doi:10.3772/j.issn.1002-0470.2020.05.005

基于深度学习的人体行为识别研究①

赵新秋② 杨冬冬③ 贺海龙 段思雨

(燕山大学工业计算机控制工程河北省重点实验室 秦皇岛 066004)

摘 要 为解决传统人体行为识别算法存在的运动前景检测不准确、特征提取模糊以及训练识别耗时长等问题,本文提出了基于深度学习的人体行为识别研究方法。利用骨架提取方法对运动前景进行检测及特征提取;针对人体行为动作的时序性,提出了连续帧组合方法;在模型训练环节,对比了不同的网络模型参数,选择了最优的激活函数、优化算法以及 dropout 系数。最后,结合网络模型,分类识别测试样本集中的各种行为,并将识别的结果和当前流行的算法进行比较,通过对比实验,最终实验结果证明了本文所提方法优于其他方法,平均识别率相比其他方法有较大的提高。

关键词 人体行为识别;卷积神经网络(CNN);运动前景检测;连续帧组合

0 引 言

近些年来,随着软硬件水平的不断提升,计算机视觉相关技术得到了爆炸式的发展。计算机视觉[1]是研究如何像人类视觉系统一样,从数字图像或视频中理解其高层内涵的一门学科,简而言之就是研究如何让计算机看懂世界,它包括对数字图像或视频进行预处理、特征提取、特征分类、分析理解几个过程,可以实现将现实世界中的高维数据向低维符号信息的映射,进而触发自主决策。计算机视觉的应用方向包括场景重建、视频跟踪、图像恢复、目标物体识别等等。其中目标物体识别还能细分为物体识别、人脸识别、姿态识别、手势识别、行为识别等。计算机视觉研究方面最近几年有了很大的发展。文献[2]提出了一种基于生成对抗网络(generative aversarial nets,GANs)的主动半监督学习算法。学习者以对抗或合作的方式相互协作,以获得对整个数据分布的全面感知。采用交替更新的方式,对整个体系结构进行端到端的训练。实验结果验证了该算法相对于现有模型的优越性。文献[3]提出了一种新的双向主动学习算法,该算法在双向过程中同时研究无标记和有标记数据集。为了获取新知识,正向学习从未标记的数据集中查询信息量最大的实例。在双向探索框架下,学习模型的泛化能力可以得到很大的提高。文献[4]提出了一种新的弱监督框架,该框架可以同时定位动作帧和识别未裁剪视频中的动作。提出的框架由2个主要组件组成。首先,对于动作帧定位,提出利用自注意机制对每个帧进行加权,从而有效地消除背景帧的影响。其次,考虑到有公开可用的裁剪视频,而且它们包含有用的信息,提出了一个额外的模块来转移裁剪视频中的知识,以提高未修剪视频的分类性能,实验结果清楚地证实了该方法的有效性。

作为机器视觉领域的重点和难点,行为识别逐渐成为研究热点。行为识别[5]技术在现实生活中有着广泛的应用,例如人机交互领域、视频监控领域、智能家居领域以及安全防护领域等。另外,行为识别技术在视频检索[6]及图像压缩等方面也有广泛的应用。人体行为识别研究在人工智能领域拥有

① 河北省自然科学基金(F2016203249)资助项目。

② 女,1969年生,博士,副教授;研究方向:冶金综合自动化,智能控制;E-mail:zxq5460@ysu.edu.cn

③ 通信作者,E-mail:2064596122@qq.com

(收稿日期:2019-06-03)

广阔的市场前景和应用价值，也充满了挑战。因此，本文致力于基于深度学习的人体行为识别研究，旨在取代人工提取特征的方法，提高识别的准确率，同时促进人体行为识别方法在实际生活中的应用价值。

1 运动目标检测

1.1 背景差分法

背景差分法[7]是一种对静止场景进行运动分割的通用方法，它将当前获取的图像帧与背景图像做差分运算，得到目标运动区域的灰度图，对灰度图进行阈值提取运动区域，而且为避免环境光照变化的影响，背景图像根据当前获取图像的帧进行更新。其算法简单易实现，一定程度上克服了环境光线的影响；但对背景图像的实时更新困难，不能用于运动背景复杂的场景。

