



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331,CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 基于多尺度特征迁移学习的步态识别研究  
作者: 徐健, 黄磊, 陈倩倩, 陆珍, 吴曙培  
网络首发日期: 2020-10-22  
引用格式: 徐健, 黄磊, 陈倩倩, 陆珍, 吴曙培. 基于多尺度特征迁移学习的步态识别研究. 计算机工程与应用.  
<https://kns.cnki.net/kcms/detail/11.2127.TP.20201021.1730.014.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于多尺度特征迁移学习的步态识别研究

徐 健, 黄 磊, 陈倩倩, 陆 珍, 吴曙培

西安工程大学 电子信息学院, 西安 710048

**摘 要:** 为了解决行人步态数据集样本量较少、单特征或多特征融合的步态识别算法特征描述不足的问题, 提出了一种基于多尺度特征深度迁移学习的行人步态识别方法。该算法步骤包括: 首先改进 VGG-16 网络, 去除网络中最后一个最大池化层(Maxpool Layer), 融合空间金字塔池化网络结构(SPP)获取行人步态能量图(GEI)的多尺度信息, 然后利用 Imagenet 数据集预训练此网络模型, 将提取特征能力迁移至行人步态识别网络模型中, 采用行人步态样本集微调网络, 修改网络中的全连接层参数, 最后应用于行人步态识别研究。此方法在中科院自动化研究所的 CASIA-B 步态数据集上的识别精度达到了 95.7%, 与单一步态特征的步态识别方法以及融合多种步态特征的识别方法相比, 步态识别率有了明显提升, 表明此方法有更好的识别性能。

**关键词:** 步态识别; 迁移学习; 步态能量图; 空间金字塔池化; 多尺度特征

**文献标识码:** A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.2006-0154

徐健, 黄磊, 陈倩倩, 等. 基于多尺度特征迁移学习的步态识别研究. 计算机工程与应用

XU Jian, HUANG Lei, CHEN Qianqian, et al. Research on Pedestrian Gait Recognition Based on Multi-scale Feature Transfer Learning. Computer Engineering and Applications

## Research on Pedestrian Gait Recognition Based on Multi-scale Feature Transfer Learning

XU Jian, HUANG Lei, CHEN Qianqian, LU Zhen, WU Shupei

School of Electronic Information, Xi'an Polytechnic University, Xi'an 710048, China

**Abstract:** In order to solve the problem of small sample size of pedestrian gait datasets and the problem of insufficient feature description of single-feature or multiple feature fusion gait recognition algorithm. A pedestrian gait recognition method based on multi-scale feature deep migration learning is proposed. The steps of the algorithm include: Firstly, improving the VGG-16 network by removing the last maximum pooling layer (Maxpool Layer) in the network and fusing the spatial pyramid pooling network structure (SPP) to obtain multiscale information of the pedestrian gait energy image (GEI). Then the network model is pretrained by using the Imagenet dataset, and the feature extraction ability is transferred to the pedestrian gait recognition network model. Using pedestrian gait sample set to fine-tune the network, modifying the full connection layer parameters in the network, and finally applying it to the pedestrian gait recognition research. This method achieves 95.7% recognition accuracy on the CASIA-B gait dataset at the CAS Institute of Automation. The method has been demonstrated higher recognition rate than single-feature or multiple feature fusion gait recognition algorithm. It is shown that this method has better recognition performance.

**Key words:** gait recognition; Transfer learning; Gait Energy Image; Spatial pyramid pooling; Multi-scale features

**基金项目:** 陕西省科技厅工业领域一般项目(No.2018GY-173)。

**作者简介:** 徐健(1963—),男,学士,教授,研究领域为机器视觉,图像处理, E-mail: 307676686@qq.com; 黄磊(1996—),男,硕士研究生,研究领域为机器视觉。

## 1 引言

随着我国智慧城市建设的深入,我国视频监控市场继续保持快速增长的态势。行人作为公共监控视频中的核心,如何在视频监控中对行人进行准确检测与有效识别成为当今研究热点。

人体步态识别作为一种新型生物识别技术,根据各人独特的行走方式来标记、描述和确定行人个体的身份。近代医学研究表明,步态是一种复杂的人类行为特征,人类的步态中存在 24 种不同的影响因素,如果将所有因素都考虑起来,则可以唯一地确定一个人<sup>[1]</sup>,此项医学研究为步态识别提供了可能。

步态识别与人脸识别和指纹识别等其他生物识别技术相比较而言,步态识别的优势在于不需要当事人同意且无需身体直接或间接接触的情况下即可进行识别行人身份。同时,步态识别高度依赖于人体的全身身体特征和行走姿态,因此步态作为识别特征也很难被隐藏或改变。随着图像处理、模式识别及计算机技术不断发展,近些年一部分研究学者采用传统机器学习方法实现通过行人步态达到识别行人身份的目的。李凯等<sup>[2]</sup>提出用步态能量图作为步态的特征图像,利用受限玻尔兹曼机自动获取步态特征,并选取支持向量机、孪生支持向量机、神经网络与 k 近邻识别方法对使用受限玻尔兹曼

机方法的提取的特征进行了研究。赵喜玲等<sup>[3]</sup>提出基于静态能量图和动态群体隐马尔可夫模型的步态识别方法,最后采用近邻法进行识别,极大程度减少了角度变化和噪声带来的影响,达到了较高的识别率。

