析，这里也采用同样的取值，取值为200，取值为0.01。公式(4.4)既惩罚同一行人图像特征对和相同属性图像特征对之间的距离较大的情况，又惩罚不同行人图像特征对和不同属性图像特征对之间的距离小于阈值的情况，同时还约束特征值的分布。一个训练批次中*M*个图像对的平均损失函数如公式(4.5)所示。

 (4.5)

所有标签的总体对比验证损失函数为公式(4.6)。

 (4.6)

其中，是用于控制行人ID的对比验证损失所占权重的参数，虽然不能与公式(3.5)中ID分类损失的权重完全等同，但是为了实验方便，这里和公式(3.5)中的一样取为3，其他所有属性的权重取相等的值。同样采用次梯度的方式对公式(4.4)求导，公式(4.4)中的三项表示为t1、t2、t3，对它们分别求导如公式(4.7)所示。其中，i的取值为1或2，公式(4.4)的总体导数为三项之和。

 (4.7)

同样地，是指示函数。最后，公式(4.8)是结合了ID和属性等多种标签、分类和对比验证等多种策略的总体损失函数，算法的优化目标是使这个总体损失函数值最小。

 (4.8)

其中，用于控制所有属性的分类损失之和占总体损失的权重，第二章对其取值进行了详细的讨论，参数的选择在相似的任务之间是可以迁移的，所以按照讨论结果在Market-1501上将其取为20，在PRW上将其取为6。

### 4.2.3 算法描述

提出的算法在训练过程中同样分为前向传播和反向传播，前向传播对原始数据利用网络参数计算各结点的特征值，将各标签的分类结果与实际标签比较获得分类的损失值，对组合而成的成对图像特征进行距离度量获得约束对比验证损失值。反向传播将上述多种损失值各自以求导的方式向后反馈，更新网络参数。多种损失的反向传播是分别对相应的特征值求导，例如分类损失是分别对fc8\_0~fc8\_6求导，而对比验证损失是分别对ip0~ip6求导，在反向传播到fc7时再按总体损失中各部分的权重对这些导数值求和后继续反向传播。训练的具体流程如算法4.1所描述。

|  |
| --- |
| **算法4.1 多属性与多策略融合的行人再识别算法（ATTR+CV-CNN）** |
| **输入：**预先训练过的网络参数，*n*个训练样本和相应的行人ID标签，以及6个属性标签、、、、、  **输出：**训练好的网络参数  1: 对一个训练批次的训练样本进行预处理；  2: 前向传播计算网络各层的值，根据公式(4.1)得到行人ID和各属性标签的分类损失值，根据公式(4.5)得到行人ID和各属性标签的对比验证损失值；  3: 若预定总体迭代次数未达到，则继续第4~5步；若已达到，则结束；  4: 根据公式(4.3)得到分类损失反向传播的梯度，根据公式(4.7)得到验证损失反向传播的梯度；  5: 反向逐层采用梯度下降，更新网络参数。重复第1~3步。 |

## 4.3 实验分析

### 4.3.1 与本文方法的比较

实验细节设置与第三章一样，将融合算法ATTR+CV-CNN与第二章、第三章中的算法进行比较，在一个图中绘出CMC曲线。Market-1501和PRW的CMC曲线分别如图4.3和图4.4所示。从CMC@1的结果来看，融合算法的性能已经远远超过了第二章、第三章提出的所有算法，与之前提出的性能最好的CV-CNN相比，在Market-1501和PRW两个数据集上的CMC@1分别超过了2.5%和3.2%，可见，融合算法非常适用于复杂的多摄像头行人再识别任务。从CMC曲线来看，融合算法在整体趋势上也比第二章、第三章的算法有更大的优势。

在PRW上的结果总体上比Market-1501低5~8%，因为PRW中的数据复杂性更高，类内差异更大。随着所取的Rank数增加，到Rank-50时多属性多策略融合的算法匹配率在两个数据集分别上升到97%和91%左右，非常适合实际应用。

