



(12) 发明专利申请

(10) 申请公布号 CN 113989920 A

(43) 申请公布日 2022.01.28

(21) 申请号 202111193385.4

(22) 申请日 2021.10.13

(71) 申请人 复旦大学

地址 200433 上海市杨浦区邯郸路220号

(72) 发明人 张立华 王顺利 杨鼎康 邝昊鹏

(74) 专利代理机构 上海科盛知识产权代理有限公司 31225

代理人 蔡彭君

(51) Int. Cl.

G06V 40/20 (2022.01)

G06V 20/40 (2022.01)

G06K 9/62 (2022.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

G06F 9/451 (2018.01)

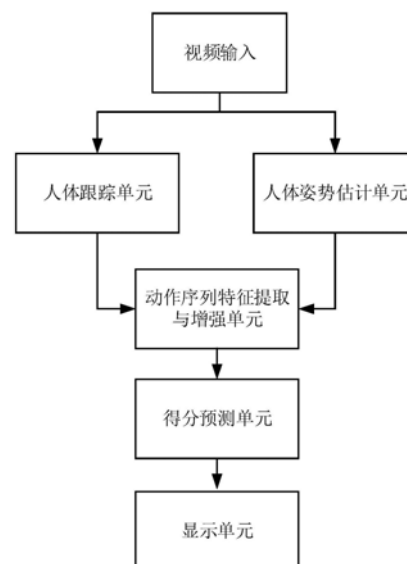
权利要求书2页 说明书9页 附图5页

(54) 发明名称

一种基于深度学习的运动员行为质量评估方法

(57) 摘要

本发明涉及一种基于深度学习的运动员行为质量评估方法,基于的评估系统包括人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元、得分预测单元和显示单元;视频输入至人体跟踪单元,对视频的各帧进行目标检测,得到各帧的检测框作为跟踪结果,在显示单元对跟踪结果进行可视化;人体姿势估计单元获取跟踪结果,对各帧中运动员的姿势进行估计,得到关键点信息作为姿势估计结果,在显示单元对姿势估计结果进行可视化;动作序列特征提取与增强单元以视频、跟踪结果和姿势估计结果作为输入,进行特征提取、特征增强和特征融合后得到视频特征;得分预测单元以视频特征为输入,进行全动作流程质量评估和阶段性行为质量评估。



1. 一种基于深度学习的运动员行为质量评估方法,其特征在于,其所基于的评估系统包括人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元,得分预测单元和显示单元,具体的,每个单元内的流程如下:

视频输入至人体跟踪单元,人体跟踪单元对视频的各帧进行目标检测,得到各帧的检测框作为跟踪结果,在显示单元对跟踪结果进行可视化;

人体姿势估计单元获取跟踪结果,对各帧中运动员的姿势进行估计,得到关键点信息作为姿势估计结果,在显示单元对姿势估计结果进行可视化;

动作序列特征提取与增强单元以视频、跟踪结果和姿势估计结果作为输入,进行特征提取、特征增强和特征融合后得到视频特征;

得分预测单元以视频特征为输入,进行全动作流程质量评估和阶段性行为质量评估。

2. 根据权利要求1所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,所述人体跟踪单元使用YOLO检测器和Siammask框架进行目标检测:在视频的初始帧使用YOLO检测器进行目标检测,得到初始帧的检测框,使用Siammask框架作为单目标跟踪器在初始帧之后的连续帧进行目标检测,得到初始帧之后各帧的检测框。

3. 根据权利要求2所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,所述人体姿势估计单元使用Alphapose框架对各帧中运动员的姿势进行估计:获取视频中各帧的检测框后,采用Alphapose框架对各帧中运动员的姿势进行估计,Alphapose框架生成各帧的关键点信息,使用卡尔曼滤波器算法对各帧的关键点信息进行处理,得到姿势估计结果。

4. 根据权利要求3所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,所述动作序列特征提取与增强单元对视频进行特征提取,得到第一特征,将第一特征结合跟踪结果使用管道自注意力机制进行特征增强,得到增强后的第一特征,动作序列特征提取与增强单元对姿势估计结果进行特征提取,得到第二特征,将第二特征与增强后的第一特征通过全连接层进行融合得到视频特征。

5. 根据权利要求4所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,所述动作序列特征提取与增强单元使用I3D神经网络对视频进行特征提取,得到第一特征,使用图卷积神经网络对姿势估计结果进行特征提取,得到第二特征。

6. 根据权利要求5所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,进行全动作流程质量评估时,将所有的视频特征进行时序全局平均池化后送入全连接层完成质量评估,进行阶段性行为质量评估时,将一段视频的视频特征送入全连接层完成质量评估。

7. 根据权利要求6所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,得分预测单元使用I3D神经网络,进行全动作流程质量评估时,将所有的视频特征进行时序全局平均池化后送入I3D神经网络的全连接层完成质量评估,进行阶段性行为质量评估时,将一段视频的视频特征送入I3D神经网络的全连接层完成质量评估。

8. 根据权利要求4所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,将第一特征结合跟踪结果使用管道自注意力机制进行特征增强具体为:将第一特征与检测框进行量化和对齐生成特征图掩膜,按照视频的帧数和第一特征的数量比例进行掩膜的融合以生成时空管道,在时空管道内部采用管道自注意力机制完成第一特征的稀疏增强操