1.2 帧间差分法

帧间差分法[8]是将视频中相邻 2 帧或相隔几帧图像的 2 幅图像像素相减，并对相减后的图像进行阈值化来提取图像中的运动区域。若相减 2 帧图像的帧数分别为第 k 帧和第 $(k+1)$ 帧，其帧图像分别为 $f_k(x, y)$ 和 $f_{k+1}(x, y)$，差分图像二值化阈值为 T，差分图像用 $D(x, y)$ 表示，则帧间差分法的公式如下：

$$D(x, y) = \begin{cases} 1 & |f_{k+1}(x, y) - f_k(x, y)| > T \\ 0 & \text{其他} \end{cases} \tag{1}$$

帧间差分法不易受环境光线的影响，但其无法识别静止或运动速度很慢的目标，当运动目标表面有大面积灰度值相似区域的情况下，再做差分时图像会出现孔洞。

1.3 ViBe 前景检测法

与背景差分法、帧间差分法不同的是，前景检测算法[9]（foreground detection algorithm，ViBe）建立了一个样本数据库，里面包含图像序列中的每一个像素点，通过比较当下视频帧像素与样本集中的像素值，来确定该点为前景点还是背景点。

背景模型的建立一般都是通过初始化完成，大多数的方法必须先学习一段视频帧图像，然后才能完成前景检测，这在一定程度上无法达到实时性的要求。另外当视频中背景环境不断变化时，还需要重新进行学习，耗时较长。而 ViBe 前景检测方法利用单帧图像就能够对模型进行初始化，具体操作是首先把单帧图像当成背景图像，然后收集背景像素点附近的像素来充当样本集，最后不断更新背景模型，以适应环境光照的变化。

本文分别利用背景差分法、帧间差分法以及 ViBe 前景检测方法对目标前景进行提取，如图 1 所示。

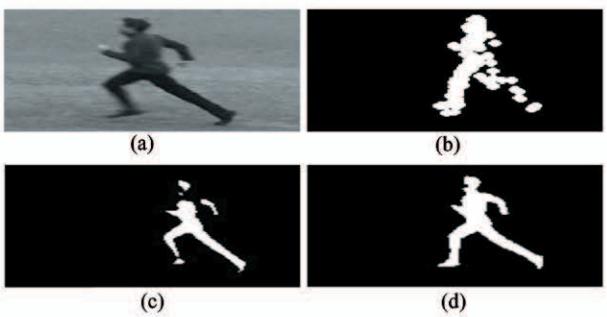


图 1 效果对比图

图 1(a)为原图，图 1(b)、(c)、(d)分别为背景差分、帧间差分法以及 ViBe 前景检测所得到的前景图像。从图中可以看出 ViBe 前景检测方法能够提取出更加清晰的目标运动特征。

1.4 Kinect 骨架提取方法

传统的运动前景检测算法虽然做出了一些改进，但本质上还是无法解决诸如有遮挡物、光照等一些环境因素的影响，所以本文提出了利用 Kinect[10]对人体骨架信息进行采集，将生成一个具有 20 个节点的人体骨架系统。

首先，为了突出骨架数据的优越性，本文测试了环境中有遮挡物时，不同方法采集的人体行为，如图 2所示为传统 ViBe 检测方法与骨架识别的效果对比。

图 2(a)显示由于目标的右臂在前，明显遮挡住了左臂的运动信息，导致最后只能显示出目标右臂的运动特征，无法描绘左臂的运动信息，目标运动特征表达不完整；图 2(b)为 Kinect 相机采集到的骨架信息，同样是右臂遮挡住了左臂的运动信息，Ki-

nect 准确地评估了被遮挡关节点的具体位置，显示出了目标左臂关节点的运动位置,保留了目标完整的运动信息。

图 2　对比图

所以当环境中有遮挡物时,传统的前景检测算法无法再完整地提取前景图像,而用 Kinect 进行采集的时候,它会根据每一个像素点来评估人体关节点所处的具体位置,通过这种方式可以最大可能地保留人体关节点信息,基本可以忽略遮挡物的影响。

2　样本数据库

2.1　KTH 数据库

KTH 数据库包含 6 种类型的人类行为分别是 Walking(步行)、Jogging(慢跑)、Running(跑步)、Boxing(拳击)、Hand waving(挥手)和 Hand clapping(拍手),由 25 名实验人员在 4 个不同的场景中进行了多次运动采集制成。目前数据库中包含 2 391 个动作序列,所有的序列都是用 25 fps 帧速率的静态相机拍摄的均匀背景，这些序列被下采样到 160 ×120 像素的空间分辨率,平均长度为 4 s。图 3 是数据库中 6 种行为在不同场景下的图像帧示意图。

