“基于深度学习的人体行为识别综述” 河南工业大学 信息科学与工程学院 邓淼磊，振东，李磊，陈斯

介绍了人体行为识别任务中传统特征提取方法和基于深度学习的特征提取方法；讨论了以元学习（meta-learning）和transformer为代表的FSL算法将成为未来模型主流算法的可能性，同时对未来基于深度学习的人体行为识别算法的发展方向进行展望。

现有技术分析：

视频智能分析的人体行为识别技术在医疗监护领域可以用于患者的实时监控和意外跌倒检测；

介绍的深度学习算法：

1. 在空间维度中引入时间维度特征，以此获得视频中相邻帧之间的上下文关系，将其命名为3D CNN，在人工智能领域表现优异但是计算成本大，用来提取行为特征容易梯度消失、梯度爆炸、过拟合。
2. 混合深度学习网络：目前主流的混合深度学习网络以3D卷积神经网络为基础，融合其他深度学习方法。
3. 双流卷积神经网络：提取人体静态表观特征和动态运动特征，时间和空间两方面对特征进行深度分析，人体识别领域具有良好的识别效果。可以识别单人行为和双人交互行为。
4. Few-shot learning（FSL）：少样本学习，使得模型在有限样本中快速适应新的类别。

深度学习人体行为识别的现状以及问题分析：

1. 跨场景识别困难。现有模型常常在指定场景训练，无法跨场景，重新训练增加工作量。如何在模型训练过程中融入环境信息，是一个研究方向。
2. 部署困难。高识别度模型常常有高复杂度，如何保证精度同时降低复杂度，是个有价值的研究方向。
3. 行为预测。大部分算法只能够事后识别，不能事前预防，对于实时视频，需要准确判断，对判断具有前瞻性，做到事前预防，需要深入研究。
4. 少样本学习模型问题。许多场景样本采集是难以解决的问题，数据少无法有效训练。FSL模型虽然减少对样本的依赖，但是识别的精度有下降。如何减少样本提高模型性能还具有高识别精度，是一个值得探索的方向。

“基于骨骼数据特征的人体行为识别方法综述” 安庆师范大学 计算机与信息学院

从基于传统机器学习的手工特征提取方法和基于深度学习的深度特征提取方法两方面对基于骨骼数据的人体行为识别相关研究进行综述。

目前应用于基于骨骼数据的人体行为识别研究的深度学习网络主要有CNN，RNN，GCN（图卷积神经网络）以及混合神经网络等。CNN、RNN、GCN三种深度学习网络处理的关节点数据表示方式依次为伪图像、长向量和拓扑图。深度神经网络提取的具体步骤是：首先将三维骨骼序列数据处理成深度学习网络能学习的数据格式，然后输入深度学习网络，通过网络自学习深度特征进行动作分类。

三维骨骼序列数据中既含有空间结构信息又含有时间动态信息，这两者是准确识别动作类型的重要因素。

图卷积神经网络：CNN和RNN均针对排列规则的欧氏数据进行学习而不能直接处理非欧式数据。人体骨骼有很多非欧氏数据的特征，可将人体骨骼视为由关节点和骨骼组成的图数据，采用图卷积神经网络直接对图数据学习。

混合神经网络采用CNN+LSTM、CNN+GCN、GCN+Bi-LSTM等不同类型的深度网络组合。通过合理组合使得整体效果达到更好。

对比之下，运用GCN能比RNN、CNN更准确地识别出动作类型，原因是将骨骼数据视为图数据比表示为伪图像和长向量更合适。

长向量表示：骨骼数据可以被表示为长向量，其中每个关节的坐标按时间序列排列。这种表示方式将骨骼数据视为一维序列，然后可以使用循环神经网络（RNN）或卷积神经网络（CNN）进行处理。但这种表示可能会忽略骨骼关节之间的拓扑结构和关联信息。

伪图像表示：伪图像表示是一种将骨骼数据呈现为类似于图像的形式，其中关节位置在二维图像上的位置反映了它们之间的关系。这可以通过在一个固定大小的网格上将关节位置映射为像素值来实现。这种方式将骨骼数据的关系视为图像中的像素点之间的关系，使得可以应用基于图的方法，如GCN，来考虑关节之间的拓扑结构和依赖关系。

Application Research of Optical Tracking Point Layout in Computer Motion Capture Technology Take Walking Motion as an Example

The layout of optical tracking points often dramatically influences the efficiency and quality of optical motion capture.

光学跟踪点的布局通常显著影响光学运动捕获的效率和质量。

Motion capture devices can be classified into sensors, signal capture devices, data transmission devices, and data processing devices.

运动捕捉设备可以分为传感器、信号捕捉设备、数据传输设备和数据处理设备。

The position where the sensor is fixed is called the tracking point.

传感器固定的位置称为跟踪点。

The most commonly used optical motion capture system is considered: the optical motion capture system relies on the tracking of sensor spots to complete motion capture, which has the advantages of high sampling rate, free sports performance, low requirements on the site, etc.

最常用的光学运动捕捉系统被认为是：光学运动捕获系统依靠对传感器斑点的跟踪来完成运动捕捉，具有采样率高、运动性能自由、对现场要求低等优点。

Joints are points, bones are rods, and the whole limbs are connected in series according to joints' kinematic relationship.

关节是点，骨骼是杆，整个肢体根据关节的运动关系串联在一起。

At the beginning of the motion, the a node drives the β node, which in turn drives the y node, while the y node does not drive the β node and a node when it moves alone, thus forming a hierarchical relationship structure which can be embodied in the operation mode.

Combined with the hierarchical relationship of joint association, we can regard the whole human body movement as the combination of movement around waist node (a node) and horizontal rotation, and that of rotation of each associated child node (y node) around parent node (β node).

在运动开始时，a节点驱动β节点，β节点反过来驱动y节点，而y节点单独运动时不驱动β节点和a节点，从而形成了可以体现在操作模式中的层次关系结构。

结合关节关联的层次关系，我们可以将整个人体的运动视为围绕腰部节点（a节点）的运动和水平旋转的组合，以及每个关联的子节点（y节点）围绕父节点（β节点）的旋转的组合。

In using motion capture technology to create walking motion, the skeleton motion of the action performer can be decomposed into the horizontal displacement of the waist node and regular motion of lower limbs.

在使用运动捕捉技术创建行走运动时，动作执行者的骨骼运动可以分解为腰部节点的水平位移和下肢的规则运动。