作,通过残差连接将第一特征与完成稀疏增强后的第一特征进行融合,得到增强后的第一特征。

9.根据权利要求8所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,获取跟踪结果和第一特征后,按照视频的帧数和第一特征的数量比例确定检测框的数量与第一特征的时序维度数量的比值 $N:1$, $N>1$,确定每一个检测框所对应的掩膜,获取检测框覆盖第一特征的特征网络的比例,如果比例大于预设置的阈值,则该第一特征被选中,否则,该第一特征被剔除,完成 N 个检测框的掩膜计算后,通过位与操作完成掩膜的融合,生成时空管道。

10.根据权利要求8所述的一种基于深度学习的运动员行为质量评估方法,其特征在于,时空管道内部管道自注意力机制表示为:

$$y_p = \frac{1}{C(x)} \sum_{\forall c} \sum_{\forall t} \sum_{\forall (i,j) \in \Omega_{c,t}} f(x_p, x_{c,t}(i,j)) g(x_{c,t}(i,j))$$

其中, p 表示需要计算的输出位置, (c,t,i,j) 对时空管道中所有的第一特征位置进行了遍历,输出特征 y 和输入特征 x 具有相同的尺寸, f 函数为距离度量函数, g 函数为特征映射函数,响应值会通过归一化因子 $C(x) = \sum_c \sum_t \Omega_{c,t}$ 进行归一化。

一种基于深度学习的运动员行为质量评估方法

技术领域

[0001] 本发明涉及运动员行为质量评估技术领域,尤其是涉及一种基于深度学习的运动员行为质量评估方法。

背景技术

[0002] 近年来,随着体育运动实时转播技术的不断发展,越来越多的正规体育赛事均对运动员的比赛进行了全程录像,并且对运动员的最终得分进行了保存。如何更好地利用这些数据为运动员后续训练中的技能提升带来更有效的帮助成为一个值得研究的问题。目前虽然已经有一些算法通过视觉信息对运动员的姿势进行评估,但是仅仅局限在姿势和短时动作序列感知的层次,不能够对长时间的复杂动作序列进行高效分析。因此体育运动训练中急需一种能够自动对运动员动作序列进行打分与评估的智能系统,以节约运动员训练阶段的人力物力成本,提高训练效率。

[0003] 现有技术中,针对运动员的姿势与行为序列评估,现有的模型重点关注两个技术点:运动感知技术和运动评估技术。

[0004] 运动感知技术往往被首先考虑。运动感知指通过原始视频和图像信息对运动员的位置进行定位、检查运动员姿势和语义分割。目前此类算法已经很多,例如面向物体检测技术的YOLO检测器、面向姿势估计任务的Alphapose算法和面向语义分割任务的Mask R-CNN。此类算法已经在各类公开数据集中取得了优良的性能,虽然能够直接被运动评估系统所采用,但是这些算法均基于深度学习技术,运行过程对设备有较高的需求,限制了其使用场景。

[0005] 在完成运动感知之后,需要使用运动评估技术对运动序列进行整体评估,得到最终预测结果,并且对低分行为进行检测与反馈。目前虽然有一些工作对行为质量评估模型进行了设计,但是这些模型统一地将整个视频信息做为输入,忽略视频中运动员和背景信息的差异性。这种无差别的特征提取与特征增强一方面会拖慢评估模型的运行效率,另一方面会导致视频信息的混杂从而影响最终的行为质量评估结果。

[0006] 因此,现有技术的缺点主要体现在以下三个方面:

[0007] 1、评估精度低:目前基于视频的行为质量评估技术都只将原始视频作为输入,通过3D卷积核(Convolution kernel)对视频进行特征提取,最终通过回归器完成分数预测。这种处理方法在特征提取过程中并没有考虑前景和背景的差异性。例如运动员的运动区域应当受到更多关注,而不是背景的广告和观众。这种统一处理方法会导致重要信息淹没在繁杂的背景信息中,最终导致模型的评估性能变差。

[0008] 2、运行速度慢:3D卷积网络的训练和推理过程会占用大量的内存,并且需要设备具备极高的算力。这些问题会导致视频分析系统运行的延时过长,无法提供及时的行为质量反馈,最终降低训练效率。因此低参数量、低计算量、高计算效率成为一个优良行为质量评估系统必备的特点。

[0009] 3、可解释性差:基于LSTM(Long Short-Term Memory)的行为质量评估系统的信息

处理过程共分为逐帧特征提取和特征联合分析两个大步骤。首先通过2D卷积神经网络对视频中的每一帧进行特征提取,之后使用LSTM对特征序列进行聚合分析,最终完成行为质量预测。此种方法只能对整个视频进行分析,无法精确到每个运动阶段,因此无法对视频进行分段评估并提供改进意见。

发明内容

[0010] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种基于深度学习的运动员行为质量评估方法,基于的评估系统包括人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元,得分预测单元和显示单元,人体跟踪单元通过对原始比赛视频中的运动员进行跟踪以获取连续的检测框;人体姿势估计单元对运动员身体的关键点进行检测;动作序列特征提取与增强单元分别使用I3D卷积神经网络和管道自注意力机制(Tube Self-attention Mechanism)完成特征的提取和特征增强,得到视频特征;得分预测单元以视频特征作为输入,并且对行为质量评估结果进行预测。整个过程中,从原始比赛视频中分别提取跟踪结果和姿势估计结果,并通过I3D神经网络进行特征提取,使用管道注意力机制完成特征的高效和有效增强,最终实现高精度行为质量评估。质量评估结果为运动员的行为质量提供局部和全局的分析,运动员可以进行针对性的训练,能够节约运动员训练阶段的人力物力成本,提高训练效率,更具指导意义。