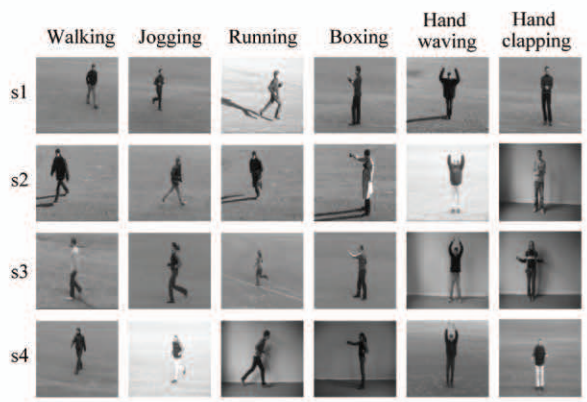


图 3　KTH 数据库

目前，国内外研究人员常用的人体行为识别数据库主要有 Weizmann、UCF101 等。不过,由于 Weizmann 数据库样本数量较少,不适合作为卷积神经网络的样本;而 UCF101 数据库样本数量又过大,对实验硬件要求较高,也不适合作为本实验样本;KTH 数据库样本充足,内容涵盖了基本的人体动作行为,因此采用 KTH 数据库作为本文的一个实验样本库。

2.2　骨架数据库

为了能够突出 Kinect 骨架数据在本文人体动作行为识别上的优越性,分别采集了与 KTH 数据库中相同的 6 种人体动作行为。本数据库分别由 3 名实验人员在 3 种不同的场景下采集而成。由于 Kinect 仅采集人体骨架信息,所以无论背景环境以及光照如何变化都不会对人体的骨架信息产生影响,采集视频的摄像头是静止的,图 4 所示是采集的人体骨架数据。

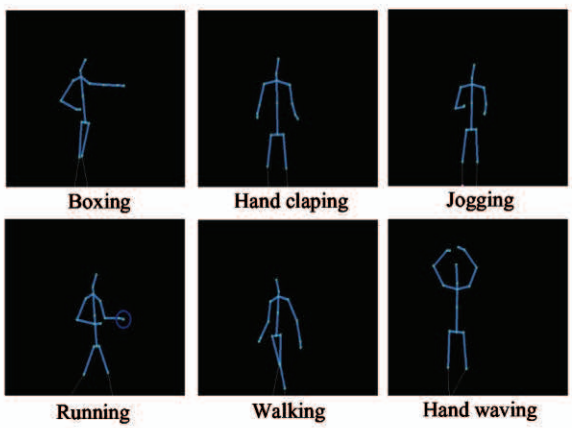


图 4　骨架数据库

3　卷积神经网络

AlexNet[11]是 2012 年 ImageNet 竞赛冠军获得者设计的,该模型层数一共 8 层,含 5 个卷积层、3 个全连接层,如图 5 所示为 AlexNet 网络结构模型。

相比于其他的卷积神经网络(convolutional neural network, CNN),该卷积神经网络主要的创新工作在于应用了如下技术。

(1) 用线性整流函数(rectified linear unit, ReLU) ReLu 作为网络的激活函数[12],并验证了在网络结构较深时效果比 sigmoid 要好。

（2）增加 Dropout 层[13]，减少过拟合问题的发生概率。

（3）在模型中使用重叠的最大池化。规定了移动步长的尺寸必须要比池化核的尺寸要小，这样输出结果相互之间可能存在重叠，有利于增加样本的特征，提高模型的识别率。

（4）增加了局部响应归一化（local response normalization，LRN）层[14]，能够对神经元的反馈进行放大或者抑制，有利于提高模型的鲁棒性。

