



(12) 发明专利

(10) 授权公告号 CN 111079658 B

(45) 授权公告日 2023. 10. 31

(21) 申请号 201911314448.X

G06V 10/764 (2022.01)

(22) 申请日 2019.12.19

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 111079658 A

CN 109446991 A, 2019.03.08

CN 110390363 A, 2019.10.29

CN 102903126 A, 2013.01.30

(43) 申请公布日 2020.04.28

CN 108648746 A, 2018.10.12

US 2017200065 A1, 2017.07.13

(73) 专利权人 北京海国华创云科技有限公司
地址 101199 北京市通州区榆西一街1号院
4号楼5层501室633

审查员 张李一

(72) 发明人 吴伟 马超 王威 关飞庆
左丹婷

(74) 专利代理机构 北京市恒有知识产权代理事
务所(普通合伙) 11576
专利代理师 郭文浩 尹文会

(51) Int. Cl.

G06V 20/40 (2022.01)

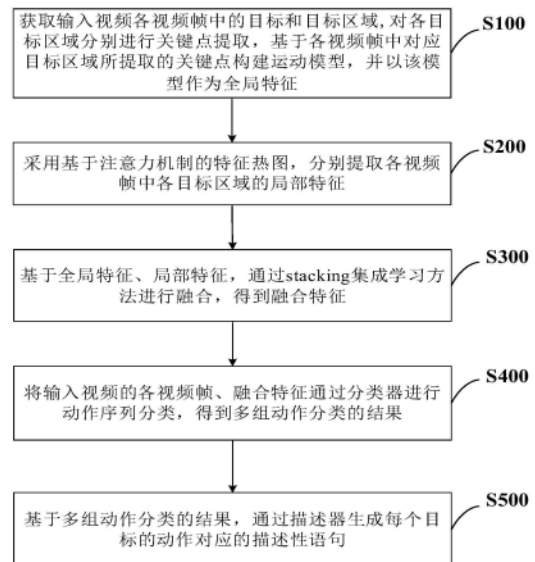
权利要求书2页 说明书7页 附图3页

(54) 发明名称

基于视频的多目标连续行为分析方法、系统、装置

(57) 摘要

本发明属于行为识别领域,具体涉及一种基于视频的多目标连续行为分析方法、系统、装置。旨在解决现有目标行为识别的准确度低的问题。本系统方法包括获取输入视频各视频帧中的目标和目标区域;对各目标区域分别进行关键点提取,并构建运动模型,以该模型作为全局特征;采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;采用stacking集成学习方法将全局特征和局部特征进行融合,获得融合特征;将输入视频的各视频帧、融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句。本发明提高了目标行为识别的准确度。



1. 一种基于视频的多目标连续行为分析方法,其特征在于,包括以下步骤:

步骤S100,获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

步骤S200,采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;

所述S200包括:步骤S201,基于各视频帧中各目标区域提取的关键点,采用基于注意力机制的特征热图方法获取各目标区域对应的特征热图;

步骤S202,通过gabor滤波器、离散小波变换对所述关键点及特征热图进行处理,得到各目标区域的局部特征;具体包括:通过gabor滤波器分别在时间维度和空间维度对所述关键点及特征热图进行滤波,获取多个兴趣点;基于各兴趣点的每一维低通和高通的滤波响应,通过离散小波变换得到多个时空显著点,作为各目标区域的局部特征;

步骤S300,基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

步骤S400,将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;分类器为基于卷积深度神经网络模型构建,包括第一神经网络、第二神经网络、分类融合网络,两个神经网络相对独立;

所述第一神经网络用于对所述输入视频进行分类,输出第一分类值;

所述第二神经网络用于对所述融合特征进行分类,输出第二分类值;

所述分类融合网络用于对第一分类值、第二分类值进行融合,将二者合并为一个向量,作为网络输入,得到最终分类输出即动作分类的结果;

步骤S500,基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句;

所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

2. 根据权利要求1所述的基于视频的多目标连续行为分析方法,其特征在于,步骤S100中“获取输入视频各视频帧中的目标和目标区域”,其方法为:通过multi-stage目标检测方法进行目标识别和定位,获取目标和目标区域。