[0011] 本发明的目的可以通过以下技术方案来实现:

[0012] 一种基于深度学习的运动员行为质量评估方法,其所基于的评估系统包括人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元,得分预测单元和显示单元,具体的,每个单元内的流程如下:

[0013] 视频输入至人体跟踪单元,人体跟踪单元对视频的各帧进行目标检测,得到各帧的检测框作为跟踪结果,在显示单元对跟踪结果进行可视化;

[0014] 人体姿势估计单元获取跟踪结果,对各帧中运动员的姿势进行估计,得到关键点信息作为姿势估计结果,在显示单元对姿势估计结果进行可视化;

[0015] 动作序列特征提取与增强单元以视频、跟踪结果和姿势估计结果作为输入,进行特征提取、特征增强和特征融合后得到视频特征;

[0016] 得分预测单元以视频特征为输入,进行全动作流程质量评估和阶段性行为质量评估;

[0017] 根据对视频中运动员的运动行为的质量评估,得到运动员的运动行为的不足,基于运动行为的不足对运动员进行专项训练。

[0018] 进一步的,所述人体跟踪单元使用YOLO检测器和Siammask框架进行目标检测:在视频的初始帧使用YOLO检测器进行目标检测,得到初始帧的检测框,使用Siammask框架作为单目标跟踪器在初始帧之后的连续帧进行目标检测,得到初始帧之后各帧的检测框。

[0019] 进一步的,所述人体姿势估计单元使用Alphapose框架对各帧中运动员的姿势进行估计:获取视频中各帧的检测框后,采用Alphapose框架对各帧中运动员的姿势进行估计,Alphapose框架生成各帧的关键点信息,使用卡尔曼滤波器算法对各帧的关键点信息进行处理,得到姿势估计结果。

[0020] 进一步的,所述动作序列特征提取与增强单元对视频进行特征提取,得到第一特

征,将第一特征结合跟踪结果使用管道自注意力机制进行特征增强,得到增强后的第一特征,动作序列特征提取与增强单元对姿势估计结果进行特征提取,得到第二特征,将第二特征与增强后的第一特征通过全连接层进行融合得到视频特征。

[0021] 进一步的,所述动作序列特征提取与增强单元使用I3D神经网络对视频进行特征提取,得到第一特征,使用图卷积神经网络对姿势估计结果进行特征提取,得到第二特征。

[0022] 进一步的,进行全动作流程质量评估时,将所有的视频特征进行时序全局平均池化后送入全连接层完成质量评估,进行阶段性行为质量评估时,将一段视频的视频特征送入全连接层完成质量评估。

[0023] 进一步的,得分预测单元使用I3D神经网络,进行全动作流程质量评估时,将所有的视频特征进行时序全局平均池化后送入I3D神经网络的全连接层完成质量评估,进行阶段性行为质量评估时,将一段视频的视频特征送入I3D神经网络的全连接层完成质量评估。

[0024] 进一步的,将第一特征结合跟踪结果使用管道自注意力机制进行特征增强具体为:将第一特征与检测框进行量化和对齐生成特征图掩膜,按照视频的帧数和第一特征的数量比例进行掩膜的融合以生成时空管道,在时空管道内部采用管道自注意力机制完成第一特征的稀疏增强操作,通过残差连接将第一特征与完成稀疏增强后的第一特征进行融合,得到增强后的第一特征。

[0025] 进一步的,获取跟踪结果和第一特征后,按照视频的帧数和第一特征的数量比例确定检测框的数量与第一特征的时序维度数量的比值 $N:1$, $N>1$,确定每一个检测框所对应的掩膜,获取检测框覆盖第一特征的特征网络的比例,如果比例大于预设置的阈值,则该第一特征被选中,否则,该第一特征被剔除,完成 N 个检测框的掩膜计算后,通过位与操作完成掩膜的融合,生成时空管道。

[0026] 进一步的,时空管道内部管道自注意力机制表示为:

$$[0027] \quad y_p = \frac{1}{C(x)} \sum_{\forall c} \sum_{\forall t} \sum_{\forall (i,j) \in \Omega_{c,t}} f(x_p, x_{c,t}(i,j)) g(x_{c,t}(i,j))$$

[0028] 其中, p 表示需要计算的输出位置, (c,t,i,j) 对时空管道中所有的第一特征位置进行了遍历,输出特征 y 和输入特征 x 具有相同的尺寸, f 函数为距离度量函数, g 函数为特征映射函数,响应值会通过归一化因子 $C(x) = \sum_c \sum_t \Omega_{c,t}$ 进行归一化。

[0029] 与现有技术相比,本发明结合人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元、得分预测单元和显示单元,实现了运动员位置与姿势信息的多通道感知,通过I3D网络完成了视频特征提取,并且提出了管道自注意力机制借助检测框信息完成运动员动作特征的高效和有效增强,通过特征融合技术完成了视频特征和第二特征的融合,并在得分预测单元设计了全局和局部两种行为质量评估模式。本系统能够在全行为层次和各个动作阶段层次高效地完成运动员行为质量评估,可以节约运动员培训过程中大量的人力和物力,提高训练效率和质量,为运动员竞技水平的不断提升和体育事业的发展提供有力基础保障。

