

文章编号: 2095-2163(2022)04-0101-04

中图分类号: TP391.4

文献标志码: A

基于 OpenPose 的滑雪动作分析

许志豪, 高 铭, 殷绍轩, 崔 杰

(北京信息科技大学 计算机学院, 北京 100101)

摘要: 为了填补动作分类系统在滑雪运动方面的空白, 为后续的研究做铺垫, 本文使用 OpenPose 和 STGCN 建立了一个滑雪动作分类框架。该框架利用 OpenPose 提取人体关键点信息, 将连续的多个单帧的人体关键点信息处理为一个骨骼时空图, 再使用时空图卷积网络 STGCN 训练模型、提取特征、建立分类器。利用构建好的框架可以做到滑雪动作类型实时分类和输入视频分类。

关键词: OpenPose; STGCN; 滑雪; 动作识别

Analysis of skiing action based on OpenPose

XU Zhihao, GAO Ming, YIN Shaoxuan, CUI Jie

(Computer School, Beijing Information Science and Technology University, Beijing 100101, China)

[Abstract] For filling the blank of action classification systems on skiing, also providing the basis for the futural research, the paper builds up a skiing action classifier framework associated with OpenPose and STGCN (Spatial-Temporal Graph Convolutional Networks). Extracting the keypoint information on the human body, the OpenPose renders continuous multiple keypoint information from each frame into a set. Furthermore, the STGCN is used in training models, extracting traits, constructing the classifier. Applying the well-built classifier can categorize the skiing action from the real-time or the prerecorded video input.

[Key words] OpenPose; STGCN; skiing; action recognition

0 引言

当前, 人体动作识别已经广泛应用于视频监控、运动分析和医疗检测等场景, 关于人体姿态识别的研究越来越多^[1]。2003年, 任海兵^[2]提出姿态识别和动作过程识别, 归纳出基于统计、模板和语法的方法。2014年, 李瑞峰等人^[3]将人体动作行为识别问题归纳为计算机经过检测动作数据而获得符号化动作信息, 继而理解其特征的过程。同年, 刘飞^[4]分析了深度图像应用于动作识别的可行性。2016年, 吴军等人^[5]针对人体动作的识别问题提出了一种运用深度卷积神经网络的方法, 在 Actitracker 开源数据库上达到了 91.26% 的识别率, 使深度学习运用于人体动作识别。

现有研究表明, 随着中国冬奥会的成功举办, 冬季运动在中国日渐流行, 其中最典型的的就是滑雪运动。2019年, 中国滑雪产业白皮书指出, 当年国内滑雪场的滑雪人次多达 2 090 万^[6]。但由于滑雪运动具有一定的安全隐患, 不正确的滑雪姿势会对人

体造成伤害。动作分类可用于规范滑雪者的动作, 减小受伤的概率, 并且目前动作识别领域中关于滑雪运动的模型仍然少见, 本文进行的滑雪动作分类框架开发可以为后续滑雪动作的识别评估等更深层次的研究提供有益基础。

1 理论方法

本文搭建的滑雪动作分类框架, 主要由 2 个部分组成: OpenPose 识别提取人体关键点信息, STGCN 模型处理数据并进行分类, 其主要流程如图 1 所示。首先使用 OpenPose 识别训练集视频中的人体关键点信息, 其次通过 Python 编写脚本, 将单个视频中连续的单帧的人体关键点信息按照时间序列分组构建骨架时空图, 每幅图中包含有 300 帧的人体关键点信息。然后, 将骨架时空图输入到 STGCN 中训练模型、提取特征、得到分类器。

1.1 OpenPose

OpenPose 是由美国卡耐基梅隆大学 (CMU) 基于卷积神经网络和监督学习算法开发的一个用于检

基金项目: 北京信息科技大学 2021 年大学生创新创业训练计划项目资助 (5102110805)。

作者简介: 许志豪 (2001-), 男, 本科生, 主要研究方向: 机器学习; 高 铭 (1995-), 男, 硕士, 实验员, 主要研究方向: 机器学习、数据挖掘、虚拟化应用。