3. 根据权利要求1所述的基于视频的多目标连续行为分析方法,其特征在于,步骤S400中所述分类器为基于卷积深度神经网络模型构建,包括第一神经网络、第二神经网络、分类融合网络,两个神经网络相对独立;

所述第一神经网络用于对所述输入视频进行分类,输出第一分类值;

所述第二神经网络用于对所述融合特征进行分类,输出第二分类值;

所述分类融合网络用于对第一分类值、第二分类值进行融合,得到最终分类输出。

4. 根据权利要求1所述的基于视频的多目标连续行为分析方法,其特征在于,步骤S500中“通过描述器生成每个目标的动作对应的描述性语句”,其方法为:将多组动作分类的结果编码成词向量,通过所述描述器生成每个目标的动作对应的描述性语句。

5. 一种基于视频的多目标连续行为分析系统,其特征在于,包括全局特征提取模块、局部特征提取模块、特征融合模块、行为分类模块、生成描述模块;

所述全局特征提取模块,配置为获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

所述局部特征提取模块,配置为采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;具体包括:基于各视频帧中各目标区域提取的关键点,采用基于注意力机制的特征热图方法获取各目标区域对应的特征热图;通过gabor滤波器、离散小波变换对所述关键点及特征热图进行处理,得到各目标区域的局部特征;具体包括:通过gabor滤波器分别在时间维度和空间维度对所述关键点及特征热图进行滤波,获取多个兴趣点;基于各兴趣点的每一维低通和高通的滤波响应,通过离散小波变换得到多个时空显著点,作为各目标区域的局部特征;

所述特征融合模块,配置为基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

所述行为分类模块,配置为将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;分类器为基于卷积深度神经网络模型构建,包括第一神经网络、第二神经网络、分类融合网络,两个神经网络相对独立;

所述第一神经网络用于对所述输入视频进行分类,输出第一分类值;

所述第二神经网络用于对所述融合特征进行分类,输出第二分类值;

所述分类融合网络用于对第一分类值、第二分类值进行融合,将二者合并为一个向量,作为网络输入,得到最终分类输出即动作分类的结果;

所述生成描述模块,配置为基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句;

所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

6. 一种存储装置,其中存储有多条程序,其特征在于,所述程序适于由处理器加载并执行以实现权利要求1-4任一项所述的基于视频的多目标连续行为分析方法。

7. 一种处理装置,包括处理器、存储装置;处理器,适于执行各条程序;存储装置,适于存储多条程序;其特征在于,所述程序适于由处理器加载并执行以实现权利要求1-4任一项所述的基于视频的多目标连续行为分析方法。

基于视频的多目标连续行为分析方法、系统、装置

技术领域

[0001] 本发明属于行为识别领域,具体涉及一种基于视频的多目标连续行为分析方法、系统、装置。

背景技术

[0002] 目标行为识别是计算机视觉领域的一个重要的课题。其在行为检测、视频监控等领域都有着广泛的应用价值。与单纯的图片识别不同,目标行为识别会受到诸多因素的干扰,例如光照、背景等。当前的目标识别方法,总体分为传统基于人工特征的模板匹配方法、基于数据驱动的端到端训练学习的方法。

[0003] 传统基于人工特征的模板匹配方法提前通过采集图像信息制作出一个原始的标准模板,然后在检测的时候,系统会将待检测目标的相关数值进行匹配,可以轻易完成一定量的目标识别。但对环境和背景要求较高,环境变化会对图像造成巨大干扰,同时手工特征描述的准确性和视觉中的动态环境、遮挡等问题对系统的性能影响较大。

[0004] 基于数据驱动的端到端学习的方法通过多层、大型的神经网络自动学习到数据中的特征,不再需要手动提取特征。但存在以下问题:

[0005] (1) 数据驱动的学习方法需要大量带标注的图像序列,且每种行为在数据中出现的位置和时间都不确定,同时还要考虑同一行为的不同表现形式以及不同行为之间的区分度,即该方法中数据需要具有多样性和全面性,因此数据的获取和标注的工作量十分巨大。