附图说明

[0030] 图1为本发明的系统原理框图;

- [0031] 图2为人体跟踪单元的原理框图；
[0032] 图3为人体姿势估计单元的原理框图；
[0033] 图4为动作序列特征提取与增强单元的原理框图；
[0034] 图5为得分预测单元的原理框图；
[0035] 图6为管道自注意力机制示意图。

具体实施方式

[0036] 下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案为前提进行实施,给出了详细的实施方式和具体的操作过程,但本发明的保护范围不限于下述的实施例。

[0037] 实施例1:

[0038] 一种基于深度学习的运动员行为质量评估方法,如图1所示,其所基于的评估系统包括人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元,得分预测单元和显示单元。

[0039] 人体跟踪单元解决了传统目标检测方案中漏检、误检的问题。一般的追踪算法首先使用逐帧检测算法获取到每一帧的约束框(即本申请中所提到的检测框,也可以称为追踪框、跟踪框等),之后通过卡尔曼滤波器(Kalman Filter)和人体重识别(Person Re-identification, ReID)算法完成跟踪任务。然而这种方法仅仅适用于一般监控环境中,要求人体不能够在短时间内发生巨大形变,并不适合人体姿势经常处于高度扭曲和高速运动状态下的体育运动场景。因此本发明将单目标跟踪器(Single Object Tracker, SOT)引入到了运动员行为质量评估任务中。和普通跟踪算法中的逐帧检测策略不同,单目标跟踪器跳过了逐帧检测阶段,在给定首帧约束框的前提下能够在连贯的时序中输出稳定的跟踪结果,为后续特征增强提供位置信息,本申请首帧(或称为初始帧)使用YOLO检测器进行目标检测,之后根据检测框大小关系进行无关目标滤除,为单目标跟踪器提供初始帧信息。

[0040] 人体姿势估计单元能够在空间和时间两个维度中对人体姿势进行联合分析,最终获取到高精度人体关键点信息。传统的姿势估计算法往往只关注单帧场景,视频中的姿势估计算法也仅仅是将算法用于逐帧检测再拼接。这种简单的迁移方法并不能妥善处理视频中运动员的身体扭曲和自遮挡问题。本发明在单帧关键点检测的基础上添加了基于卡尔曼滤波器的跟踪机制,用于处理关键点的误检与漏检。此外,本发明还将人体姿势估计单元的神经网络进行3-bit量化,以节约计算资源,且提高神经网络的推断效率。

[0041] 动作序列特征提取与增强单元以视频、跟踪结果和姿势估计结果作为输入,输出视频特征。首先使用I3D神经网络对运动视频片段进行特征提取,之后将检测框和第一特征(也可以成为视频特征图)进行量化和对齐生成特征图掩膜(Mask),并且按照视频帧数与特征图数量的比例进行掩膜的融合以生成时空管道(Spatio-temporal Tube)。在时空管道内部采用管道自注意力机制完成第一特征的稀疏增强操作,通过残差连接(Residual link)将增强之后的第一特征和原来的第一特征进行融合,得到增强后的第一特征。使用了图卷积神经网络(Graph Convolutional Network, GCN)用于姿势估计结果的特征提取,得到第二特征。第二特征(即姿势估计结果的特征)和增强之后的第一特征将通过I3D神经网络的全连接层进行信息融合生成视频特征,用于后续的行为质量评估。

[0042] 得分预测单元(进行行为质量评估)将动作序列特征提取与增强单元输出的视频特征作为输入,在I3D神经网络进行预测,输出聚合之后的视频特征,并完成最终得分预测。得分预测单元共分为两种模式:全动作流程质量评估和阶段性行为质量评估。在全动作流程质量评估模式中,需要首先将所有视频段提取到的视频特征进行时序全局平均池化,之后再送入全连接层完成预测;在阶段性行为质量评估模式中,将每段视频的视频特征直接送入全连接层完成预测,因此可以观察运动员在每个阶段中动作的优劣程度。

[0043] 显示模块能够将上述执行过程中每一个阶段的结果和最终结果进行可视化,并针对阶段得分提供动作改进意见。

[0044] 具体的,每个单元内的流程如下:

[0045] (一)视频输入至人体跟踪单元,人体跟踪单元对视频的各帧进行目标检测,得到各帧的检测框作为跟踪结果,在显示单元对跟踪结果进行可视化;

[0046] 本实施例中,人体跟踪单元的原理图如图2所示,所述人体跟踪单元使用YOLO检测器和Siammask框架进行目标检测:在视频的初始帧使用YOLO检测器进行目标检测,得到初始帧的检测框,使用Siammask框架作为单目标跟踪器在初始帧之后的连续帧进行目标检测,得到初始帧之后各帧的检测框。

[0047] (1)给定首帧追踪框:传统的跟踪方法由单帧目标检测和连续帧检测结果融合两个模块构成,这种方法使得追踪结果严重受制于检测结果(即在单帧中检测到目标,本申请中,待检测的目标为运动员),而体育视频中的检测效果不容乐观。运动员的高速运动和严重形变会导致检测器的漏检和误检,同时背景中的观众也会干扰运动员的目标检测。因此本发明采用了基于单目标跟踪器的策略。通常体育动作执行的准备阶段运动员处于静止状态,较容易被目标检测器识别。因此本发明首先使用YOLO检测器在初始帧完成目标检测,之后根据检测框大小关系进行无关目标滤除,为单目标跟踪器提供初始帧信息。