近些年来,深度学习<sup>[4]</sup>发展迅速,在语音识别、机器视觉、图像识别等诸多领域取得了巨大成功。卷积神经网络 (Convolutional Neural Network ,CNN) 是深度学习中一种端到端的人工智能模型, CNN 在海量的 Imagenet 数据集上识别和检测达到了比人眼更高的识别率与准确率,但在实际学习工作中,由于往往不具备官方数据集那么大的数据量,因此应用于实际样本集时效果很差,通常表现为训练集识别率和测试集识别率相差甚大,也经常导致过拟合现象发生。Sinno 等<sup>[5]</sup>提出迁移学习的思想,解决了样本数据量不够的问题,并且还可避免大量昂贵的数据标注工作。基于迁移学习的思想,龙满生等<sup>[6]</sup>将 AlexNet 模型在 ImageNet 图像数据集上学习到的知识迁移到油茶病害识别任务中,对于常见的 5 种油茶病害都达到了极高的识别率。罗娟等<sup>[7]</sup>人采用深度卷积神经网络通过迁移学习分别训练植物的单个器官,并且结合多个神经网络的多线索进行植物识别,对传统植物都可以有效地自动进行识别。关胤<sup>[8]</sup>采用 152 层残差网络结构,在花卉数据集上对模型进行

迁移学习训练,对于用户上传的花卉照片达到了较高的识别率。

针对当前公布的步态数据集,如果使用步态能量图作为网络输入,会导致样本集偏小易产生过拟合的问题。因此本研究引入深度迁移学习思想,利用 VGG-16 网络模型作为特征提取器,考虑到原网络提取图像特征时仅针对单一尺度这样会导致丢失部分图像细节信息,故在网络模型上融合空间金字塔池化网络 (SPP)<sup>[9]</sup> 优化算法,提取行人步态中的多尺度信息特征。最终提出一种基于融合空间金字塔池化微调 VGG-16 网络模型的行人步态识别方法。通过对比现有多种识别方法,该方法不仅避免过拟合现象发生,而且测试集识别率有所提升。

## 2 步态能量图

步态能量图是一种混合步态轮廓序列中静态和动态信息的运动模板,通过计算一个步态周期中步态轮廓图像像素的平均强度得到模板中每个像素点的能量<sup>[10]</sup>。步态能量图提取方法简单,并且能较完整的表现步态的形状、速度等特征,因此行人步态识别过程中获取完整的步态能量图对整个识别过程至关重要。

### 2.1 步态周期

人体的步态被认为是具有周期性的循环运动,一个完整的步态周期是从一侧足的足跟着地到同一侧足跟

再次着地为止的连续过程所用的时间。

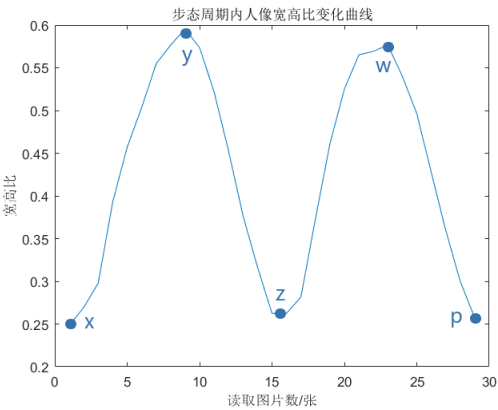
步态周期常用的检测方法是基于人体侧影的宽高比<sup>[11]</sup>, 对于一个连续的步态过程, 人体的高度和宽度会随着身体的运动不断变化。

如图 1(a)所示,当两腿分开的角度逐渐变大时,人体的宽度变大,高度变小,此时宽高比变大,当两腿逐渐靠近时,人体的高度变大,宽度变小,此时宽高比变小,因为整个步态周期呈现周期变化,因此宽高比变化也呈现周期变化。

图 1(b)中的 x 点、z 点与 p 点对应双腿并拢时所对应的宽高比曲线, y 点、w 点分别对应左脚迈出最大步伐时与右脚迈出最大步伐时所对应的宽高比曲线,由 x 点到 p 点即为一个完整的步态周期。



(a) 半个周期步态序列



(b) 步态周期内人像宽高比变化曲线

图 1 步态宽高比变化曲线

Fig.1 Variation curve of gait aspect ratio

## 2.2 提取步态能量图

相较于将步态轮廓图或者步态序列作为深度网络的输入,步态能量图将整个步态过程用一幅特征图像进行表示,保留了运动过程中的绝大部分特征信息,同时对于因某一帧轮廓图出现的轮廓分割变形而产生的噪声有很强的鲁棒性,并且卷积神经网络对于彩色图像十分敏感,网络会被彩色特征吸引,行人衣服变更会极大影响识别准确率,步态能量图能够忽视衣物背包和协变量的特征干扰,只关注步态特征本身,包括关节、身高、体型、行走的姿态等,因此选择将步态能量图作为深度神经网络的输入。GEI 的定义如下:

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

其中  $N$  为单个步态周期内目标图像的总帧数,  $B_t(x, y)$  是步态周期中的  $t$  时刻的二值步态图像序列,  $t$  代表步态周期中的第  $t$  帧。

一个完整周期内的人体目标图像合成步态能量图的过程如图 2 所示。

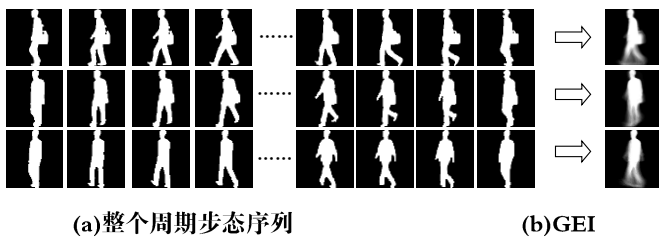


图 2 合成步态能量图

Fig.2 Synthetic gait energy map

## 3 识别算法模型结构

### 3.1 卷积神经网络

传统的机器学习识别方法,主要分为三个步骤:目标图像预处理、特征提取、特征处理,然后利用分类器等方法对特征向量进行分类操作。这种方法的弊端是每个过程需要人为参与费时费力且特征提取的好坏直接影响分类结果,而提取特征时需要工作人员丰富经验,且最终识别率有待提高。