（5）使用 cuda 加速模型的训练，利用 GPU 强大的并行计算能力，处理神经网络训练时大量的矩阵运算。

（6）增加了数据。随机地从 256 × 256 的原始图像中截取 224 × 224 大小的区域，相当于增加了 2 048倍的数据量。

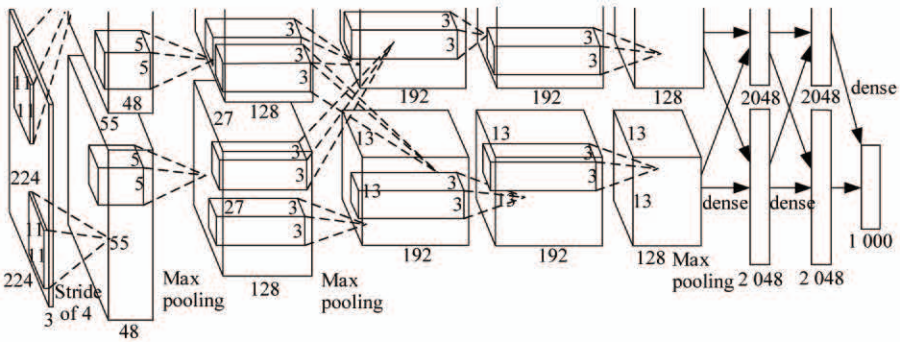


图 5　AlexNet 网络模型

4　实验分析

本实验中使用的处理器为 Intel（R）Xeon（R）Gold 5118 CPU@2.30 GHz，操作系统 Ubuntu 16.04 64 位，基于 Pytorch 0.4 框架，GPU 为 NVIDIA Telsa P100 GPU @ cudnn7。实验数据按照 4∶1 的比例分为训练集和测试集，每次迭代次数为 5 000 代，网络模型为 AlexNet 卷积神经网络模型。

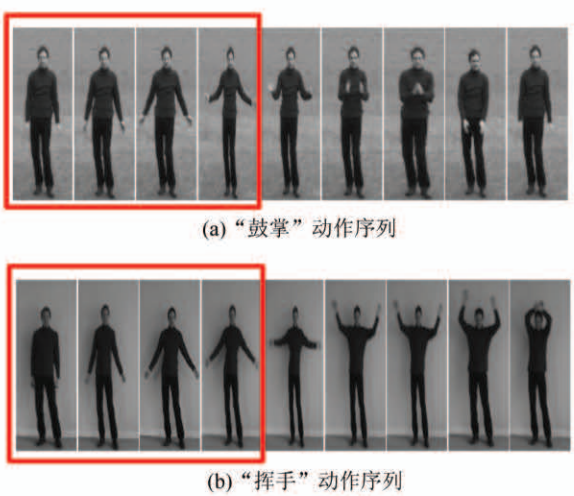
4.1　连续帧组合实验

考虑到卷积神经网络（CNN）识别的是单帧图像，而人体动作行为可以看作是一个连续的图像序列，并且单帧图像不能真正代表一个具体动作行为，特别是当 2 个或几个动作相近的时候，单帧图像识别很容易造成混淆，对模型识别造成很大的干扰。如图 6所示，图 6（a）与（b）分别为鼓掌和挥手 2 种人体行为的连续 9 帧动作序列，从图中可以明显发现，2 种行为的前 4 帧视频序列即用方框标注的序列行为动作极其相似，如果是单帧图像输入模型训练的话，很容易造成模型判断不准确。

所以根据人体行为动作识别的时序性，本文提出了基于连续帧的组合实验方法。连续帧组合实验的公式如下：

$$B_n = \frac{1}{N}(f_N + f_{N-1} + \cdots + f_1) \tag{2}$$

其中，B_n 表示某种行为的预测概率值，N 表示的是输入网络结构的帧数。f_N 表示每一帧预测的概率值。



(a)“鼓掌”动作序列

(b)“挥手”动作序列

图 6　不同行为中相似片段示意图

首先，将模型按照原始的训练方式完成训练。

在测试的时候,不仅仅输入单帧图像,而是将连续几帧图像同时测试,每帧图像都会有一个测试结果,然后将这几帧图像的测试概率相加求平均取最大值,即某种行为的预测概率最大就代表输入的图像是某种行为。连续帧组合根据概率相加求平均最大值方法,可以有效避免单帧输入图像造成模型判断不准确问题,使训练结果更接近于真实行为。具体方式如图 7 所示。

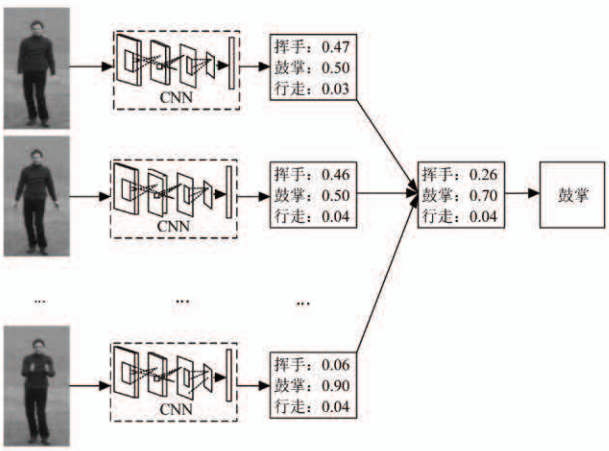


图 7　连续帧组合

为了验证连续帧组合实验的有效性,本文分别在单帧图像与连续帧图像识别上进行了实验,结果如表 1 所示。

从实验结果中可以发现,在一些特征较明显的动作识别上,比如拳击,组合帧识别效果与单帧识别效果几乎没有什么差别。但是在容易产生混淆的动作识别上,例如慢跑和跑步,挥手和拍手,利用组合帧实验方法可以提高模型的准确率。相比之下,在全部测试集中组合帧实验的识别准确率提高了 4.11%,实验结果证明了本文方法的优越性。