[0048] (2)单目标跟踪:目前单目标跟踪领域中已经有很多成熟的框架相继被提出,本申请考虑到运动员目标跟踪的特殊性,采用了Siammask框架作为单目标跟踪器。Siammask是一种能够实时完成视觉目标跟踪和半监督目标分割的简单方法。Siammask在训练过程中采用了全连接孪生网络结构,并使用二进制分割任务对损失函数进行增强;在测试时能够以55FPS生成目标分割掩膜和旋转约束框。单目标跟踪器策略的采用弥补了传统基于单帧检测方法中的检测难题,可以获得更加精确的跟踪框与分割信息,为后续特征增强和行为质量评估提供重要参考信息。

[0049] (3)跟踪结果可视化:此部分是显示模块中的组成部分。Siammask能够同时生成跟踪目标的约束框和分割掩膜,因此本发明在显示模块中对跟踪结果进行可视化,为教练和运动员提供参考。

[0050] (二)人体姿势估计单元获取跟踪结果,对各帧中运动员的姿势进行估计,得到关键点信息作为姿势估计结果,在显示单元对姿势估计结果进行可视化;

[0051] 本实施例中,人体姿势估计单元的原理图如图3所示,所述人体姿势估计单元使用Alphapose框架对各帧中运动员的姿势进行估计:获取视频中各帧的检测框后,采用Alphapose框架对各帧中运动员的姿势进行估计,Alphapose框架生成各帧的关键点信息,使用卡尔曼滤波器算法对各帧的关键点信息进行处理,得到姿势估计结果。

[0052] (1)单帧姿势估计:本发明采用Alphapose框架完成单帧中运动员的姿势估计。

Alphapose框架的设计初衷是处理两种问题,定位错误问题和冗余检测问题。定位错误问题由指检测器所给定的框和真实框之间存在差异引起,即虽然两者交并比(IoU>0.5)满足筛选要求,但是检测框可能只包含部分人体信息,导致最终的误检和漏检;冗余检测问题由NMS产生的重复检测框造成。为了处理这些问题,Alphapose使用区域多人姿势估计框架完成约束框位置不确定情况下的姿势估计。

[0053] (2) 卡尔曼滤波器平滑:完成单帧姿势估计之后,需要在时序上对关键点信息进行平滑处理。本发明采用卡尔曼滤波器算法对Alphapose框架生成的关键点信息进行处理。卡尔曼滤波器的本质是一组数学方程,用递归的方式来估计过程的状态,也即最小化根误差的均值。卡尔曼滤波器能够在系统精度未知的情况下支持对过去、现在和未来状态的估计。其目的是对系统的状态列向量 x 进行估计,一般通过一个含有随机量的差分方程来对系统的状态进行估计:

$$[0054] \quad x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}$$

[0055] 其中的 x_{k-1} 是当前时刻的状态, x_k 是下一时刻的状态。 A 是大小为 $n \times n$ 的转移矩阵, B 是大小为 $n \times 1$ 的控制矩阵。 w_{k-1} 是状态转移过程种的噪音。由于对系统的观测并不是完美的,会存在一些测量噪音,故观测方程为:

$$[0056] \quad z_k = Hx_k + v_k$$

[0057] 其中 H 是大小为 $m \times n$ 观测矩阵,将 $n \times 1$ 状态转化为 $m \times 1$ 观测值,同时添加观测过程的偏差 v_k 。假设状态转移过程噪声 w 和测量噪声 v 都服从正态分布:

$$[0058] \quad \begin{cases} p(w) \sim N(0, Q) \\ p(v) \sim N(0, R) \end{cases}$$

[0059] 其中 Q 称为过程噪声协方差矩阵, R 称为测量噪声协方差矩阵。

[0060] (3) 姿势估计结果可视化:此部分是显示模块中的组成部分。关键点的位置和置信度信息均以不同颜色和不同透明度在原始视频中进行标注,为教练和运动员提供参考。

[0061] (三) 动作序列特征提取与增强单元以视频、跟踪结果和姿势估计结果作为输入,进行特征提取、特征增强和特征融合后得到视频特征;

[0062] 本实施例中,动作序列特征提取与增强单元的原理图如图4所示,动作序列特征提取与增强单元对视频进行特征提取,得到第一特征,将第一特征结合跟踪结果使用管道自注意力机制进行特征增强,得到增强后的第一特征,动作序列特征提取与增强单元对姿势估计结果进行特征提取,得到第二特征,将第二特征与增强后的第一特征通过全连接层进行融合得到视频特征。

[0063] (四) 得分预测单元以视频特征为输入,进行全动作流程质量评估和阶段性行为质量评估;

[0064] (1) 全动作流程质量评估:本发明可以在全局层次完成运动视频的整体评估,并且给定最终动作得分。本发明将视频特征的提取与增强和得分预测阶段进行解耦,并通过在两部分之间添加时序平均池化层将特征(即整个视频的视频特征)变换为1024-d,实现全动作流程质量评估。

[0065] (2) 阶段性行为质量评估:本发明可以在局部层次对每个视频片段的行为质量进行评估。每个视频片段在经过特征提取和特征增强之后会变为1024-d特征向量(即每个视频片段的视频特征),基于此特征向量实现阶段性行为质量评估。

[0066] (3) 行为质量预测模块: 本发明使用多层全连接层完成从特征向量到行为质量的映射, 本实施例中采用的网络层为: {FC (1024-512), ReLU}, {FC (512-128), ReLU}, {FC (128-1)}。

