基于3D卷积神经网络的行为姿态识别文献综述

**一、引言**

在计算机技术广泛应用的今天，机器已经大规模的代替人力用于各种事务和工作，而安防工作也是其中的一个热点，如果计算机能够自动的识别特定的动作信息并产生应急信息，将很大程度上的提高发现率并减少人力的消耗。

相比于原始的Human Action Recognition方法进行人体行为识别所需要的苛刻环境条件，CNN的价值就在于它是一种可以直接在原始输入就最大程度运算的深度模型。而高于2D卷积神经网络之上的，3D卷积神经网络应用于动态运动识别，可以更加广泛而实际的应用于现实生活。

**二、方法**

人体姿态识别的一种比较简单的思路是将数据集中视频及分类标签转换为图像（视频帧）和其对应的分类标签，再采用卷积神经网络对图像进行训练学习和测试，将视频分类问题转化为图形分类问题。这种工作和逻辑较为简单的方法同时也使得测试的效率和准确率相对的较低。

通过对每一帧视频运用卷积神经网络来识别，最大的弊端就在于没有考虑到连续帧间的运动信息。为了有效的综合运动信息，所阅读论文中提出了一种高效的3D卷积的方法。通过在CNNs的卷积层进行3D卷积，以捕捉在时间和空间维度都具有区分性的特征。新的3D卷积是通过堆叠多个连续的帧组成一个立方体，然后在立方体中运用3D卷积核，对空间和时间维度上的信息进行操作，在这个结构中，卷积层中每一个特征map都会与上一层中多个邻近的连续帧相连，因此捕捉运动信息。接下来，基于3D卷积特征提取器构造了一个3D卷积神经网络。这个架构可以从连续视频帧中产生多通道的信息，然后在每一个通道都分离地进行卷积和下采样操作。最后将所有通道的信息组合起来得到最终的特征描述。在增强模型方面，模型通过计算高层运动特征得到的辅助输出，同时为应对不同环境综合不同的卷积模型架构来综合判断数据。这种方法在准确率和效率上都将是一个有效的提高。

两种方法的对比可以看出，动态的行为识别与静态的姿态识别相比最大的难度就在于如何将时间维度地数据有效的与空间维度的数据相结合，使相邻卷积层之间产生有利于识别工作的关联，产生多通道信息，使得3D卷积神经网络可以从时间和空间的三个维度上进行卷积。

**三、总结**

目前采用所阅读文献中的第一种方法，在UCF-101数据集上进行研究，到目前位置所能够识别的动作有101个，但是由与设备有限，而且模型比较简单，效果并不是非常理想，不过这也验证了论文中说法。之后需要将上文所述的第二种方法加以应用，从目前的分析看，如何将多帧图像联系起来是技术上的一大难点，而如何能够减小甚至去除不包含动作的帧的干扰也是需要考虑的问题之一。除了上述问题，还有一些较小的问题不在此加以赘述。从目前的研究及文献的查阅看，利用CNN进行姿态识别，需要对视频的处理方法加以改进，同时需要去研究目前更为前沿的方法。

**参考文献**：

**[1]** Beyond Short Snippets: Deep Networks for Video Classification Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, George Toderici

**[2]** 基于3D卷积神经网络的行为识别 余凯