而深度学习大大减少了发现特征的成本,其本质是可以利用庞大丰富的数据,自主学习图像深层特征,完成模型的训练,泛化能力更强,更容易应用于实际场景。卷积神经网络(CNN)是一种经典的深度学习模型,常见的卷积神经网络由输入层、卷积层、激活层、池化层、全连接层和最后的输出层构成,此网络可以有效的从大量样本中学习到相应地特征,避免了复杂的特征提取过程。

VGG-16<sup>[12]</sup>是 2014 年提出的一种经典卷积神经网络模型,相比之前其他网络模型,通过增加更多的卷积层来增加网络的深度,卷积层全部采用  $3 \times 3$  小卷积核代替以往网络中的  $5 \times 5$  大卷积核,多卷积核串联与池化层共同作用使得网络结构简单性能良好。其结构如图 3 所示。



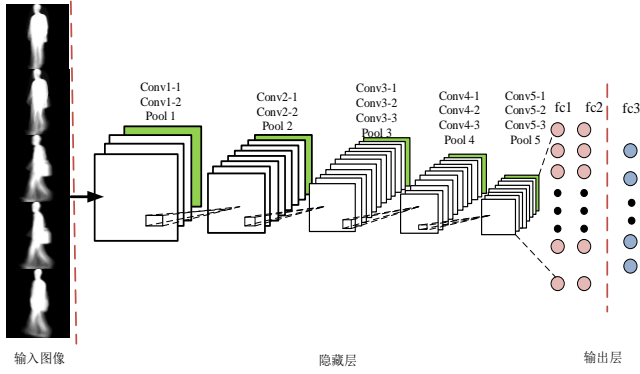


图3 VGG-16 网络模型

Fig.3 VGG-16 network model

原网络输入采用  $224 \times 224$  的 RGB 彩色图像，其中隐藏层主要由 13 个卷积层，5 个池化层和全连接层组成。卷积层通过卷积运算完成特征提取工作，卷积核通过滑动窗口形式在输入图像中以确定的步长从左到右从上到下进行扫描，计算卷积核中的参数与输入图像卷积后的结果，每扫描完一张输入图像后输出一张特征图，再通过激活函数进行非线性映射后输出作为下一层卷积层的输入。卷积层的表达式为：

$$y_i^{n+1} = f\left(\sum_{i \in M_j} y_i^n * w_{ij}^{n+1} + b_j^{n+1}\right) \quad (2)$$

其中  $y_i^{n+1}$  为第  $n+1$  层第  $i$  个输出特征， $f()$  为激活函数， $M_j$  为从第  $n$  层选取的若干个作为第  $n+1$  层输入特征图的集合， $y_i^n$  为第  $n$  层第  $i$  个输出特征， $w_{ij}^{n+1}$  为第  $n+1$  层的权值参数， $b_i^{n+1}$  为  $n+1$  层的偏置， $*$  代表卷积操作。

池化层又称为下采样层，对每个卷积后的特征图进行下采样操作，不仅减少了模型参数，加快网络运算速度，还能滤除部分图像信息，起到特征选择的作用，以

此来提升网络泛化能力。

经过若干次卷积+激励+池化后，网络对前面提取的特征图进行“平展拉伸”，将其转换为一维特征向量，通过全连接层进行相互连接。为防止出现过拟合现象，采用 Dropout 方法随机丢弃网络中的部分神经元，稀疏全连接层，最后通过 Softmax 激活函数分类得到最终的输出，Softmax 函数表达式为：

$$\delta(z_j) = \frac{\exp(z_j)}{\sum_{k=1}^k \exp(z_k)} \quad (3)$$

式中： $z_j$  表示第  $j$  个神经元的输入， $k$  为输出节点的个数即样本类别数； $j = 1 \cdots k$ 。

### 3.2 网络模型结构的改进

一方面由于经过一系列卷积池化操作后得到的特征信息较为单一，另一方面因为 VGG 网络需要输入特定尺寸图像，在检测其他大小的图片的时候，图像需要进行裁剪或者缩放，但是裁剪缩放会减少图像的全局特征信息，降低识别检测的精度。因此为了有效提取行人步态图像中的多尺度特征，本文通过将空间金字塔池化层融合到 VGG-16 网络结构中，对卷积特征进行不同尺度最大池化操作，并提取出每个尺度下的特征，然后将生成的多尺度特征根据网络需要聚合成特定维度，从而提高特征的空间尺度不变性。