[0067] 本实施例中, 得分预测单元的原理图如图5所示, 得分预测单元使用I3D神经网络, 进行全动作流程质量评估时, 将所有的视频特征进行时序全局平均池化后送入I3D神经网络的全连接层完成质量评估, 进行阶段性行为质量评估时, 将一段视频的视频特征送入I3D神经网络的全连接层完成质量评估。

[0068] 在动作序列特征提取与增强单元与得分预测单元中, 所述动作序列特征提取与增强单元使用I3D神经网络对视频进行特征提取, 得到第一特征, 使用图卷积神经网络对姿势估计结果进行特征提取, 得到第二特征。

[0069] 得分预测单元进行全动作流程质量评估时, 将所有的视频特征进行时序全局平均池化后送入全连接层完成质量评估, 进行阶段性行为质量评估时, 将一段视频的视频特征送入全连接层完成质量评估。

[0070] (1) I3D神经网络全称为双流膨胀3D卷积网络 (Two-Stream Inflated 3D ConvNet), 此网络将2D网络中的滤波器和池化核都进行了3D扩展, 因此视频网络的参数初始化可以通过图像数据集上的预训练模型完成。I3D由Inception网络扩展而来, 其基础构成部件为Inception模块。从整体架构上看, I3D由卷积层、Inception模块和池化层构成。普通的视频识别任务将I3D网络视为一个整体, 视频被转化为特征向量之后用于分类任务、回归任务等。此发明将I3D网络划分为两个阶段, 并且在两个阶段之间通过跟踪结果和姿势估计估计完成特征增强。

[0071] 在I3D网络的第一阶段, 进行特征提取, 设输入的视频共含有L帧, 首先使用Siammask追踪器对视频中的运动员进行跟踪获取到检测框信息。在视频特征提取阶段, 视频将会被划分为N个片段, 每个片段含有M帧连续图像, 本实施例中 $N=10$, $M=16$ 。视频片段会被送入I3D网络的第一阶段完成特征提取过程得到特征, 用于后续特征增强。

[0072] (2) 特征增强: 目前方法无法完成有效且高效的特征增强。卷积操作的感受野有限导致无法对长时间的依存关系建模, RNN需要存储隐状态的特性导致无法高效并行计算。本发明提出将管道机制和自注意力机制进行融合的方法完成行为特征的高效增强。

[0073] 管道自注意力模块将检测框和第一特征作为输入进行特征增强, 并最终生成增强的第一特征。管道自注意力机制并不会改变特征图尺寸, 此特性使得管道自注意力机制能够被嵌入在网络中任意两个层之间, 并且可以进行多层堆叠。

[0074] (3) 特征融合: 本发明通过特征连接的方式将由管道自注意力机制增强得到的第一特征和由图卷积网络得到的第二特征 (即姿势特征) 进行融合, 生成融合特征 X' 。特征 X' 会被送入到I3D的第二阶段完成后续特征提取, 最终生成H, H表示视频特征, 表征了运动员的行为质量。

[0075] 图6为管道自注意力机制的示意图, 展示了检测框的量化和掩膜生成过程:

[0076] 将第一特征结合跟踪结果使用管道自注意力机制进行特征增强具体为: 将第一特征与检测框进行量化 (或者称为离散化) 和对齐生成特征图掩膜, 按照视频的帧数和第一特征的数量比例进行掩膜的融合以生成时空管道, 在时空管道内部采用管道自注意力机制完成第一特征的稀疏增强操作, 通过残差连接将第一特征与完成稀疏增强后的第一特征进

行融合,得到增强后的第一特征。

[0077] 具体的,获取跟踪结果和第一特征后,按照视频的帧数和第一特征的数量比例确定检测框的数量与第一特征的时序维度数量的比值 $N:1$, $N>1$,确定每一个检测框所对应的掩膜,获取检测框覆盖第一特征的特征网络的比例,如果比例大于预设置的阈值,则该第一特征被选中,否则,该第一特征被剔除,完成 N 个检测框的掩膜计算后,通过位与操作完成掩膜的融合,生成时空管道。

[0078] 约束框的量化和掩膜生成过程见图6:在获取到跟踪结果和第一特征(也可以理解为特征图)之后,需要对特征图中被选中的特征进行筛选。而由于I3D网络中第一阶段含有两个时序池化层,检测框数量和特征图的时序维度数量之比并非 $1:1$,而且由于Siammask生成的检测框是歪斜的,因此无法直接完成特征图筛选。针对此问题,本发明提出了一种基于追踪框离散化和聚合的特征图掩膜生成方法,用于构建时空管道。本实例中,检测框数量和特征图时序维度数量之比为 $4:1$ 。需要首先确定每一个检测框所对应的掩膜。之后根据特征网格被检测框所覆盖的比例和预设置的阈值 τ 判断此位置的第一特征是否能够被选中,如果此比例大于阈值 τ ,则此位置的第一特征被选中;如果小于则被剔除。在本实施例中阈值 $\tau=0.5$ 。在完成四个检测框的掩膜计算之后,需要通过位与操作完成掩膜的聚合,得到总掩膜。此掩膜包含所有被选中的第一特征位置,为了表示更加简洁明确,本实施例将其转化为下标集合再参与后续运算。