[0006] (2) 端到端学习的方法只能给出多个可能标签,并且需要耗费大量的计算力资源,且目前的端到端学习的方法在行为识别的精度方面依然不能明显优于传统的基于手工设计特征的方法。以上两种方法均不能满足精确的行为识别需求,因此,本发明提出了一种基于视频的多目标连续行为分析方法。

发明内容

[0007] 为了解决现有技术中的上述问题,即为了解决现有目标行为识别方法由于在训练过程中数据量要求过大、过度依赖环境和背景等原因,导致识别准确度低的问题,本发明第一方面,提出了一种基于视频的多目标连续行为分析方法,该方法包括:

[0008] 步骤S100,获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

[0009] 步骤S200,采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;

[0010] 步骤S300,基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

[0011] 步骤S400,将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;

[0012] 步骤S500,基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句;

[0013] 所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

[0014] 在一些优选的实施方式中,步骤S100中“获取输入视频各视频帧中的目标和目标区域”,其方法为:通过multi-stage目标检测方法进行目标识别和定位,获取目标和目标区域。

[0015] 在一些优选的实施方式中,步骤S200中“分别提取各视频帧中各目标区域的局部特征”,其方法为:

[0016] 步骤S201,基于各视频帧中各目标区域提取的关键点,采用基于注意力机制的特征热图方法获取各目标区域中目标的关键点及特征热图;

[0017] 步骤S202,通过gabor滤波器、离散小波变换对所述关键点及特征热图进行处理,得到各目标区域的局部特征。

[0018] 在一些优选的实施方式中,步骤S202中“通过gabor滤波器、离散小波变换对所述关键点及特征热图进行处理,得到各目标区域的局部特征”,其方法为:

[0019] 通过gabor滤波器分别在时间维度和空间维度对所述关键点及特征热图进行滤波,获取多个兴趣点;

[0020] 基于各兴趣点的每一维低通和高通的滤波响应,通过离散小波变换得到多个时空显著点,作为各目标区域的局部特征。

[0021] 在一些优选的实施方式中,步骤S400中所述分类器为基于卷积深度神经网络模型构建,包括第一神经网络、第二神经网络、分类融合网络,两个神经网络相对独立;

[0022] 所述第一神经网络用于对所述输入视频进行分类,输出第一分类值;

[0023] 所述第二神经网络用于对所述融合特征进行分类,输出第二分类值;

[0024] 所述分类融合网络用于对第一分类值、第二分类值进行融合,得到最终分类输出。

[0025] 在一些优选的实施方式中,步骤S500中“通过描述器生成每个目标的动作对应的描述性语句”,其方法为:将多组动作分类的结果编码成词向量,通过所述描述器生成每个目标的动作对应的描述性语句。

[0026] 本发明的第二方面,提出了一种基于视频的多目标连续行为分析系统,该系统包括全局特征提取模块、局部特征提取模块、特征融合模块、行为分类模块、生成描述模块;

[0027] 所述全局特征提取模块,配置为获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

[0028] 所述局部特征提取模块,配置为采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;

[0029] 所述特征融合模块,配置为基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

[0030] 所述行为分类模块,配置为将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;

[0031] 所述生成描述模块,配置为基于多组动作分类的结果,通过描述器生成每个目标

的动作对应的描述性语句；

[0032] 所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

[0033] 本发明的第三方面,提出了一种存储装置,其中存储有多条程序,所述程序应用由处理器加载并执行上述的基于视频的多目标连续行为分析方法；

[0034] 本发明的第四方面,提出了一种处理装置,包括处理器、存储装置;处理器,适用于执行各条程序;存储装置,适用于存储多条程序;所述程序适用于由处理器加载并执行上述的基于视频的多目标连续行为分析方法。