本文融合的空间金字塔池化网络结构如图 4 所示。

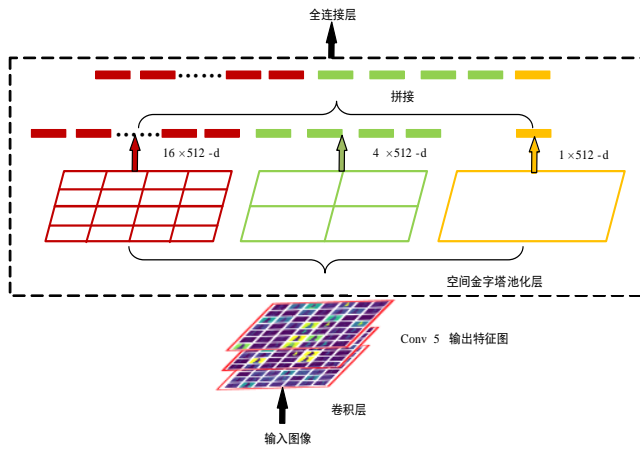


图4 空间金字塔池化结构

Fig.4 Space pyramid pooling structure

此池化结构将 VGG-16 最后一个卷积层 (conv5) 输出的 512 张特征图作为空间金字塔池化层的输入,此时对于每一张  $m \times m$  的特征图可以划分为  $n \times n$  的区域块,对于每一个区域块采取最大池化方式进行池化,其中池化层滑动窗口大小为  $\text{ceiling}(m/n) \times \text{ceiling}(m/n)$ ,步长为  $\text{floor}(m/n)$ 。本文中  $n$  分别取 4、2、1,即将特征图分别划分为  $4 \times 4$ ,  $2 \times 2$ ,  $1 \times 1$  共 21 个不同大小的区域块,将每个区域块中的最大值保留进行级联融合得到  $21 \times 512$  的多尺度特征,最后将这个多维特征“平展拉伸”成一维向量送入全连接层用于后续分类工作。

### 3.3 深度迁移学习

迁移学习的初衷是节省样本标注时间,把处理源任务获取的知识,应用于新的目标难题。随着卷积神经网络的兴起迁移学习逐渐引起研究者的关注,并得到广泛应用。深度迁移学习即采用深度学习的方法进行迁移学

习,相较于采用非深度学习方法拥有两个优势:自动化地提取更具表现力的特征;满足实际应用中的端到端 (End-to-End) 需求。其优点是可以对已有的知识进行利用,解决目标领域中标签数据匮乏的问题起到举一反三的效果,并且明显减少训练时间提高精度。结合本文步态识别过程中训练样本数据量较少的现状,将深度迁移学习思想用于本次研究,即可克服过拟合现象发生,提升识别率,加快训练速度。

迁移学习有域  $D$  和任务  $T$  之分,以小数据量的样本集作为目标域,以大量标注的数据集作为源域,针对不同的任务,常规的卷积神经网络都需要利用要识别的目标数据集从头训练,这是个很耗时间的任务且在数据集数量较小时极易产生过拟合现象。因此可以根据源域与目标域的极大相似性,先利用源域数据训练神经网络,将学到的能力通过某种方式共享迁移至新模型,缩短目标域训练时间且提升识别率。迁移学习的方式有很多种,根据迁移内容的不同主要分为基于实例的迁移,基于特征的迁移和基于参数的迁移等。本文采用基于参数的迁移方式,将 Imagenet 数据集作为源域,行人步态数据集作为目标域,在融合 SPP 网络的 VGG-16 神经网络模型上进行预训练,迁移参数,目标域微调参数,达到最后步态识别的目标。基于改进 VGG-16 网络的迁

移学习模型结构如图 5 所示。

首先将 VGG-16 网络模型最后的最大池化层去掉，用空间金字塔池化（SPP-Net）代替，且此时输入图像相比于原 VGG 网络不要求必须是  $224 \times 224$  的 RGB 彩色图像，可以输入任意尺寸大小的图像以获得多尺度信息且维度固定的特征，其他网络参数不变，用于源域的预训练操作。我们将加入 SPP 的网络称为 sppvgg 网络模型。待该网络在 Imagenet 数据集上得到的损失函数值与准确率趋向稳定时保存全连接层前面的特征提取层权值参数。将改进的 VGG 网络结构中的全连接层神经元个数按本研究需要进行修改，第一层全连接层由

4096 改为 512，第二层根据行人步态分类需要改为 124，去除了第三层全连接层，经过 softmax 函数得到 124 类行人结果的输出，将其称为 Im-sppvgg 网络模型。将该网络作为行人步态图像识别最终模型，在训练测试过程中由于神经网络中的卷积层和池化层主要承担着提取图像特征的作用，因此先将由源域训练获得的提取特征能力由 sppvgg 网络迁移至识别行人的 Im-sppvgg 网络中，然后将 Im-sppvgg 网络中卷积层池化层的权重参数“冻结”使其处于不被训练状态，再输入小数据量的行人步态数据集至 Im-sppvgg 网络中，微调网络中的全连接层参数使其具有分析识别行人的能力。