表 1　组合帧组合对比实验

行为类别	单帧识别率(%)	组合帧识别率(%)	增幅(%)
Walking	95.61	95.27	-0.34
Jogging	80.58	88.59	8.01
Running	77.67	89.47	11.80
Boxing	98.46	98.55	0.09
Hand waving	79.39	88.04	8.65
Hand clapping	78.81	88.16	9.35
All	87.14	91.25	4.11

4.2　前景检测提取实验

考虑到前景特征提取效果的好坏可能会对网络的识别产生影响,本文在连续帧组合方法的基础上,分别运用传统的基于 ViBe 前景检测法与基于 Kinect 的骨架提取方法对人体运动前景进行提取,然后制作成数据集分别输送到 AlexNet 网络中进行训练。

模型训练曲线如图 8 所示,在平均识别准确率方面,随着迭代次数的增加,模型趋于收敛,相比于传统的基于 ViBe 图像二值化法得到的91.25% 准确率,Kinect 骨架提取方法得到了 93.16% 的准确率,提高了 1.91%;而在损失函数方面,ViBe 图像二值化法得到的损失函数值为 0.257,Kinect 骨架损失函数值为 0.22,下降了 3%。

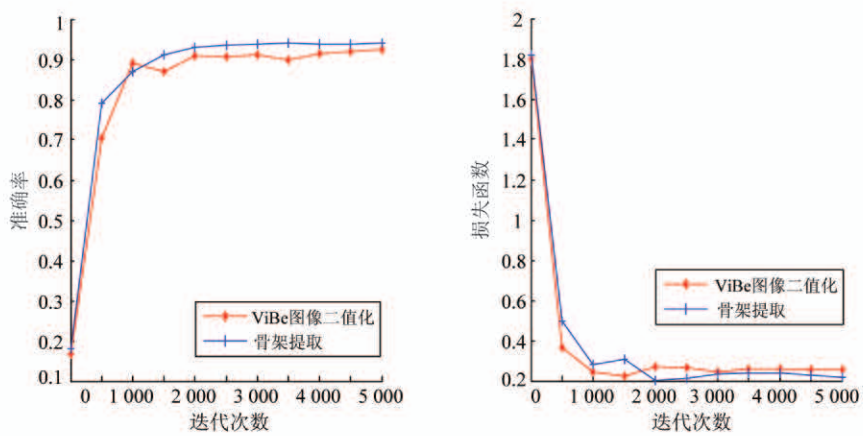


图 8　模型训练曲线

在不同动作的识别方面，本文分别对 6 种不同的动作进行测试，实验结果如表 2 所示。

表 2　不同前景特征提取方法的比较

行为类别	ViBe 识别率（%）	骨架识别率（%）	增幅（%）
Walking	95.27	97.76	2.03
Jogging	88.59	90.16	1.57
Running	89.47	91.71	2.24
Boxing	98.55	98.02	−0.53
Hand waving	88.04	89.93	1.89
Hand clapping	88.16	90.26	2.10
All	91.25	93.16	1.91

从实验结果中可以发现，相比于传统的目标前景提取算法，本文利用 Kinect 相机提取的人体行为骨架信息在步行、慢跑、跑步、挥手、拍手 5 种动作的识别率上均有不同的提高，在全部测试集上的平均识别率上优于传统的前景检测算法。验证了输入图像特征对模型识别率的影响，也说明了骨架信息能够更加清晰地表达人体运动特征。