[0035] 本发明的有益效果：

[0036] 本发明提高了目标行为识别的准确度。本发明在视频的每帧图像上,检测出目标和目标区域,对各视频帧中各目标区域分别进行关键点提取,并构建运动模型,以该模型作为全局特征。根据提取的关键点,采用基于注意力机制的特征热图获取各目标区域的局部特征。将全局特征和局部特征通过stacking集成学习方法进行融合,既充分使用了全局特征的高信息量,同时将局部特征对视角、遮挡等问题不敏感,且具备较强的抗干扰性的优点结合在一起,提高了目标姿态识别的准确性。

[0037] 同时,通过基于卷积深度神经网络模型构建的分类器对目标动作和位姿进行分类,根据分类结果,通过描述器生成了目标行为的描述性语句,使多目标连续行为分析方法具有鲁棒性,更能满足在实际应用中的行为识别需求。

附图说明

[0038] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显：

[0039] 图1是本发明一种实施例的基于视频的多目标连续行为分析方法的流程示意图；

[0040] 图2是本发明一种实施例的基于视频的多目标连续行为分析系统的框架示意图；

[0041] 图3是本发明一种实施例的检测目标和目标区域的示意图；

[0042] 图4是本发明一种实施例的提取的目标的关键点和特征热图的示意图。

具体实施方式

[0043] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0044] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0045] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。

[0046] 本发明的一种基于视频的多目标连续行为分析方法,既可以处理多目标,也可以处理单目标,如图1所示,包括以下步骤：

[0047] 步骤S100,获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

[0048] 步骤S200,采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;

[0049] 步骤S300,基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

[0050] 步骤S400,将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;

[0051] 步骤S500,基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句;

[0052] 所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

[0053] 为了更清晰地对本发明基于视频的多目标连续行为分析方法进行说明,下面结合附图对本发明方法一种实施例中各步骤进行展开详述。

[0054] 步骤S100,获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征。

[0055] 全局特征是把一对象当做成一个整体,这是一种从上到下的研究思维。这种情况下,视频中的人必须先被定位出来。因此,本实施例中,通过multi-stage(多阶段)目标检测方法进行目标识别和定位,获取目标和目标区域,即得到二维全局特征,如图3所示。具体获取过程为:获取输入视频S中第i个视频帧 S_i 中第j个目标 S_i^j 、第j个目标对应的目标区域 DS_i^j ,第j个目标在各视频帧的目标区域集合构成了第j个目标的二维全局特征。

[0056] 本实施例通过multi-stage(多阶段)目标检测方法进行目标识别和定位之前,先通过VGG-19分类网络进行视频帧中目标的识别。

[0057] 基于获取的各视频帧中对应的目标区域,进行关键点提取并构建运动模型。本实施例中得到的运动模型为一或多个,分别为各目标对应的运动模型,该模型为对应目标的三维全局特征。例如,对输入视频S第i帧中第j个目标 S_i^j 对应的目标区域 DS_i^j 进行关键点检测提取作为其运动模型,第j个目标的所有目标区域处理完成后,得到的即为第j个目标的三维全局特征。即三维全局特征是二维全局特征进行关键点检测之后的结果。

[0058] 步骤S200,采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征。

[0059] 本实施例中,提取图像中目标的感兴趣的点或块,不需要精确的目标定位和跟踪,对目标的表观变化,视觉变化和部分遮挡问题不敏感。这些特征点多是图像中的二维信息,因此可以采用基于注意力机制的特征热图的提取方法对各视频帧中各目标区域的进行局部特征的提取,由于图像的特征热图显示了图像中的特征对结果的影响程度,所以,提取其中对结果影响比较大的特征来作为局部特征,忽略不重要特征对结果的干扰。具体处理如下:

[0060] 基于卷积神经网络或卷积神经网络模型,对目标关键点进行检测,在图像中产生

特征热图,获取可以表征目标局部特征的关键点和热图,之后在单独的时间维和空间维上使用gabor滤波器进行滤波,这样检测出的兴趣点数目会随着时间和空间的局部领域尺寸的改变而改变,并通过离散小波变换对每一维的低通和高通滤波响应选择时空显著点,同时,可以整合颜色和运动信息来计算显著点。其中,卷积神经网络在分类任务中训练过程中最后一层使用softmax激活函数,在所有的全连接层之前的卷积层的输出为特征热图。