通讯作者: 高 铭 Email: gm@bistu.edu.cn

收稿日期: 2021-11-08

测 2D 图片中人体关键点(面部,四肢等)的开源实时系统^[7]。

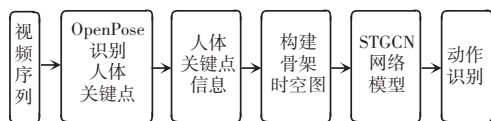


图1 滑雪动作分类框架工作流程图

Fig. 1 Work flow chart of skiing action classification framework

由于在实际滑雪中,一张照片中极可能会有多人同时出现的情况,且人与人之间存在接触、遮挡等复杂的空间干扰问题。传统的自顶向下人体关键点识别算法,由于是先识别到人,再识别身体关键点,使得其结果严重依赖行人检测器,且人数越多的情况下,所花费的时间就越多,导致关键点信息不能实时输出。



图2 OpenPose 算法流程

Fig. 2 OpenPose algorithm flow chart

考虑到 OpenPose 的高鲁棒性和可实时输出的优点,研究中用其来完成滑雪运动时人体关键点的提取工作,但 OpenPose 不具备动作分类的能力,因而本文选用了 STGCN 算法训练分类器。

1.2 STGCN

对于 OpenPose 识别输出的人体关键点信息可以被视为一种拓扑结构。由于传统的卷积神经网络(CNN)无法很好地解决非欧氏空间数据带来的问题,不能有效提取出其空间特征,且提取出的特征缺乏泛化性,而 GCN 在提取拓扑图的空间特征方面具有优越的性能。另经研究发现,在滑雪过程中,完成某个动作都需要一定的时间,即在完成该动作的过程中,人体关

OpenPose 区别于传统识别算法的是,研究时使用了自下而上的人体关键点信息识别,即寻找出人体的各个部位,再通过部分亲和域(Part Affinity Fields)连接各部分。这使得 OpenPose 具有高鲁棒性、可实时输出的优点。所以本文选择使用 OpenPose 识别人体关键点信息。

OpenPose 的算法流程如图 2 所示。由图 2 可知,先将整个图片作为卷积神经网络的输入(见图 2(a)),来联合预测身体部位检测的置信度(见图 2(b))和研究求得用于部位关联的部分亲和域(见图 2(c),图 2(d))。参见图 2(e)中,解析方法执行一组二分匹配在候选的关键点里寻找最为接近的。最后参见图 2(f),将图像中所有人的身体关键点组合起来^[7]。

键点在时间维度上具有一定的特征。因此选用时空图卷积网络 STGCN 作为本文动作分类的模型。

STGCN 模型是由香港中文大学提出的一种时空图卷积神经网络模型,即将时域卷积神经网络 TCN 和图神经网络 GCN 结合,处理有时序关系的图结构数据。如图 3 所示^[8]。STGCN 是同时在时序和空间两个维度上进行卷积并提取特征,由此得出模型。图卷积神经网络 GCN 是对输入的数据进行空间卷积,即不考虑时间的因素,在同一时序的不同点的数据进行卷积运算。时间卷积网络 TCN 是对输入数据进行时序卷积,并考虑不同时序同一特征点的关系,卷积作用于不同时序同一点的数据。

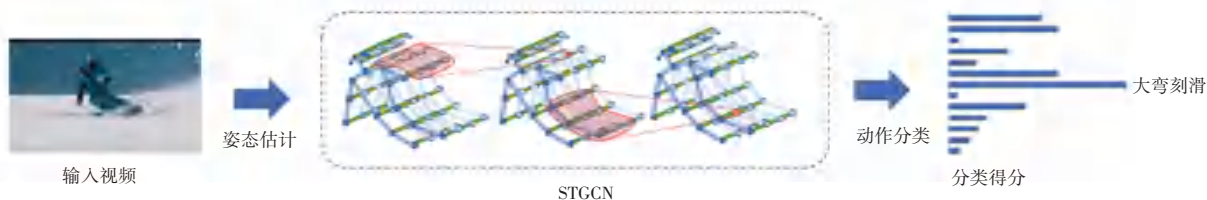


图3 STGCN 算法流程^[8]

Fig. 3 STGCN algorithm flow chart^[8]

2 实验

2.1 实验设置

本文使用自制数据集,共有 200 个视频,分为

11 个类别:平地滑行、平行直滑降、犁式直滑降、犁式制动、半犁式转弯、犁式转弯(深弧)、平行转弯(中弯)、变向横滑降、点仗小弯滑行、大弯刻滑、蘑菇雪道滑行。训练集与测试集比例为 8:2。