本文利用骨架识别算法与现今主流的行为识别算法进行了比较，如表 3 所示。从表 3 可以看出，本文所选择的深度学习模型识别结果优于其他算法。

表 3　准确率对比

算法	准确率（%）
文献[15]算法	90.2
文献[16]算法	90.5
文献[17]算法	91.8
本文算法	93.2

4.3　优化算法对比实验

优化算法就是在网络模型的训练环节，通过调整网络训练方式，来不断更新网络的内部参数比如权重系数、偏置等，使得网络能够达到最优的收敛效果。网络优化算法的选择对模型最终的识别效果起着至关重要的作用。如图 9 所示为网络权重 w 与梯度误差 E 的关系，从图中可以看出，当网络的权重过大或者过小时，梯度误差都非常大，此时需要重新对网络进行训练。而优化算法的目的就是用最快的方法找到一个局部最优点 k，使得当权重取得一定值时，网络的梯度误差最小。

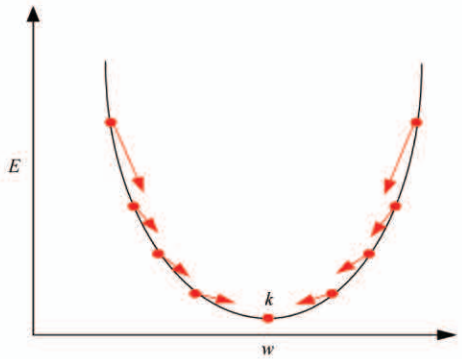


图 9　权重与梯度误差的关系

本文对比了当下比较流行的 5 种优化算法，随机梯度下降法（stochastic gradient descent，SGD）、自适应梯度算法（adaptive gradient algorithm，Adagrad）、自适应的学习率方法（an adaptive learning rate method，AdaDelta）、自适应矩估计（adaptive moment estimation，Adam）、梯度下降法（coreuclbelgium，Nesterov），并在实验平台上分别进行了测试。

随机梯度下降法[18]是在批量梯度下降法的基础上发展起来的，批量梯度下降在每一次网络迭代的时候需要使用所有样本来进行梯度更新，虽然一定能够得到全局最优，但当样本数量非常大的时候，训练速度会变得很慢。

而随机梯度下降法也叫最速下降法是指在网络训练时，每迭代一次会用一个样本对模型的参数进行一次更新，训练速度快，对于很大的数据集，也能够以较快的速度收敛。但由于频繁地对参数进行更新，误差函数可能会按照不同的强度大幅波动。所以每一次迭代的梯度受抽样的影响比较大，不能很好反映真实的梯度。如图10所示，随着迭代次数的

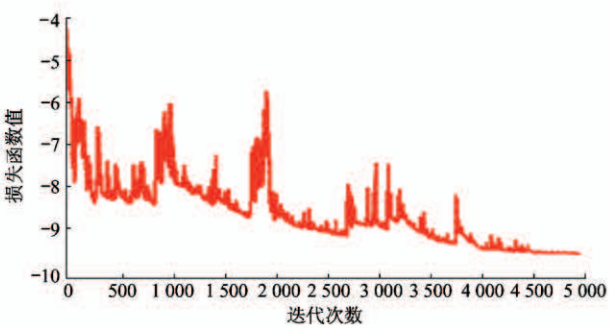


图 10　SGD 损失函数曲线

增加损失函数的大幅度振荡可能会导致最终的结果不是损失函数的最小值。

后来针对随机梯度下降算法存在的弊端,又提出了一种小批量梯度下降法(mini batch gradient descent,MBGD)[19],是指每一次迭代的时候需要使用 n 个样本即 batch _ size,来对梯度进行更新。这样既可以提高运行速度,又能够减少参数的波动更加接近真实的全局最优值。后来将小批量梯度下降算法与随机梯度下降算法统称为随机梯度下降算法。

与随机梯度下降算法不同的是,在 Adagrad[20] 的更新规则中,对于学习率不在设置固定的值,每次迭代过程中,每个参数优化时将会使用不同的学习率,学习率 η 会随着每次迭代而根据历史梯度的变化而变化,在处理稀疏数据时效果非常好。但是随着学习率的不断衰减,网络的学习能力逐渐下降,耗时较长,很难达到最优收敛效果。

相对于 Adagrad 算法来说,AdaDelta [21] 是对学习率进行自适应约束,不需要再设置一个默认的学习率,简化了网络的计算,能够有效地解决 Adagrad 学习率衰减问题。把历史梯度累计窗口限制到固定的尺寸 w,而不是累加所有梯度的平方和。

Adam[22] 是有动量项(Momentum)的一种均方根反向传播算法,动量项可以强化相关方向的振荡以及弱化无关方向的振荡来加速网络的训练。它的学习率调整是根据梯度的一阶、二阶估计来完成的。相比于其他算法对于参数学习率的规定,Adam 的优势体现在每个参数的学习率都有一个固定的范围,这样能够防止参数震荡现象的发生。