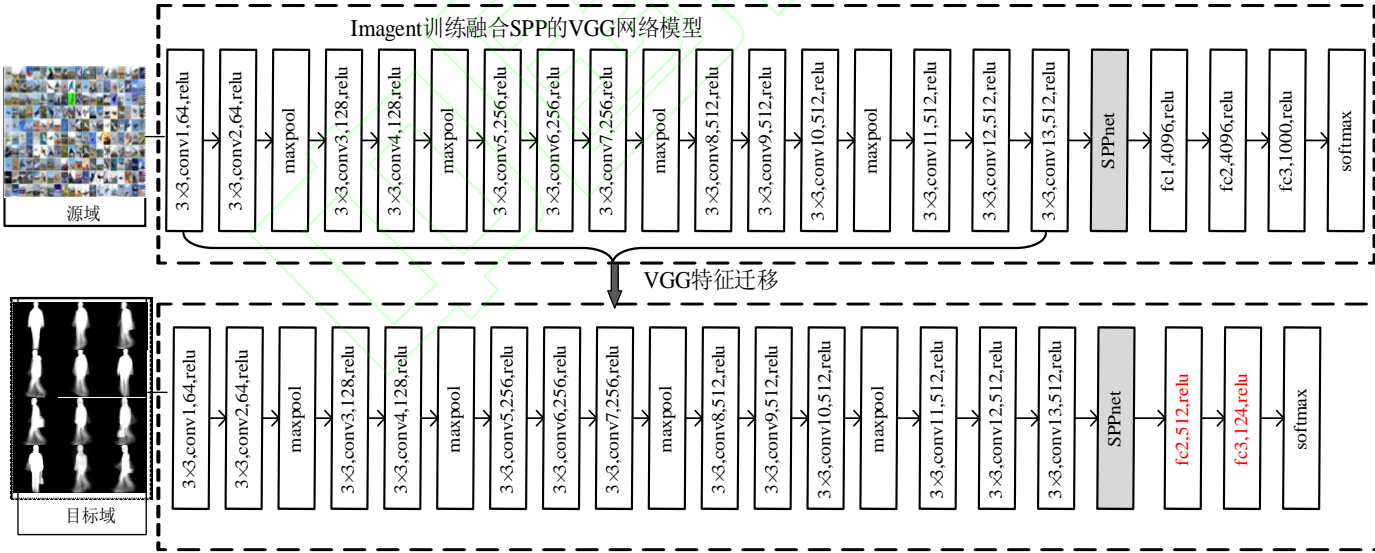


图 5 迁移学习模型

Fig.5 Transfer learning mode

#### 4 实验与分析

本文算法实验在 windows10 操作系统下，CPU 为 Intel i7-7700K，GPU 为 NVIDIA gtx1080，16G 内存，

基于 Tensorflow1.14 深度学习框架，编程语言为 python3.6。



4.1 实验数据集

本文采用的数据集是中国科学院自动化研究所提供的 CASIA 步态数据库 Dataset-B 数据库<sup>[13]</sup>(多视角库)。此数据库提供了 124 位受测者在 11 种视角( $0^{\circ}$  ,  $18^{\circ}$  ,  $36^{\circ}$  , …… ,  $180^{\circ}$  ) , 在三种行走条件下 ( 普通条件, 穿大衣, 携带包裹条件 ) 的行人步态序列。此每种视角选取 1 个完整步态周期作为样本数据集, 每一位行人包含 110 张图像, 一共 13640 张行人数据集图像。数据集存储了行人步态序列的二值图像, 因为本文选择步态能量图作为深度网络的输入, 所以对人体步态序列的预处理(目标检测, 图像分割, 图像归一化等操作)本文将不再论述。

按照前述获取步态能量图的方法获取每个行人样本的步态能量图, 每位行人样本包含 10 个行走条件 ( 2 个穿大衣状态、2 个携带包裹条件、6 个普通行走条件 ), 每种状态包含 11 种视角, 。部分样本如图 6 所示。

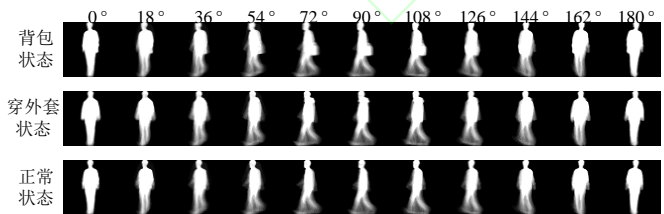


图 6 行人样本集  
Fig.6 Pedestrian sample set

4.2 实验流程及实验参数设置

实验流程如图 7 所示, 首先使用 Imagenet 数据集预训练改进 VGG 后的网络模型 ( sppvgg ), 然后保存

在源域上学习到的权重参数, 迁移预训练的权重参数并修改网络的全连接层, 初始化参数后用行人步态图像样本训练修改后的网络模型 ( Im-sppvgg ), 微调全连接层的参数, 经过 softmax 函数输出识别结果。在整个迁移学习过程中冻结的是预训练模型的低层网络参数, 虽然 Imagenet 数据集中的图像和行人步态图像不一样, 但是它们都是具有相似性的彩色图像, 从大数据集学习得到的低层滤波器往往描述了各种不同的局部边缘和纹理信息, 随着卷积层数增加学习的特征信息越来越抽象化, 因此得到的图像滤波器更具有普适性。而最终本文只根据研究对象需要, 修改全连接层, 使其具备识别行人的能力。