|  |
| --- |
| 图4 |
| 图4.3 Market-1501的CMC曲线图 |
| 图4 |
| 图4.4 PRW的CMC曲线图 |

### 4.3.2 与现有方法的比较

实验对提出的结合了属性标签的分类与对比验证、行人ID标签的分类与对比验证的融合算法（ATTR+CV-CNN）进行了训练与测试，并将结果与第二章、第三章中提出的算法以及用于比较的方法进行充分对比，Market-1501的结果如表4.1所示。与Wang[54]、DLCE[55]等采用分类结合验证的方法相比，融合算法有压倒性的性能优势。SSDAL[37]和JLSLA[49]、APR[50]是结合了属性的深度学习方法，前两者对属性的利用方式是直接在属性识别网络上微调行人再识别网络，对于属性信息的保持不是很好，而APR[50]在方法上与属性算法很相似，但他们用手工方式为再识别数据集标注标签，工作量大且不一定准确。CAN[33]和LSTM-Siamese[32]在验证网络中融合LSTM，Gated SCNN[35]精心设计与控制其网络结构。由结果可知，融合算法的CMC@1达到了70.0%，mAP达到了45.7%，在这两种评价标准上都比表中所有的方法要好，比目前准确率最高的Gated SCNN在CMC@1上高4.1%，在mAP上高6.1%，说明提出的融合算法有效地将多种属性和多种策略结合在了一起，可以很好地利用它们的优点。

|  |  |  |
| --- | --- | --- |
| 表4.1 与其他方法在Market-1501上的实验对比 | | |
| **方法** | **CMC@1** | **mAP** |
| Wang[54] | 58.3 | 34.0 |
| DLCE[55] | 62.1 | 39.6 |
| SSDAL[48] | 39.4 | 19.6 |
| JLSLA[49] | 47.4 | 21.1 |
| APR[50] | 57.5 | 32.9 |
| CAN[33] | 48.2 | 24.4 |
| LSTM-Siamese[32] | 61.6 | 35.3 |
| Gated SCNN[35] | 65.9 | 39.6 |
| C-CNN | 57.9 | 32.6 |
| V-CNN | 57.1 | 33.6 |
| CV-CNN | 67.5 | 42.0 |
| ATTR-CNN | 61.9 | 37.2 |
| ATTR+C-CNN | 64.8 | 40.4 |
| ATTR+CV-CNN | **70.0** | **45.7** |

同样地，与其他方法在PRW上的比较结果如表4.2所示。由比较结果可见，在PRW上与第二章中列出的采用分类结合对比验证的方法、第三章中列出的结合属性或其他技巧的方法相比，融合算法都有很大的性能优势。在PRW上，融合算法的CMC@1达到了58.3%，mAP达到了35.1%，优于表中列出的所有方法，比目前准确率最高的Gated SCNN在CMC@1上高4.4%，在mAP上高3.5%，说明提出的多属性多策略融合算法在PRW上也非常有效。

|  |  |  |
| --- | --- | --- |
| 表4.2 与其他方法在PRW上的实验对比 | | |
| **方法** | **CMC@1** | **mAP** |
| Wang[54] | 53.1 | 29.2 |
| DLCE[55] | 55.3 | 29.6 |
| SSDAL[48] | 37.6 | 14.4 |
| APR[50] | 49.9 | 26.2 |
| Gated SCNN[35] | 53.9 | 31.6 |
| Re-ranking[58] | 52.5 | 31.2 |
| C-CNN | 49.8 | 26.0 |
| V-CNN | 50.6 | 26.4 |
| CV-CNN | 55.1 | 30.4 |
| ATTR-CNN | 53.4 | 30.7 |
| ATTR+C-CNN | 54.6 | 31.9 |
| ATTR+CV-CNN | **58.3** | **35.1** |

### 4.3.3 摄像头关系研究

融合算法的效果有很大一部分来源于查询图像与候选集图像来自同一个摄像头的情况，在视觉分布相似的图像之间，识别准确率自然比较高。为了进一步理解融合算法良好性能的来源，这里给出摄像头之间的行人再识别结果，图4.5和图4.6分别是Market-1501和PRW的6个不同摄像头之间的CMC@1和mAP，显然同一个摄像头之间的识别率最高。从整体上看，摄像头之间的再识别准确率比较均衡，去除同一摄像头下的识别结果后，在两个数据集上的CMC@1和mAP分别有38.1%和32.3%、36.3%和28.1%，也是不错的性能表现。

|  |
| --- |
|  |
| 图4.5 Market-1501摄像头之间的CMC@1和mAP |
|  |
| 图4.6 PRW摄像头之间的CMC@1和mAP |