[0061] 如图4所示,左图为原始视频帧,右图为所提取的各目标的包含关键点的特征热图。

[0062] 本实施例中,采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征,处理过程可参考文献:“王培森,基于注意力机制的图像分类深度学习方法研究,硕士学位论文,中国科学技术大学,2018”。

[0063] 步骤S300,基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征。

[0064] 经过全局和局部特征提取之后,每个目标我们可以得到N+1个特征向量,其包括1个全局特征向量和N个局部特征向量。由于这些特征向量的维数较高,为了避免维数的影响和降低后续步骤的计算量,先对获取的N+1个特征向量进行降维处理。基于降维处理后的特征向量,根据集成学习理论,将获取的特征向量进行集成,从而有效降低分类错误率。本实施例中,通过stacking集成学习方法对获取的全局特征、局部特征进行融合。其中,在训练过程中,该步骤具体处理如下:

[0065] 在步骤S100和步骤S200进行目标识别和检测时将训练数据划分为多个组,如图3中,共包括5组。

[0066] 使用K折交叉验证方法和机器学习算法进行监督学习,即使用k-fold方法将训练数据切分为k份,每一份包括验证集和测试集,每一次取其中的k-1份进行训练,本实施例中,优选采用极端梯度提升方法xgboost进行监督训练学习,在其他实施例中,可以选用其他机器学习算法发,用剩下的一份进行验证。

[0067] 在stacking集成学习方法中,我们对第一个模型,如上述进行全局特征提取的VGG模型,进行训练得到k个分类模型,这k个模型的种类一致,但其中的权重和超参数不同。同样的这里的VGG网络可以替换为RESNET或者InceptionNET等预训练的网络,从而每一种网络可以得到k个分类模型。这样可以得到一个更加鲁棒且泛化能力更强的分类模型,从而为后续多stage的识别和检测网络提供较好的先验知识。

[0068] 基于全局特征提取模型和局部特征提取模型得到的分类模型,通过stacking集成学习方法进行并行集成,得到整体分类模型。即基于全局特征、局部特征,通过stacking集成学习方法得到融合特征。

[0069] 本实施例中,通过全局特征和局部特征的融合,既充分使用了全局特征的高信息量,同时将局部特征对视角、遮挡等问题不敏感、且具备较强的抗干扰性的优点结合在一起。

[0070] 步骤S400,将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果。

[0071] 本实施例中,分类器根据特征融合模块的输出和各视频帧图像进行一个多任务协同的分类,根据融合后的特征对位姿进行识别,并通过位姿和图像信息对目标动作进行识

别,分类器的输出信息包括目标位姿、动作类别。

[0072] 分类器为基于卷积深度神经网络模型构建,包括第一神经网络、第二神经网络、分类融合网络,两个神经网络相对独立;

[0073] 所述第一神经网络用于对所述输入视频进行分类,输出第一分类值;

[0074] 所述第二神经网络用于对所述融合特征进行分类,输出第二分类值;

[0075] 所述分类融合网络用于对第一分类值、第二分类值进行融合,将二者合并为一个向量,作为网络输入,得到最终分类输出。

[0076] 步骤S500,基于多组动作分类的结果,通过描述器生成每个目标的动作对应的描述性语句。

[0077] 本实例中,描述器基于LSTM神经网络构建,预训练成一个文本生成神经网络架构。即对多组动作分类的结果进行编码成词向量,将编码后的词向量输入基于LSTM神经网络构建的描述器,训练神经网络预测下一个词,训练完成后,输入由分类器的结果转换的词向量,通过描述器自动生成描述性语句。

[0078] LSTM神经网络具有学习序列数据之间关系的能力,可用于自动生成文本。可使用部分MSCOCO图像说明文字数据集与动作有关的数据,起始时刻LSTM的输入为与说明文字数据集中对应的动作编码后的词向量,之后输入为其说明文字的每一个单词的词向量,他们编码后为维度相同的词向量。

[0079] 如果分类结果为A,对应说明文字为[B1,B2,B3,B4],完整的输入即为[A,B1,B2,B3,B4],每次输入的输出为其下一个单词。经过训练,LSTM神经网络可学习到输入词向量与说明文字之间的关系。