[0079] 具体的,完成选中第一特征的定位即构建了时空管道,此时可以通过自注意力机制的引入完成特征增强。和Non-local模块保持类似形式,时空管道内部管道自注意力机制表示为:

$$[0080] \quad y_p = \frac{1}{C(x)} \sum_{\forall c} \sum_{\forall t} \sum_{\forall (i,j) \in \Omega_{c,t}} f(x_p, x_{c,t}(i,j)) g(x_{c,t}(i,j))$$

[0081] 其中, p 表示需要计算的输出位置, (c,t,i,j) 对时空管道中所有的第一特征位置进行了遍历,输出特征 y 和输入特征 x 具有相同的尺寸, f 函数为距离度量函数, g 函数为特征映射函数,响应值会通过归一化因子 $C(x) = \sum_c \sum_t \Omega_{c,t}$ 进行归一化。

[0082] 为了减少计算量,本实施例采用点乘操作作为相似性度量函数;

[0083] 为了增加后续I3D网络的特征提取能力,本发明为管道自注意力模块添加了残差连接:

$$[0084] \quad x'_p = W_z y_p + x_p$$

[0085] 其中 x'_p 和 x_p 具有完全相同的尺寸, W_z 是连接的参数,因此管道自注意力机制能够被嵌入在网络中的任何位置。为了能够达到高性能和高计算效率的平衡,本实施例将管道自注意力模块放置于Mixed_4e层之后,因此 $T=4$, $H=W=14$ 。

[0086] 和Non-local模块相比,管道自注意力模块在完成特征增强时并没有将所有的特征纳入考虑,而是在时空管道的基础上使用自注意力机制完成特征增强。此种策略在极大程度上削减了计算量。

[0087] (五) 根据对视频中运动员的运动行为的质量评估,得到运动员的运动行为的不足,基于运动行为的不足对运动员进行专项训练。

[0088] 本实施例结合人体跟踪单元、人体姿势估计单元、动作序列特征提取与增强单元、

得分预测单元和显示单元,实现了运动员位置与姿势信息的多通道感知,通过I3D网络完成了视频特征提取,并且提出了管道自注意力机制借助检测框信息完成运动员动作特征的高效和有效增强,通过特征融合技术完成了视频特征和第二特征(即基于姿势估计结果得到姿势特征,姿势估计结果可以简单理解为运动员身体的关键点)的融合,并在得分预测单元设计了全局和局部两种行为质量评估模式。本系统能够在全行为层次和各个动作阶段层次高效地完成运动员行为质量评估,可以节约运动员培训过程中大量的人力和物力,提高训练效率和质量,为运动员竞技水平的不断提升和体育事业的发展提供有力基础保障。

[0089] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思作出诸多修改和变化。因此,凡本技术领域技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案,皆应在由权利要求书所确定的保护范围内。