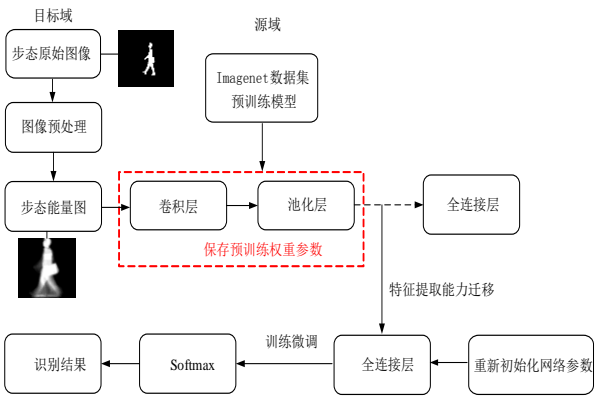


图 7 迁移学习流程图  
Fig.7 Process map of transfer learning

本实验对迁移特征后的 Im-sppvgg 网络使用样本集共 13640 张行人步态图像进行训练微调。训练过程中, 网络超参数如表 1 所示。

表 1 网络超参数

Table 1 Hyper-parameter of network

训练参数	参数设置
Epoch	10
Batch_Size	256
Optimizer	AdamOptimizer
Dropout	0.8

表 1 中各参数具体含义如下：

**Epoch:** 网络训练迭代周期数，整个样本集数据被训练一次计为一个 Epoch，一般取的数值能确保损失函数值和准确率值趋于最终稳定。

**Batch\_Size:** 每次输入神经网络的数据大小，每 Batch\_Size 个数据输入后，进行一次梯度下降法对权重参数更新。

**Optimizer:** 优化器，在比较 SGD（随机梯度下降）等优化器所得效果后，最终选择了 AdamOptimizer 优化器。

**Dropout:** 用于使全连接层的部分神经元失效，丢失一些信息，避免过拟合现象。本实验设置为 0.8。

**损失函数:** 训练过程中通过 Softmax 函数将神经网络的输出变成一个概率分布，使用交叉熵(CrossEntropy)作为网络的损失函数，交叉熵刻画了两个概率分布之间

的距离，是分类问题中常用的评判方法之一。

当  $p$  表示正确答案， $q$  表示网络预测值时，交叉熵可以表示为：

$$H(p, q) = - \sum_x p(x) \ln(q(x)) \quad (4)$$

### 4.3 实验结果及分析

图 8 是使用本文所提出的深度迁移学习网络模型在源域学习到的特征提取能力应用在行人步态上所得到的部分卷积特征图。

从图 8 中看出，随着卷积层不断加深，得到的特征图不断抽象化，在低层特征图可以明显看出行人具体纹理、边缘、形状等特征。在高层次特征图中行人图像已经很抽象，人眼无法看出具体内容，用来滤除不相关内容，提取重要特征。低层卷积核和高层卷积核的局部细节信息表明，卷积神经网络的低层特征抽象能力是可以共享的，因此使用 ImageNet 的大量数据训练基于 VGG 网络改进的 Im-sppvgg 网络能够更详尽、精准地学习到样本的抽象特征。总的来说，将已在 ImageNet 数据集上训练好的模型迁移到行人步态识别问题上能有效地提取图像的抽象特征，进而提高模型的泛化能力。

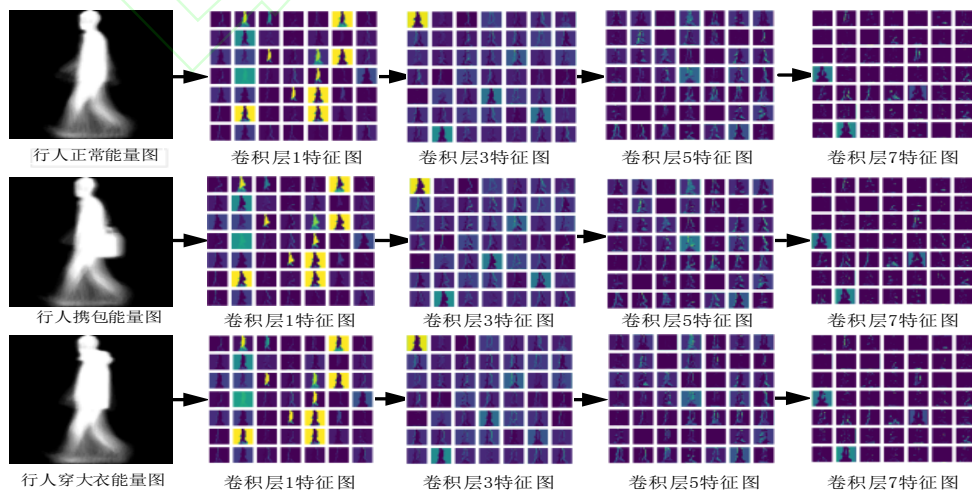
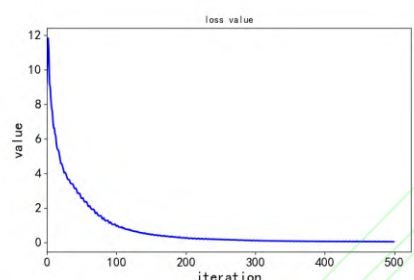


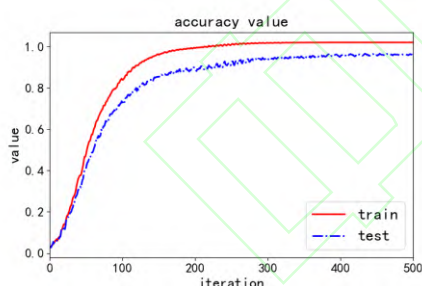
图 8 卷积特征图

Fig.8 Convolution feature maps

为直观描述网络模型的有效性,使用行人步态能量图样本集对模型进行训练,得到损失函数值及准确率曲线如图 9 所示。因本文网络利用改进的 VGG 网络在结合深度迁移学习思想构建而成。为客观比较本文所提网络的性能,对原 VGG-16 网络模型进行迁移学习,其损失函数值及准确率曲线如图 10 所示。其中针对本文 Im-sppvgg 网络输入图像无尺寸要求,而 VGG-16 需对输入图进行尺寸裁剪至  $224 \times 224 \times 3$  大小以满足网络要求。