[0080] 因此,将动作分类结果转化为词向量输入第一层LSTM,输出对应描述性语句。由于此时LSTM并不知道对应的描述文本,所以每次利用之前的词预测下一个词。

[0081] 另外,用于实现本发明方法的系统包括硬件设备、软件系统、数据库;硬件设备包括计算机、摄像头;软件系统包括全局特征提取模块、局部特征提取模块、特征融合模块、行为分类模块、描述器模块;数据库包括目标行为的图像模板库及特征字典;计算机通过USB驱动摄像头采集图像,并通过安装于计算机内的软件系统进行分析。

[0082] 本发明一种实施例的基于视频的多目标连续行为分析系统,如图2所示,包括全局特征提取模块100、局部特征提取模块200、特征融合模块300、行为分类模块400、描述器模块500;

[0083] 所述全局特征提取模块100,配置为获取输入视频各视频帧中的目标和目标区域,分别对各目标区域进行关键点提取,基于各视频帧中对应目标区域所提取的关键点构建运动模型,以该模型作为全局特征;

[0084] 所述局部特征提取模块200,配置为采用基于注意力机制的特征热图,分别提取各视频帧中各目标区域的局部特征;

[0085] 所述特征融合模块300,配置为基于所述全局特征、所述局部特征,通过stacking集成学习方法进行融合,得到融合特征;

[0086] 所述行为分类模块400,配置为将所述输入视频的各视频帧、所述融合特征通过分类器进行动作序列分类,得到多组动作分类的结果;

[0087] 所述生成描述模块500,配置为基于多组动作分类的结果,通过描述器生成每个目

标的动作对应的描述性语句；

[0088] 所述描述器基于LSTM神经网络构建,用于根据分类器的分类结果获取对应的描述性语句。

[0089] 所属技术领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统的具具体工作过程及有关说明,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0090] 需要说明的是,上述实施例提供的基于视频的多目标连续行为分析系统,仅以上述各功能模块的划分进行举例说明,在实际应用中,可以根据需要而将上述功能分配由不同的功能模块来完成,即将本发明实施例中的模块或者步骤再分解或者组合,例如,上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块,以完成以上描述的全部或者部分功能。对于本发明实施例中涉及的模块、步骤的名称,仅仅是为了区分各个模块或者步骤,不视为对本发明的不当限定。

[0091] 本发明第三实施例的一种存储装置,其中存储有多条程序,所述程序适于由处理器加载并执行以实现上述的基于视频的多目标连续行为分析方法。

[0092] 本发明第四实施例的一种处理装置,包括处理器、存储装置;处理器,适于执行各条程序;存储装置,适于存储多条程序;所述程序适于由处理器加载并执行以实现上述的基于视频的多目标连续行为分析方法。

[0093] 所属技术领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的存储装置、处理装置的具体工作过程及有关说明,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0094] 本领域技术人员应该能够意识到,结合本文中所公开的实施例描述的各示例的模块、方法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,软件模块、方法步骤对应的程序可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。为了清楚地说明电子硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以电子硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。本领域技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0095] 术语“第一”、“第二”等是用于区别类似的对象,而不是用于描述或表示特定的顺序或先后次序。

[0096] 术语“包括”或者任何其它类似用语旨在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备/装置不仅包括那些要素,而且还包括没有明确列出的其它要素,或者还包括这些过程、方法、物品或者设备/装置所固有的要素。

[0097] 至此,已经结合附图所示的优选实施方式描述了本发明的技术方案,但是,本领域技术人员容易理解的是,本发明的保护范围显然不局限于这些具体实施方式。在不偏离本发明的原理的前提下,本领域技术人员可以对相关技术特征作出等同的更改或替换,这些更改或替换之后的技术方案都将落入本发明的保护范围之内。