Nesterov[23] 加速梯度法作为凸优化中最理想的方法,其收敛速度非常快。在动量项的基础之上,Nesterov 在梯度进行大的跳跃后,进行计算对当前梯度进行校正,避免前进太快以及出现大幅度振荡现象错过最小值,同时提高灵敏度。

本小节分别对 5 种优化算法 SGD、Adagrad、AdaDelta、Adam、Nesterov 在骨架数据集上进行实验,实验结果如表 4 所示。

从表中能够看出,Nesterov 加速梯度法在本文的实验能够取得 93.3% 的准确率,明显高于其他 4 种优化算法。

表 4　不同优化算法的比较

优化算法	准确率(%)
SGD	92.3
Adagrad	90.8
AdaDelta	79.2
Adam	89.1
Nesterov	93.3

4.4　Dropout 网络优化

本文分别在 KTH 数据集和骨架识别数据集上针对 dropout 进行实验,通过设定不同的 dropout 系数得到最终的准确率,如图 11(a)所示是标准的一个全连接的神经网络,图 11(b)是对标准的全连接神经网络应用 dropout 后的结果。从图中可以看出,当加入 dropout 系数后,神经网络会以一定的概率随机丢弃掉一些神经元。

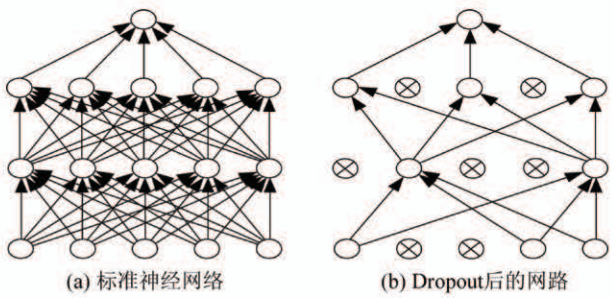


图 11　神经网络

传统的神经网络训练经验一般会把 dropout 系数范围设定为 0.2 ~ 0.5,本文为了更有效地对比 dropout 的作用,将其系数设定为 0 ~ 0.7。识别效果如图 12 所示。

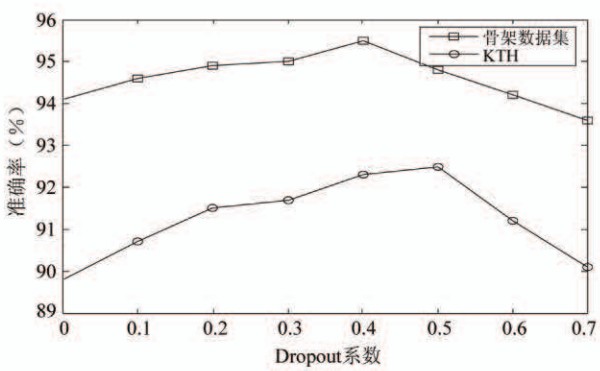


图 12　Dropout 系数对结果的影响

从图中能够总结出在合理的参数选择范围内，随着 dropout 系数的增加，模型准确率不断提高；但当 dropout 系数增加到一定数值后，模型的准确率开始下降。由此可以得出，适当的 dropout 系数可以减少模型的过拟合问题，有利于提高模型的准确率。

5 结 论

本文提出了一种基于骨架识别的人体行为研究方法，将最近比较流行的骨架识别与深度学习相结合。该方法利用 Kinect 相机对人体的骨架信息进行采集，有效地忽略了光照和复杂背景的影响，采集的人体骨架信息能够更好地表达人体的行为特征。然后介绍了 AlexNet 卷积神经网络，将采集的骨架信息制作成训练集输入到 AlexNet 神经网络中去训练，并根据人体行为的时序性，提出了基于连续帧的组合方法，极大地提高了模型的准确率。最后，通过理论结合实际，本文分别在前景特征提取、网络优化算法和基于 dropout 网络优化方面进行了实验。实验结果表明，本文采用的网络参数最终的识别准确率明显高于同等环境下的其他参数模型。

参考文献

[1] 轩中. 中国在计算机视觉领域的人工智能公司[J]. 互联网周刊, 2018, 668(14): 48-50

[2] Zhang X Y, Shi H, Zhu X, et al. Active semi-supervised learning based on self-expressive correlation with generative adversarial networks [J]. *Neurocomputing*, 2019, 345:103-113