本文的实验环境配置如下:操作系统为 Ubuntu 18.04LTS, GPU 为 GeForce RTX 2080。

2.2 实验及结果分析

由于 OpenPose 只是在二维图像上提取人体关键点信息, 这些信息中只包含了动作的空间特征而

没有时间特征。为了弥补这一不足, 本文使用 Python 编写脚本程序, 程序功能如图 4 所示, 将 OpenPose 识别得到的单帧的人体关键点信息构建为带有时间序列的骨架时空图。



图 4 骨架时空图

Fig. 4 Skeleton spatial-temporal diagram

在 STGCN 训练的过程中, 通过调整和训练参数, 最终选取其中准确率最好的一组参数: 学习率为 0.1、批处理量为 32、训练 epoch 数为 50。训练过程中的平均损失值如图 5 所示。从图 5 中可以看出 STGCN 算法基本收敛, 从中选取平均损失值较低的 3 个模型进行比较。比较结果见表 1。由表 1 可以看出 epoch 为 39 时, 各项数值都达到了最优的结果。所以选择 epoch 为 39 的模型作为本文框架所使用的滑雪动作分类器。

现与验证集基本吻合, 具有较高的准确率。

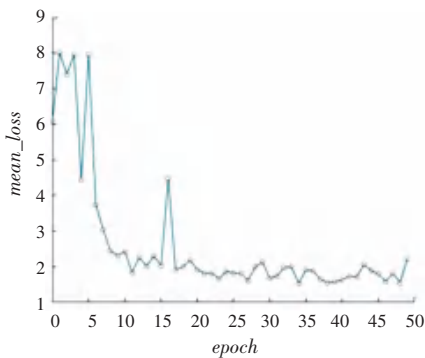


图 5 训练过程 mean_loss 随训练 epoch 的变化图

Fig. 5 The change of mean_loss with epoch

表 1 模型训练评估

Tab. 1 Model training evaluation

epoch	mean_loss	Top1/%	Top5/%
24	1.613 997 70	40.91	77.27
34	1.555 755 00	50.00	86.36
39	1.555 910 83	50.00	90.91

使用该分类器在测试集上进行测试, 输入动作已知的视频, 验证输出结果。测试样例如图 6 所示。由图 6 分析可知, 识别结果正确, 为大弯刻滑 (Big bend craving)。经过测试, 分类器在测试集上的表

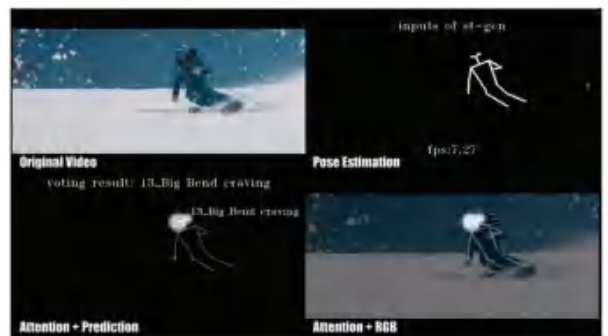


图 6 模型测试效果图

Fig. 6 Model test effect diagram

实验结果表明, 本文建立的滑雪动作分类框架, 优于直接使用传统卷积神经网络 CNN 进行动作分类, 可以做到高准确率、且实时地对滑雪视频进行分类。传统 CNN 由于模型的限制并不能直接分析视频, 这导致滑雪动作在时间维度上的特征被忽略。并且由于 CNN 不能识别非欧氏数据, 只能将整张图片作为输入, 而本文框架使用了 OpenPose 提取人体关键点信息, 使得需要运算的数据量大为降低, 运算时间缩短到了可实时的程度。

3 结束语

本文使用 OpenPose 处理视频, 得到单帧的人体关键点信息, 然后通过自制脚本合并为带有时间序列的多帧集合, 最后使用 STGCN 算法训练模型, 得到分类器。使用该分类器可以完成滑雪视频中的动作识别和实时滑雪动作的识别。本文为后续滑雪动作方面的研究做了铺垫, 填补了这方面的空白。基于 OpenPose 和 STGCN 的滑雪动作分类器具有较高 (下转第 109 页)