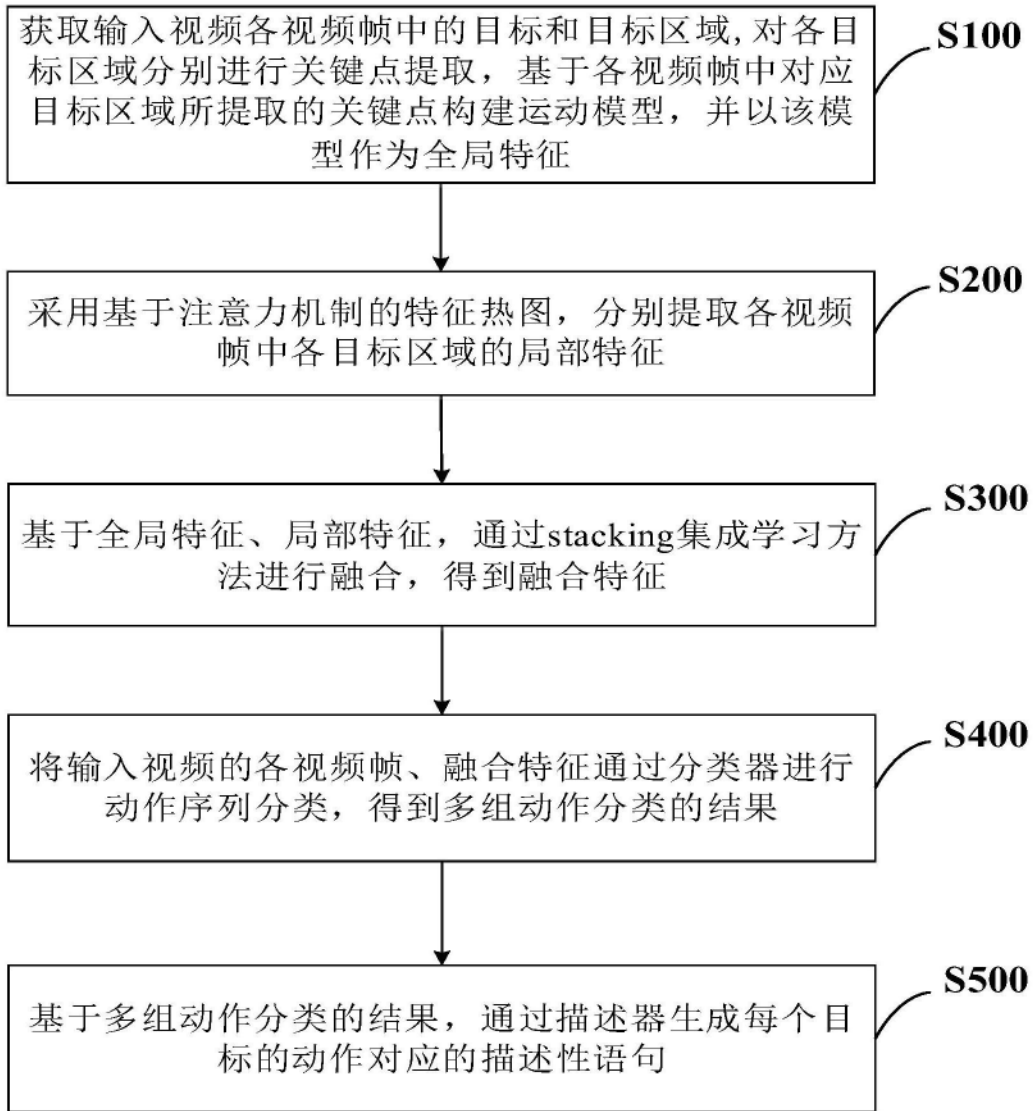


图1

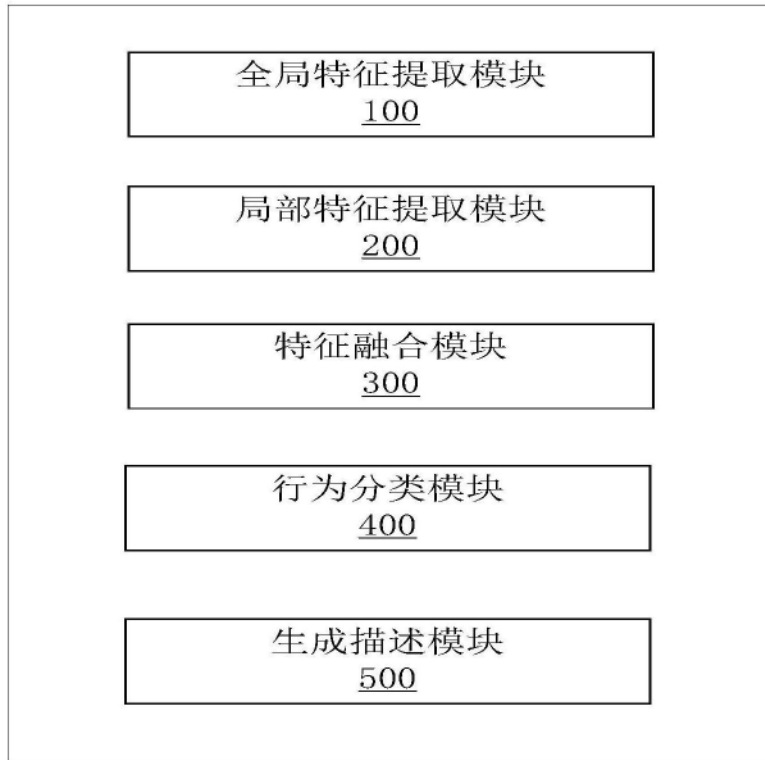


图2