### 4.3.4 结果展示

下面将提出的三种行人再识别算法应用于实际，测试时提取特征大约需要4.82秒/100张图像，计算距离与排名大约需要0.11秒/100张图像，可以满足实际应用的需求。在Market-1501和PRW的查询集中分别随机选择3张图像，采用融合算法ATTR+CV-CNN、第三章的属性算法ATTR+C-CNN、第二章的分类加验证算法CV-CNN进行再识别，在候选集中找出与其距离得分最高的8个图像，结果分别如图4.7和图4.8所示，正确的匹配用蓝色框标出，错误的匹配用红色框标出。

|  |
| --- |
|  |
| 图4.7 三种算法在Market-1501上的效果展示 |
|  |
| 图4.8 三种算法在PRW上的效果展示 |

由结果可以看出，匹配错误的图像与查询图像在外观上很相似，而且显然采用多属性多策略融合的算法比第二章的分类加验证算法CV-CNN、第三章的行人ID加属性标签分类算法ATTR+C-CNN的表现更好。

## 4.4 本章小结

本章首先对第二章和第三章中所用的多种策略进行了梳理，并总结了这些策略的不足之处，然后融合这两章中所有的基本方法，对行人ID标签和多个属性标签都采用同时分类与对比验证的网络结构设计，实验结果显示，结合多种属性和多种策略的算法进一步提高了行人再识别的准确率，已经超过了绝大多数报告出来的较好的方法和第二章、第三章提出的所有算法，从该网络对查询图像检索其正确匹配的实例也可以看出，该算法操作简单且不失效率，可以很好地应用于实际中。

本章的主要贡献如下：

（1）基于CaffeNet提出融合行人ID、多种属性和分类、对比验证的等多种策略的行人再识别框架，在一个网络中既学习行人ID和多种属性具有区分性的特征，又约束它们之间的关系；

（2）在大型数据集Market-1501和PRW上进行实验，与现有的采用了各种策略的方法、第二章与第三章的算法分别进行了对比，并用实例说明了该框架的有效性。

# 5 总结与展望

## 5.1 全文工作总结

行人再识别技术主要研究如何在多个不同摄像头得到的大量行人图像中，准确、快速地找到与查询图像属于同一行人的目标。在实际应用中，由于分辨率、光照、背景等拍摄条件变化和姿态、角度、遮挡等行人自身变化的影响，行人的视觉外观差异较大，所以行人再识别任务非常具有挑战性。为了更好地适应监控视频中复杂的外观变化，本文基于深度卷积神经网络，从行人身份识别、行人身份对比验证、行人多属性识别等多种研究角度着手，提出了三种新颖的行人再识别算法，并对它们分别在大型数据集Market-1501和PRW上进行了有效性验证。本文的主要研究工作与贡献如下：

（1）设计与实现基于分类与对比验证的行人再识别算法

提出了一种约束对比验证损失函数，用于在对比验证时约束特征值和特征间距离的大小，对提高行人对比验证的准确率有很大作用。再结合行人的分类与对比验证的优点，提出了一种联合的深度CNN学习框架，在一个网络中同时学习行人自身极具表达能力的CNN特征与行人图像高层特征之间的相似性关系，设计优化目标同时最小化分类损失和对比验证损失，在一定程度上解决光照强度、分辨率、平移缩放、姿态变化等较小视觉变化造成的问题。验证的结果显示该联合分类与对比验证的算法有效提高了行人再识别的性能，与其他方法相比具有一定的竞争力。

（2）设计与实现基于属性的行人再识别算法

提出了一种用于行人多属性识别问题的多标签属性识别网络（MLAIN），基于层次较深、表达能力较好的结构，在现有属性数据集上训练可同时识别多种属性标签的深度网络结构，为每个属性构建一个分类损失函数，优化使这些损失函数之和最小。在PETA数据集上的实验结果证明了MLAIN方法的优越性，可用于为行人再识别数据集添加属性标签信息，辅助行人再识别任务。本文还将行人属性标签与ID标签结合起来，共同学习一个同时实现行人多属性识别和行人再识别的网络框架。验证结果显示，结合属性的算法也是一个有效的行人再识别研究思路。