[3] Zhang X Y, Wang S, Yun X. Bidirectional active learning: a two-way exploration into unlabeled and labeled data set[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(12):3034-3044

[4] Zhang X Y, Shi H, Li C, et al. Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision[J]. *arXiv*: 1902.07370, 2019

[5] 人体动作行为识别研究综述[J]. 模式识别与人工智能, 2014, 27(1): 35-48

[6] 朱爱红, 李连. 基于内容的视频检索中的镜头分割技术[J]. 情报杂志, 2004, 23(3): 66-68

[7] Herrero-Jaraba E, Orrite-Uruñuela C, Senar J. Detected motion classification with a double-background and a neighborhood-based difference[J]. *Pattern Recognition Letters*, 2006, 24(12):2079-2092

[8] Gao M J, Jin W Q, Wang X, et al. Inter-frame difference oversample reconstruction for optical microscanning thermal microscope imaging system[J]. *Transactions of Beijing Institute of Technology*, 2009, 29(8):704-707

[9] Barnich O, Van D M. ViBe: a universal background subtraction algorithm for video sequences. [J]. *IEEE Transactions on Image Processing*, 2011, 20(6):1709-1724

[10] 李红波, 丁林建, 吴渝, 等. 基于 Kinect 骨骼数据的静态三维手势识别[J]. 计算机应用与软件, 2015(9): 161-165

[11] 苏思悦, 付莹, 来林静. 结合单目标多窗口检测器和 AlexNet 的螺母检测[J]. 中国科技论文, 2018, 13(4): 414-419

[12] 刘杰辉, 范冬雨, 田润良. 基于 ReLU 激活函数的轧制力神经网络预报模型[J]. 锻压技术, 2016, 41(10):162-165

[13] Wahlbeck K, Tuunainen A, Ahokas A, et al. Dropout rates in randomised antipsychotic drug trials[J]. *Psychopharmacology*, 2001, 155(3):230-233

[14] 何高兴. 局部扭立方体环互连网络及其性质[J]. 计算机应用研究, 2014(11):3401-3404

[15] 基于稠密光流轨迹和稀疏编码算法的行为识别方法[J]. 计算机应用, 2016, 36(1):181-187

[16] Fathi A, Mori G. Action recognition by learning mid-level motion features[C]// IEEE Conference on Computer Vision & Pattern Recognition, Anchorage, USA, 2008: 1-8

[17] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]// IEEE Conference on Computer Vision & Pattern Recognition, Anchorage, USA, 2008: 1-8

[18] Niu F, Recht B, Re C, et al. HOGWILD: a lock-free approach to parallelizing stochastic gradient descent[J]. *Advances in Neural Information Processing Systems*, 2011, 24: 693-701

[19] Omoregie E O, Niemann H, Mastalerz V, et al. Microbial methane oxidation and sulfate reduction at cold seeps of the deep Eastern Mediterranean Sea[J]. *Marine Geology*, 2009, 261(1):114-127

[20] 贺昱曜, 李宝奇. 一种组合型的深度学习模型学习率

策略[J]. 自动化学报, 2016, 42(6): 953-958

[21] 刘程浩, 李智, 徐灿,等. 基于深度神经网络的空间目标常用材质 BRDF 模型[J]. 光学学报, 2017, 37(11): 350-359

[22] Newey W K. Adaptive estimation of regression modelsvia moment restrictions [J]. *Journal of Econometrics*, 1988, 38(3): 301-339

[23] 胡长胜, 詹曙, 吴从中. 基于深度特征学习的图像超分辨率重建[J]. 自动化学报, 2017, 43(5): 814-821

Human action recognition based on deep learning

Zhao Xinqiu, Yang Dongdong, He Hailong, Duan Siyu

(Key Laboratory of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao 066004)

Abstract

In order to solve the problems of inaccurate motion foreground detection, feature extraction blur and training recognition time-consuming in traditional human behavior recognition algorithm, a deep learning-based human behavior recognition research method is proposed. The skeleton extraction method is used to detect the motion foreground and feature extraction; for the timing of human behaviors, a continuous frame combination method is proposed; in the model training process, different network model parameters are compared and the optimal activation function, optimization algorithm and dropout coefficient are selected. Finally, combined with the network model, this study classifies and identifies various behaviors in the test sample set and compares the recognition results with the current popular algorithms. Through comparative experiments, the experimental results show that the proposed method is superior to other methods and there is a big improvement in the average recognition rate.

Key words: human behavior recognition, convolutional neural network (CNN), motion foreground detection, continuous frame combination