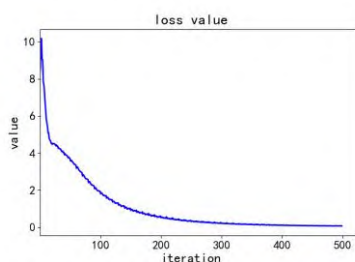
(a) 损失函数值



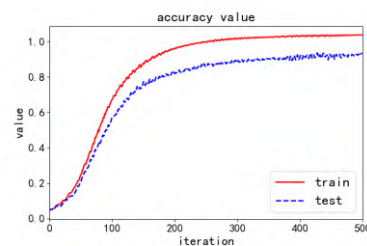
(b) 训练和测试准确率

图 9 Im-sppvgg 网络损失函数值和准确率曲线图

Fig.9 Loss function value and accuracy graph of Im-sppvgg network



(a) 损失函数值



(b) 训练和测试准确率

图 10 VGG-16 网络损失函数值和准确率曲线图

Fig.10 Loss function value and accuracy graph of VGG-16 network

结果表明,图 9 的 Im-sppvgg 网络损失函数值下降很快,大约第 300iteration 时损失函数值就趋向于 0,准确率曲线大概在第 300iteration 时趋向稳定,图中红色曲线代表行人步态图像训练集准确率,稳定后基本保持在 100%左右,蓝色曲线代表行人步态图像测试集准确率,稳定后保持在 95%左右。而基于 VGG-16 网络得到的图 10 中损失函数值下降较慢,大约在 400 iteration 才趋向于 0。而在准确率曲线图中大约 450 iteration 时趋向稳定,此时代表训练集红色曲线保持在 100%左右,代表测试集的蓝色曲线保持在 91%左右。直观上来看本文提出的 Im-sppvgg 网络模型在行人步态图像样本集上精度表现更佳,收敛速度更快。这是因为本文修改 VGG-16 网络时去掉网络最后的最大池化层用空间金字塔池化网络代替。正是由于这种结构上的改进,在 Im-sppvgg 网络输入样本图像时,能够获取多尺度信息特征,使得识别精度提升。而在最后卷积层输

出时,原 VGG16 网络设置三个全连接层,第一个全连接层 FC1 有 4096 个节点,第二个全连接层 FC2 层有 4096 个节点,第三层全连接层 FC3 有 1000 个节点,因为原 VGG 网络输出的卷积特征共  $7 \times 7 \times 512 = 25088$  个节点,而全连接层权重参数数目计算方法:前一层节点数  $\times$  本层的节点数。所以原 VGG16 全连接层参数为: $7 \times 7 \times 512 \times 4096 = 1027645444$ ,  $4096 \times 4096 = 16781321$ ,  $4096 \times 1000 = 4096000$  共 1048522756 个权重参数。

而本文在改进的 VGG16 网络中,经过空间金字塔池化后输出  $(1+4+16) \times 512$  共 10752 个参数,所以修改后 2 层全连接层的权重参数  $10752 \times 512 = 5505024$ ,  $512 \times 124 = 63488$  共 5568512 个权重参数。

较多的权重参数在小样本集上会产生过拟合现象,同时会导致训练时间过长,内存消耗量大,存储容量大等问题。正是不丢失特征信息的参数减少使得网络收敛更快且提高行人步态的分类性能。

本文将实验数据随机打乱分为训练集,测试集及验证集,随机打乱以消除过拟合现象,其中 70%作为训练集,10%作为验证集,20%作为测试集。评价指标使用平均识别率( $ARR$ )。 $ARR$  表示正确识别的目标人物数占总数的比例,  $ARR$  的值越大表明识别率越高,算法性能越好。

$$ARR = \frac{\text{正确识别的目标数}}{\text{目标总数}} \times 100\% \quad (5)$$

为了客观对本文算法进行分析,将本文算法与其他步态识别算法进行比较,表 2 列出了不同方法平均识别率。

表 2 不同算法性能比较

Table 2 Comparison of performance of different algorithms		
对比算法	训练集准确率/%	平均识别率/%
文献[14]	—	82.4
文献[15]	—	93.1
文献[16]	—	94.3
VGG-16	98.9	72.4
VGG-16(迁移学习)	99.7	91.5
本文算法	99.8	95.7

从表 2 对比其他几种文献针对行人步态识别的正确识别率可以看出,本文提出的算法识别性能总体较好,识别率更高。文献[14]采用融合策略将通过数据降维后的步态能量图、梯度直方图、小波描述子和部分图像信息熵特征进行融合后完成步态识别任务,虽然与单一步态特征的识别方法相比,步态识别率有明显提高,但是融合后的特征权重仍通过融合前的单一特征的识别率决定,缺乏特征之间的关联性以及特征个异性在识别过程中的体现。文献[15]采用分解能量图与频谱分析构建步态特征,此方法提取的特征降低了数据复杂度并且能较完整表达行人个体的步态信息,在 CASIA 步态数据库上达到较好的识别性能。文献[16]在步态能量图的基础上分三层求得目标图像的局部二值模式(LBP)与



方向梯度直方图(HOG),使用特征融合的方法每一层的LBP特征和HOG特征进行融合得到最终的特征进行实验验证,在CASIA与USF数据集上都达到较好的识别率。根据分析可知这些方法都是基于传统机器学习方法,采用融合策略融合几种行人步态的图像特征,对融合后的特征进行分类。

传统机器学习方法避免了深度学习大规模参数运算弊端,使得测试时间较短。但在整个提取特征信息时需要人为时刻参与费时费力,并且这种基于单一特征或多种特征的识别算法丢失了大部分步态过程中的信息。