（3）设计与实现多属性与多策略融合的行人再识别算法

将行人ID、属性和基于CNN分类、对比验证的行人再识别有机结合起来，在分类加对比验证方法的基础上加入属性信息，对多个融合的属性都采用同时分类与验证的网络结构与损失函数，在一个端到端的网络中，既学习行人ID和各属性的深度特征，又学习行人之间和属性之间的关系，有效减少了拍摄角度、背景变化、部分遮挡等较大的外观变化给识别方法带来的不确定性。实验结果显示，采用多属性与多策略融合的算法进一步提高了行人再识别的准确率，已经超过了绝大多数报告出来的性能较好的方法，非常适合应用于实际中。

## 5.2 未来工作展望

虽然本文提出的三个行人再识别算法在大型数据集上比现有算法有一定的性能优势，但是本文的研究工作还存在局限与不足，未来可以从如下几个方面入手研究：

（1）对辅助信息的进一步研究

考虑到现有行人属性数据集中存在样本类别分布不平衡的问题，只选择研究了少量属性的识别问题，下一步可以加入更多的属性，采用为类别加权重的方式平衡样本分布。此外，还可以加入部件、角度等辅助信息进一步提高行人再识别率。

（2）对更深的CNN结构的研究

为了方便横向比较，采用了结构简单、容易微调的CaffeNet作为基础结构，但其实通常情况下，网络层次越深，学到的知识越深刻，特征表达能力越好，适合实际应用，下一步可以将GoogLeNet、ResNet等层次更深的网络作为基础结构。

（3）对特征度量方式的研究

与很多采用深度学习的方法一样，将研究重点放在特征提取网络结构上，在测试使用时采用简单的欧氏距离对特征进行度量排序，得到最佳匹配结果，实际上还应该研究与深度特征配合较好的特征度量方法。

（4）与其他深度学习方法结合的研究

目前深度学习研究领域有许多前沿方法，例如可挖掘图像时空关系的长短期记忆网络、可生成大量样本并放大样本间关系约束的生成对抗网络，这些方法都可以与本文中的CNN结合，是下一阶段的研究重点。

# 致 谢

转眼间我已在华中科技大学度过了七个春秋，母校美丽的校园环境、温暖的人文情怀、严谨的治学态度伴随了我人生中最美好的青春岁月，我为自己的华科人身份感到骄傲。在视觉计算与智能认知实验室攻读硕士学位的三年时间里，我的知识储备越来越丰富，动手实践能力也得到了快速提高，这是一生都享用不尽的财富。在求学生涯即将结束之际，我要感谢老师、同学、家人和朋友这些年对我的帮助。

首先特别感谢我的导师凌贺飞教授，凌老师对学生秉持着“严师出高徒”的要求，非常关心大家的科研情况，在周会时总能对每个学生的不同任务提出有针对性的建议，并且经常组织学术交流活动，让大家互相了解不同的研究方向，拓展学术眼界。在完成毕业设计的过程中，凌老师经常给我指导性的意见，对我起到了很大的鼓励与督促作用。我十分钦佩凌老师认真的工作态度和深厚的学术水平，衷心感谢凌老师三年来的辛勤授业与耐心指导。感谢李平老师，李老师和蔼可亲，负责组织实验室的活动，让大家在紧张的科研之余放松心情，为实验室营造出良好的学习与科研氛围。

还要感谢实验室的欧新宇、雷洁博士，在刚进入实验室时就为我这三年的学习指明了道路，特别是欧新宇师兄在我的毕业设计中给予了许多帮助，从选题到方案的设计与实现，他一直为我提供指导，不厌其烦地答疑解惑。感谢李深、杨力两位师兄，他们教会了我很多工程技术上的技能。感谢同级的李叶、赵航、马冬冬、施剑同学，我们虽然研究方向各不相同，但是常有交流，三年来我们共同进步。另外还要感谢余成跃、陆竭、罗巍、唐堃等师弟师妹，感谢他们在科研与项目中给我的大力支持。

感谢我的父母家人，他们用辛苦换来我无忧无虑的成长，对我的学业鼎力支持，在我完成毕业设计时一直给予我真挚的关心与不断的鼓励，是我前进路上最大的动力。还要感谢陪伴我学习生活的同学室友们，感谢你们带来的欢乐与美好。