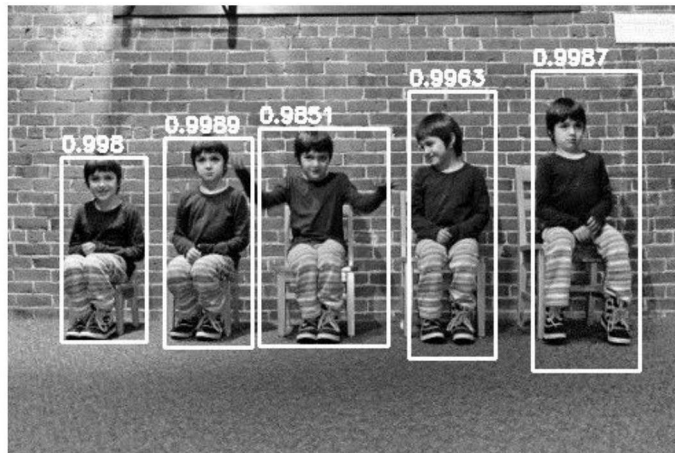


图3

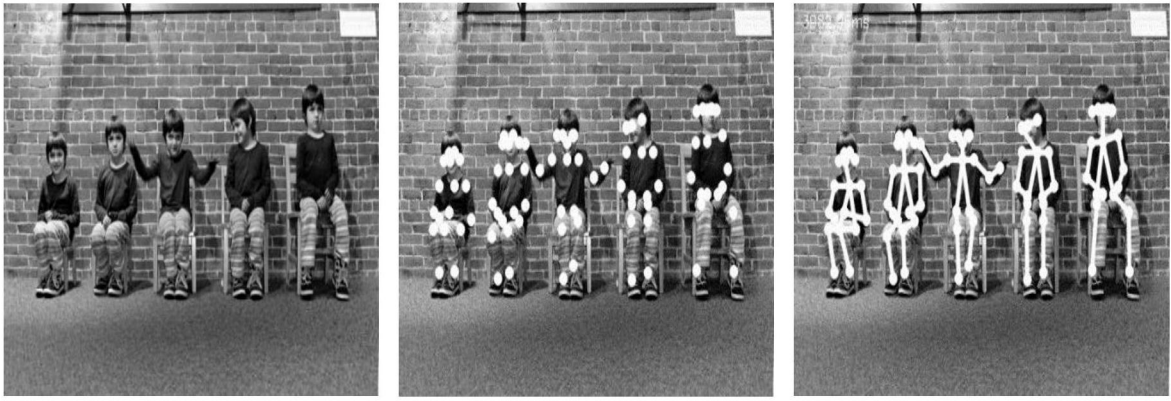


图4