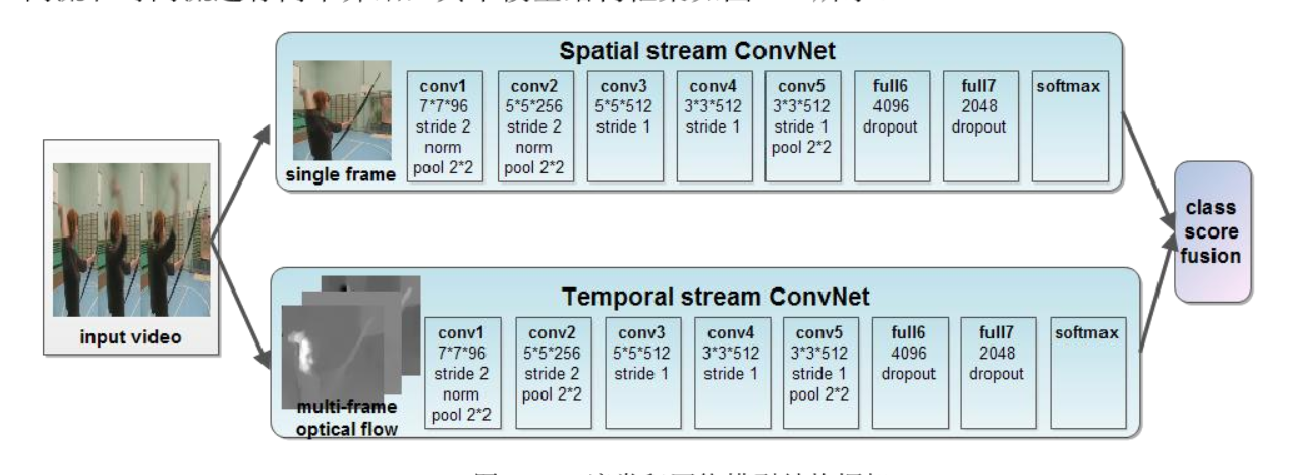
人工特征通常是基于受控环境的领域知识而设计，大多数方法遵循两步法，**第一步是从原始视频帧计算特征；第二步是根据获得的特征来学习分类器。**在实际应用中，很难知道那种特征比较适合于需要研究的内容，因此特征的选择与问题能否较好的解决密切相关。特别是对于人类动作识别而言，不同的动作类别在外观和动作模式上可能会显着不同，因此人工特征的泛化能力不足够。

而**深度学习模型可以通过样本学习特征**，从而具有比传统行为识别方法更好的优势。随着深度学习方法在图像分类和视觉目标检测等领域取得了巨大的成功，人们逐渐开始将深度学习运用于视频分析，行为识别。

提出了一种 3D CNN 网络模型，通过将 3D 卷积核取代 2D 卷积核进行卷积操作，进而同时获取空间信息和时间信息。该网络模型是将连续多帧视频图像生成多个信息通道，并在每个通道上进行卷积处理，下采样处理，最后将所有通道的信息合并得到时间-空间特征。



该模型结构由两个网络模型组成，分别为空间流卷积网络、时间流卷积网络。空间流卷积网络的输入数据是单一视频帧图像，该操作能够有效地在静态图像上实现对行为的识别。由于某些运动与特定目标对象有较强的联系，因此静态图像本身就是一个有用的线索。事实上，前期深度学习下的行为识别方法大部分是基于静止的视频帧图像，并取得了一定的效果。空间流卷积网络本质上就是用于图像分类的网络模型结构，因此该网络可以直接采用大量样本图像分类的深度网络模型。时间流卷积网络的输入数据是堆叠的连续**视频帧间光流位移场**，该输入能够明确地描述视频帧图像的运动信息，使得网络不需要进行运动估计，从而让行为识别变操作变的比较简单。下面简单介绍时间流卷积网络的输入：

本节对一些常用的特征进行介绍，其中包括稠密轨迹特征，光流直方图特征，梯度直方图特征，运动边界直方图特征。

融合方法主要分为两个层面,特征融合和结果融合。特**征融合的方法主要是串行特征融合，加权特征融合，以及基于这两种融合方法衍生的一系列求特征相关系数后进行串行特征融合或者加权特征融合。**

直到最近，香港中文大学的Wang Li-Ming博士在欧洲计算机视觉（ECCV）会议上提出的视频时间分割思想，把视频的时间变化性引入到深度网络的计算中，突破了以往卷积网络只能学习视频片段特征的局限性，大大地提高了行为识别的性能。进而推动了深度学习在视频应用上的发展。

香港中大-商汤科技联合实验室，汤晓鸥教授是人工智能和深度学习领域内的领军人物，目前供职于香港中文大学，是香港中文大学信息工程系系主任，兼任中国科学院深圳先进技术研究院副院长。

颜水成供职于：新加坡国立大学、360人工智能研究院擅长领域：深度学习、人脸识别应用：智能摄像头、行车记录仪新加坡国立大学副教授，2015年加入360担任首席科学家，负责建立和领导360人工智能研究院。颜水成教授曾入选汤森路透发布的“2014年世界最具影响力科学家”，同时2014和2015两次入选ISI Highly-cited researchers。此外，他的团队在五年内曾7次问鼎计算机视觉领域“World Cup”竞赛 PASCAL VOC 和 ILSVRC的世界冠军和亚军奖项。由于在计算机视觉，机器学习和多媒体等领域的杰出贡献，颜水成被授予新加坡青年科学家奖 (YSA)和青年教授研究成就奖。

受深度神经网络在图像描述生成应用研究的启发，基于深度学习的视频描述生成的研究于2015年前后开始成为新的研究热点。与此同时，为了促进深度神经网络在视频描述生成领域的应用，微软亚洲亚洲研究院的Mei Tao博士在2016年的计算机视觉和模式（CVPR）会议上公布了大规模的视频描述数据集（MSR-VTT）。自此，用于视频描述生成的深度模型有了很好的语料库，而训练深度的视频描述生成模型也成为了可能。

为了解决数据量不足问题，迁移学习是训练深度网络的实用技巧。具体来说，就是将深度卷积网络在其他大数据集上做预训练（pre-training）。然后用训练好的模型作为初始化模型，在小数据集的任务上进行微调（fine-tune）训练。

最近，Kai-Ming博士通过对网络的结构研究发现，固定网络输入的原因来自于全连接层。全连接的运算，在本质上是一个线性的矩阵乘法运算。而矩阵运算要求参数矩阵的维度是固定的。不同尺度的图像经过卷积层后会得到不同尺度的特征图谱。为了接入全连接层，需要将不同尺度的特征图谱映射成固定维度的特征向量。为了解决这个问题，Kai-Ming博士在引文中提出了**空间金字塔池化**技术，将不同尺度的2维特征图谱通过多尺度池化方法，映射成固定维度特征向量。然而，这个方案还没在视频领域上得到验证。

1.经过多尺度训练的模型在识别性能上要比经过单尺度模式下训练的模型要高1.5个百分点。

2.在所有的实验环境设定中，STPP RGB模型在识别性能上都比C3D模型高。