最后，衷心感谢评审老师对本论文的悉心指正。

# 参考文献

1. Gheissari N, Sebastian T B, Hartley R. Person Re-identification Using Spatiotemporal Appearance. in: 2006 IEEE Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE Computer Society, 2006. 1528~1535
2. Zheng L, Shen L, Tian L, et al. Scalable Person Re-identification: A Benchmark. in: 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE Computer Society, 2015. 1116~1124
3. Farenzena M, Bazzani L, Perina A, et al. Person Re-identification by Symmetry-Driven Accumulation of Local Features. in: 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE Computer Society, 2010. 2360~2367
4. Gray D, Tao H. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. in: Computer Vision - ECCV 2008. Marseille, France: Springer, 2008. 262~275
5. Prosser B, Zheng W S, Gong S, et al. Person Re-identification by Support Vector Ranking. in: British Machine Vision Conference. Aberystwyth, UK: British Machine Vision Association, 2010. 1~11
6. Mignon A, Jurie F. PCCA: A New Approach for Distance Learning from Sparse Pairwise Constraints. in: 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Computer Society, 2012. 2666~2672
7. Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971~987
8. Zhao R, Ouyang W, Wang X. Unsupervised Salience Learning for Person Re-identification. in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA: IEEE Computer Society, 2013. 3586~3593
9. Lowe D G. Distinctive Image Features From Scale-Invariant Keypoints. International Journal of Computer Vision, 2004, 60(2): 91~110
10. Hamdoun O, Moutarde F, Stanciulescu B, et al. Person Re-identification in Multi-Camera System by Signature Based on Interest Point Descriptors Collected on Short Video Sequences. in: 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras. Stanford, USA: IEEE, 2008. 1~6
11. Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features. in: Computer Vision - ECCV 2006. Graz, Austria: Springer, 2006. 404~417
12. Li Z, Chang S, Liang F, et al. Learning Locally-Adaptive Decision Functions for Person Verification. in: 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE Computer Society, 2013. 3610~3617
13. Liao S, Hu Y, Zhu X, et al. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015. 2197~2206
14. Zhang Y, Li B, Lu H, et al. Sample-Specific SVM Learning for Person Re-identification. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 1278~1287
15. Chen D, Yuan Z, Chen B, et al. Similarity Learning with Spatial Constraints for Person Re-identification. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 1268~1277
16. Koestinger M, Hirzer M, Wohlhart P, et al. Large Scale Metric Learning from Equivalence Constraints. in: 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Computer Society, 2012. 2288~2295
17. 李航. 统计学习方法. 第一版. 北京: 清华大学出版社, 2012. 55~153
18. LeCun Y, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 1998, 86(11): 2278~2324
19. Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks. Science, 2006, 313(5786): 504~507
20. Deng J, Dong W, Socher R, et al. ImageNet: A Large-Scale Hierarchical Image Database. in: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009. 248~255
21. Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks. in: Advances in Neural Information Processing Systems. Lake Tahoe, Nevada: NIPS, 2012. 1097~1105
22. Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015. 1~9
23. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv, 2014, 1556(1409): 1~14
24. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 770~778
25. Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. in: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014. 675~678
26. Yi D, Lei Z, Liao S, et al. Deep Metric Learning for Person Re-identification. in: 2014 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE Computer Society, 2014. 34~39
27. Li W, Zhao R, Xiao T, et al. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. in: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE Computer Society, 2014. 152~159
28. Bromley J, Bentz J W, Bottou L, et al. Signature Verification Using A “Siamese” Time Delay Neural Network. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 669~688
29. Ahmed E, Jones M, Marks T K. An Improved Deep Learning Architecture for Person Re-identification. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015. 3908~3916
30. Wu L, Shen C, Hengel A. PersonNet: Person Re-identification with Deep Convolutional Neural Networks. arXiv preprint arXiv, 2016, 07255(1601): 1~7
31. Xu Y, Ma B, Huang R, et al. Person Search in a Scene by Jointly Modeling People Commonness and Person Uniqueness. in: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014. 937~940