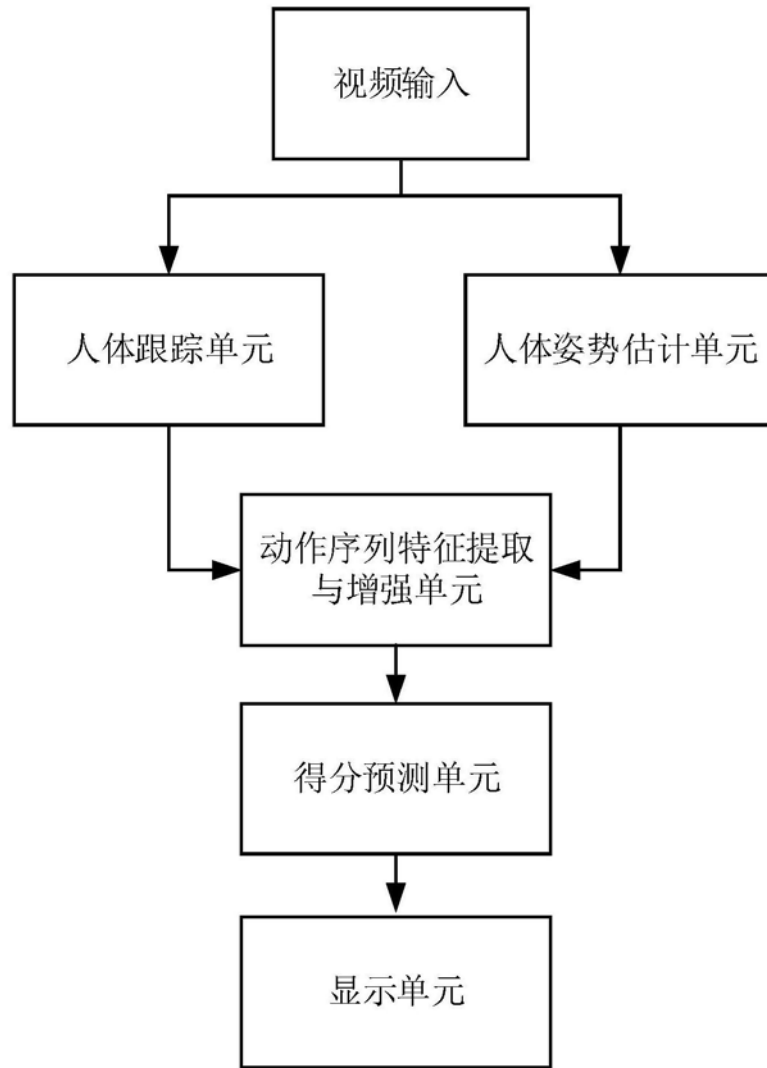


图1

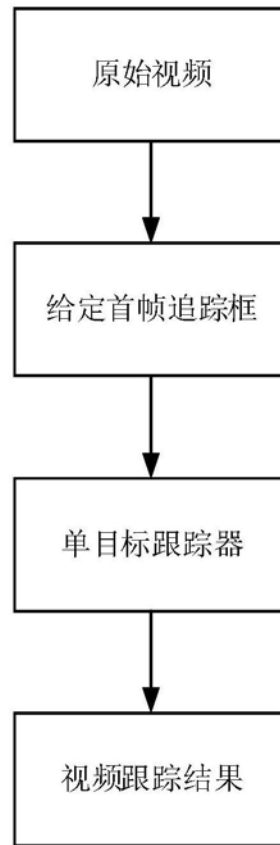


图2

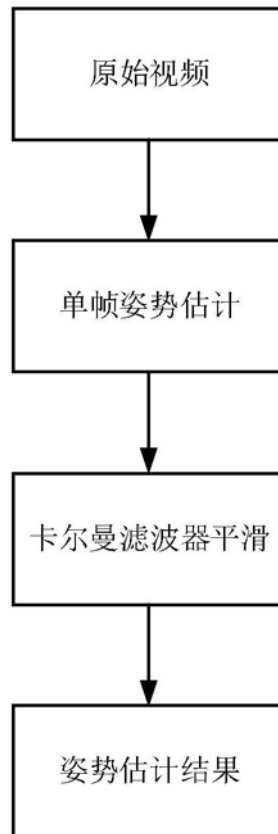


图3

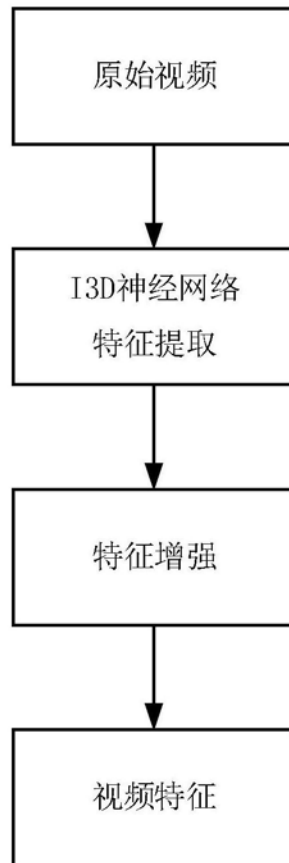


图4

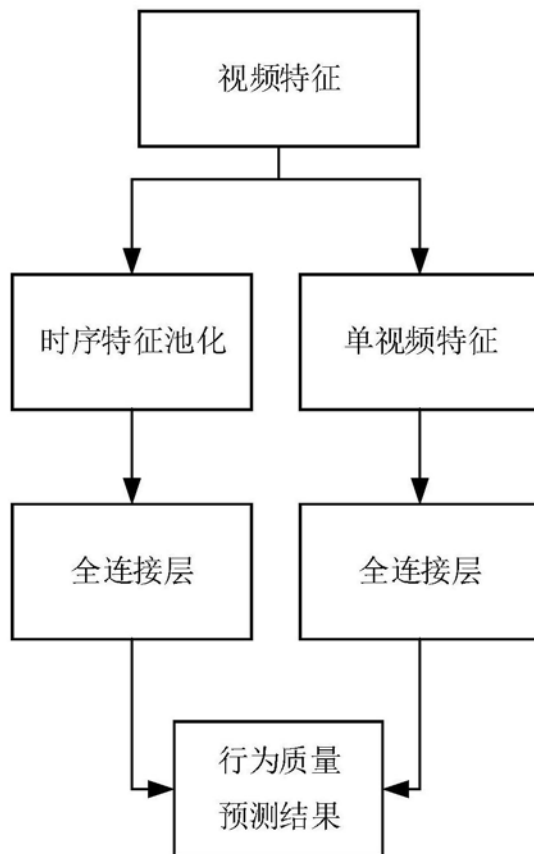


图5

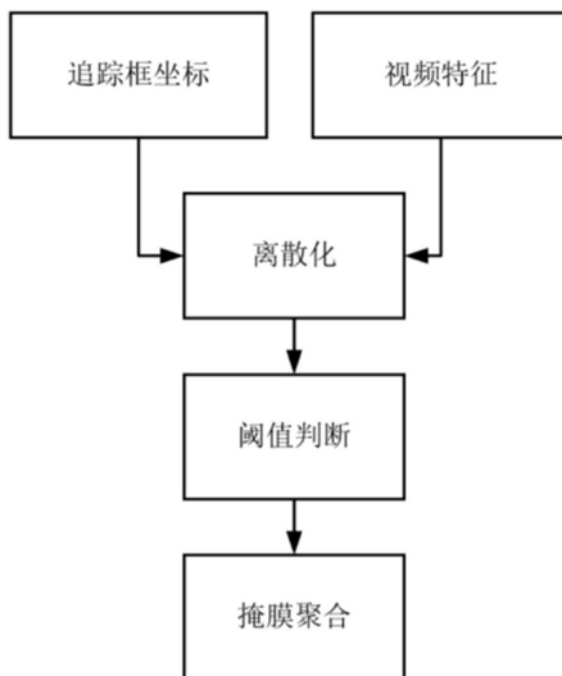


图6