算法4采用的是原始VGG-16神经网络模型,从训练集和测试集识别精度可以看出,在不加修改的经典网络模型上直接用小数据量的能量图作为输入极易产生过拟合现象,这也符合实验之初的设想,因此本文提出采用深度迁移学习的方法。算法5是在VGG-16网络模型结构上采用深度迁移学习方法,相比于算法4消除了过拟合现象,使得测试集精度大幅度提升。而本文算法在算法5基础上,进一步改进,结合空间金字塔池化思想,摆脱VGG-16网络模型对于输入图像尺寸要求的束缚,不仅获得多尺寸多维度信息特征提升识别精度,还使得全连接层权重参数减少,加快算法训练和测试时间。

## 5 结语

针对使用步态能量图作为深度卷积网络的输入,会导致样本集偏小易产生过拟合的问题。本文在VGG-16网络结构基础上进行改进,将大规模数据集训练该网络后将学习得到提取能力迁移,再用行人步态样本集微调网络,用于识别研究。根据行人步态训练集精度和测试集精度曲线,证明该网络识别的有效性。通过对比其他识别算法精度,本文提出的算法的识别率均高于其他算法,证明该网络的优越性。本文算法的提出,不仅克服了传统机器学习方法识别率低,需要人为时刻参与的弊端,而且解决了深度学习在小数据集上产生过拟合现象的问题,在CASIA跨视角跨穿着步态数据库上对本文算法进行实验,结果表明此方法是一种有效的步态识别方法。

## 参考文献:

- [1] Winter D A. Biomechanics and motor control of human movement[M].John Wiley & Sons, 2009.
- [2] 李凯,曹可凡.受限玻尔兹曼机的步态特征提取及其识别[J].河北大学学报(自然科学版),2019,39(6):657-665.  
Li Kai, Cao Kefan. Gait feature extraction based on restricted Boltzmann machine and its recognition method[J].Journal of Hebei University (Natural Science Edition), 2019, 39(6): 657-665
- [3] 赵喜玲,张晓惠.基于动态特征和静态特征融合的步态识别方法[J].湘潭大学自然科学学报,2017,39(4):89-91.  
Zhao Xiling, Zhang Xiaohui. Gait Recognition Based on Dynamic and Static Feature Fusion[J].Natural Science Journal of Xiangtan University, 2017,39(4):89-91.



- [4] 尹宝才,王文通,王立春.深度学习研究综述[J].北京工业大学学报,2015,41(1):48-59.  
Yin Baocai, Wang Wentong, Wang Lichun. Review of Deep Learning[J].Journal of Beijing University of Technology, 2015, 41(1):48-59.
- [5] SINNO J L, PAN QIANG Y. A Survey on Transfer Learning[J]. IEEE Transactions on knowledge and data engineering, 2010,22(10):1345-1359.
- [6] 龙满生,欧阳春娟,刘欢等.基于卷积神经网络与迁移学习的油茶病害图像识别[J].农业工程学报,2018,34(18):194-201.  
Long Mansheng, Ouyang Chunjuan, Liu Huan, et al. Image recognition of Camellia oleifera diseases based on convolutional neural network & transfer learning[J].Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(18): 194-201.
- [7] 罗娟,蔡骋.多线索植物种类识别[J].计算机工程与应用, 2020, 56(5): 160-165.  
Luo Juan, Cai Cheng. Multi-Cue Plant Species Identification[J]. Computer Engineering and Applications,2020,56(5):160 -165.
- [8] 关胤.基于残差网络迁移学习的花卉识别系统[J].计算机工程与应用, 2019, 55(1): 174-179.  
Guan Yin. Flower Species Recognition System Based on Residual Network Transfer Learning[J].Computer Engineering and Applications, 2019, 55(1): 174-179.
- [9] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2015, 37( 9 ):1904 -1916.
- [10] 何逸伟,张军平.步态识别的深度学习:综述[J].模式识别与人工智能,2018,31(5):442-452.  
He Yiwei, Zhang Junping. Deep Learning for Gait Recognition: A Survey[J].Pattern Recognition and Artificial Intelligence, 2018,31(5):442-452.
- [11] 刘志勇,冯国灿,陈伟福.基于局部二值模式和辨识共同向量的步态识别[J].计算机科学,2013,40(9):262-265.  
Liu Zhiyong, Feng Guocan, Chen Weifu. Gait Recognition Based on Local Binary Pattern and Discriminant Common Vector[J].Computer Science,2013,40(09):262-265.
- [12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J].Computer Science,2014.
- [13] YU S Q,TAN D L,TAN T N.A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition[C]//Proc of the IEEE International Conference on Pattern Recognition, 2006:441-444.
- [14] 李林杰,顾广华,王成儒等.能量图和典型相关分析相结合的特征融合步态识别[J].小型微型计算机系统,2014, 35(11): 2558-2561.  
Li Linjie, Gu Guanghua, Wang Chengru, et al. Features Fusion Gait Recognition by Combining Energy Image and Canonical Correlation Analysis[J].Journal of Chinese Computer Systems, 2014,35(11):2558-2561.
- [15] 李玲,罗正平,杨天奇.能量图空间分解法和 DSW 相结合的步态识别[J].计算机工程与应用,2016,52(3):127-132.  
Li Ling, Luo Zhengping, Yang Tianqi. Gait recognition method based on space mass energy image and Dynamic Space Warping[J].Computer Engineering and Applications,2016,52 (3):127-132.
- [16] 刘文婷,卢新明.基于 LBP 和 HOG 特征分层融合的步态识别[J].计算机工程与应用,2018,54(24):168-175.  
Liu Wenting, Lu Xinming. Gait recognition based on layered fusion of LBP and HOG features[J].Computer Engineering and Applications, 2018,54(24):168-175.