32. Varior R R, Shuai B, Lu J, et al. A Siamese Long Short-Term Memory Architecture for Human Re-identification. in: Computer Vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016. 135~153
33. Liu H, Feng J, Qi M, et al. End-to-End Comparative Attention Networks for Person Re-identification. IEEE Transactions on Image Processing, 2016, 14(8): 1~15
34. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation, 1997, 9(8): 1735~1780
35. Varior R R, Haloi M, Wang G. Gated Siamese Convolutional Neural Network Architecture for Human Re-identification. in: Computer Vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016. 791~808
36. Cheng D, Gong Y, Zhou S, et al. Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 1335~1344
37. Zhang R, Lin L, Zhang R, et al. Bit-Scalable Deep Hashing with Regularized Similarity Learning for Image Retrieval and Person Re-identification. IEEE Transactions on Image Processing, 2015, 24(12): 4766~4779
38. Xiao T, Li H, Ouyang W, et al. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 1249~1258
39. Zheng L, Zhang H, Sun S, et al. Person Re-identification in the Wild. arXiv preprint arXiv, 2016, 02531(1604): 1~10
40. Layne R, Hospedales T M, Gong S, et al. Person Re-identification by Attributes. in: British Machine Vision Conference. Surrey, UK: BMVA Press, 2012. 1~11
41. Shi Z, Hospedales T M, Xiang T. Transferring a Semantic Representation for Person Re-identification and Search. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015. 4184~4193
42. Deng Y, Luo P, Loy C C, et al. Pedestrian Attribute Recognition at Far Distance. in: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014. 789~792
43. Zhu J, Liao S, Lei Z, et al. Pedestrian Attribute Classification in Surveillance: Database and Evaluation. in: 2013 IEEE International Conference on Computer Vision Workshop. Sydney, Australia: IEEE Computer Society, 2013. 331~338
44. Zhang N, Paluri M, Ranzato M A, et al. PANDA: Pose Aligned Networks for Deep Attribute Modeling. in: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE Computer Society, 2014. 1637~1644
45. Chen Q, Huang J, Feris R, et al. Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015. 5315~5324
46. Sudowe P, Spitzer H, Leibe B. Person Attribute Recognition with a Jointly-Trained Holistic CNN Model. in: 2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE Computer Society, 2015. 87~95
47. Zhu J, Liao S, Yi D, et al. Multi-Label CNN Based Pedestrian Attribute Learning for Soft Biometrics. in: International Conference on Biometrics. Phuket, Thailand: IEEE, 2015. 535~540
48. Su C, Zhang S, Xing J, et al. Deep Attributes Driven Multi-Camera Person Re-identification. in: Computer Vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016. 475~491
49. Peng P, Tian Y, Xiang T, et al. Joint Learning of Semantic and Latent Attributes. in: Computer Vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016. 336~353
50. Lin Y, Zheng L, Zheng Z, et al. Improving Person Re-identification by Attribute and Identity Learning. arXiv preprint arXiv, 2017, 07220(1703): 1~10
51. Zhu J, Liao S, Lei Z, et al. Multi-Label Convolutional Neural Network Based Pedestrian Attribute Classification. Image and Vision Computing, 2017, 58(1): 224~229
52. Xiao Q, Cao K, Chen H, et al. Cross Domain Knowledge Transfer for Person Re-identification. arXiv preprint arXiv, 2016, 06026(1611): 1~10
53. Gray D, Tao H. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. in: Computer Vision - ECCV 2008. Marseille, France: Springer, 2008. 262~275
54. Wang F, Zuo W, Lin L, et al. Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016. 1288~1296
55. Zheng Z, Zheng L, Yang Y. A Discriminatively Learned CNN Embedding for Person Re-identification. arXiv preprint arXiv, 2016, 05666(1611): 1~10
56. Felzenszwalb P F, Girshick R B, McAllester D, et al. Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627~1645
57. Schumann A, Gong S, Schuchert T. Deep Learning Prototype Domains for Person Re-identification. arXiv preprint arXiv, 2016, 05047(1610): 1~12
58. Zhong Z, Zheng L, Cao D, et al. Re-ranking Person Re-identification with k-reciprocal Encoding. arXiv preprint arXiv, 2017, 08398(1701): 1~10

# 附录1 攻读学位期间发表的学术论文目录

1. **Liu Maolin**, Yu Chengyue, Ling Hefei, Lei Jie. Hierarchical Joint CNN-Based Models for Fine-Grained Cars Recognition. in: International Conference on Cloud Computing and Security. Nanjing, China: Springer, 2016. 337~347（单位为华中科技大学）
2. Ou Xinyu, Ling Hefei, Yan Lingyu, **Liu Maolin**. Convolutional Neural Codes for Image Retrieval. in: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Chiang Mai, Thailand: IEEE, 2014. 1~10（单位为华中科技大学）
3. Ou Xinyu, Yan Lingyu, Ling Hefei, Liu Cong, **Liu Maolin**. Inductive Transfer Deep Hashing for Image Retrieval. in: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014. 969~972（单位为华中科技大学）