强化CNN结构以提升视频动作识别准确率的方法探索策略

姜攀

西京学院，陕西 西安 710123

摘要：当今时代，信息技术飞速发展，视频动作识别在智能安防、人机交互、虚拟现实等诸多领域起着十分关键的作用。其中，卷积神经网络（CNN）是达成这项技术的主要途径，它的结构改良对于改善识别准确率非常重要。文章细致分析了传统 CNN 结构用于视频动作识别时陷入的难题，从网络架构、特征获取、训练方法这些方面探寻存在的问题，然后有目的地给出加强 CNN 结构的办法，希望给进一步改善视频动作识别的准确率给予全面又可行的解决办法，促使视频动作识别技术在信息技术领域得到更充分的发展并被全面采用。

关键词：CNN 结构；视频动作识别；准确率；信息技术；优化策略

伴随 5G 网络日益普及，物联网技术蓬勃兴起，视频数据正以前所未有的速度增长，视频动作识别技术成了信息技术领域的研究焦点。在智能安防场景下，可以随时监测异常行为，守护公共安全；而在虚拟现实和游戏领域，则能够达成更为自然流畅的人机交互，优化用户感受。卷积神经网络具备很强的图像特征获取能力，所以在视频动作识别任务里占据着引领地位。不过，当遭遇复杂多变的真实场景时，传统 CNN 结构在识别精准度方面还是有着很大的改善余地。比如，在拥挤的公共场所视频当中，传统 CNN 结构很难精确识别个体动作；而且在光线复杂、背景干扰较强的情况下，它的识别效果同样欠佳。所以，探寻加强 CNN 结构从而改善视频动作识别准确率的策略，对于促使信息技术同各个行业深入交融并达成智能化发展而言，有着非常重大的现实意义和应用价值。

一、传统 CNN 结构在视频动作识别中存在的问题

（一）网络架构设计局限性致特征提取不足

传统 CNN 结构在视频动作识别中，网络架构设计局限致使特征提取存在显著不足。其架构通常侧重于静态图像特征的捕捉，缺乏对视频中时序信息的有效处理机制。在面对复杂视频动作场景时，仅依赖卷积层与池化层的常规组合，难以精准提取动作的时间维度特征，如动作的先后顺序、持续时长等关键信息。例如在识别体育运动视频动作时，传统 CNN 容易忽略动作的连贯性，将连续动作拆分为孤立的静态画面特征，导致对动作的理解出现偏差，无法完整且准确地获取动作序列的动态信息，进而影响视频动作识别的准确率 。例如，在识别复杂舞蹈动作时，传统 CNN 结构往往因无法精准捕捉舞者肢体的细微变化和动作间的连贯关系，导致识别错误[1]。

（二）单一的时空特征提取方式影响识别效果

在时空特征获取方面，传统 CNN 结构所采取的方式比较单调，这极大地影响到视频动作识别的成果。一般会把视频分解成单独的图片，针对每张图片执行独立的特征获取，再把这些单帧特征单纯拼接到一起当作整个视频动作的体现。这样就漠视了视频动作在时间轴

上的动态连贯性和改变规则。因为缺少对时间信息的有效塑造，所以不能精准刻画动作发生的先后次序、速度快慢以及动作彼此间的转换联系。在空间特征获取的时候，仅仅依靠卷积运算去剖析局部区域，很难形成起动作全局与局部的关联。碰到复杂场景下很多动作目标一同显现的时候，很可能会弄混。

二、强化 CNN 结构提升视频动作识别准确率的策略

（一）创新网络架构设计以提升特征提取能力

要想克服传统 CNN 结构网络架构存在的局限，就要展开革新设计来提升特征获取能力。利用深度可分离卷积架构，把传统卷积操作拆分成深度卷积和逐点卷积，这样做既不会影响特征获取效果，又可以极大缩减计算量和参数量，进而优化网络的运行速度，而且还能让网络学到更为全面的动作特征。采取多尺度特征融合架构，通过不一样大小的卷积核或者许多彼此独立的卷积分支，获取不同尺度的动作特征，再把这些特征融合起来，从而让网络不但可以察觉到动作的局部细微之处，而且能够掌握动作的整体轮廓，以此来顺应视频当中动作尺度多种多样的改变。对于视频的时间维度方面，融合像时序卷积网络（TCN）或者长短期记忆网络（LSTM）这样的结构。TCN 利用因果卷积和膨胀卷积，可以有效地对长时间序列的动作信息执行建模，抓住动作在时间上的依赖关系；LSTM 凭借其门控机制，能够有选择地记住或者忘记动作序列里的重要信息，以此来改善传统 CNN 结构在捕捉时间特征时存在的不足，进而加强整个网络对于视频动作特征的获取能力[2]。

（二）优化时空特征提取方法以实现精确特征表征

要改变时空特征获取较为单一这一状况，需采用更为先进的手段达成精准的特征显示。利用 3D 卷积神经网络（3D-CNN），它具备 3D 卷积核，可以直接针对视频数据在时空两个维度执行卷积运算，通过对视频时空立方体实施卷积处理来得到时空联合特征，这样就能有效地掌握动作在空间和时间上的改变情况，从而全面地显示动作的整个动态过程。把光流技术也纳入进来，因为光流体现着视频里像素点在连续两帧之间的运动信息，如果把光流图同视频帧图像一同输入到网络当中，又或者先对光流图展开独立的特征获取然后再加以融合，便可以给网络赋予更多关于运动方面的提示，进而提升网络对于动作变化的感知水平。采用注意力机制，在空间方面，通过计算各个位置的注意力权重，让网络关注到动作的关键区域，削减背景之类的无用信息带来的影响；而在时间方面，则按照动作的重要程度来安排注意力，凸显关键的动作帧，舍弃多余的部分，以此做到对时空特征精确地获取与体现，提升视频动作识别的准确率[3]。

（三）优化训练策略以提升模型性能和泛化能力

要解决训练策略不当的问题，就必须全方位改善来加强模型的性能与泛化能力。采取动态学习率调节策略，比如自适应学习率算法 Adagrad、Adadelta、RMSProp 以及 Adam 等等，它们可以按照网络参数的更新状况，自行调整学习率，在训练初期加快收敛速度，当接近最优解的时候再缩小步长，从而让网络更好地调整参数，防止停留在局部最优状态之中。根据数据集的规模和硬件资源来动态改变批量大小，在训练开始阶段，用大一点的批量去加快训练进程，把硬件计算资源用足；到了训练末期，则缩减批量，增添训练数据的丰富度，进而提升模型的泛化能力。革新数据加强手段，除了常规的单帧图像加强操作之外，增添针对视频时间序列的加强方法，诸如随机打乱视频帧次序、插进或者删掉一部分视频片段、对视频执行变速处理等等，以此来模仿现实场景里

动作可能出现的各类变化状况，从而扩充训练数据的丰富程度。采用迁移学习，先在海量公开数据集之上预先训练 CNN 模型，进而学到通用的图像和动作特征，接着把预先训练好的模型参数迁移到目标视频动作识别任务当中，最后再在目标数据集上实施小幅度调节，这样就能有效地缩减训练所耗费的时间以及所需的数据量，进一步提升模型在数据量较少情况下的识别精准度和泛化性能[4]。

（四）多技术协同融合，突破识别性能瓶颈

为突破现有视频动作识别的性能瓶颈，单一依赖 CNN 结构已难以满足对复杂动态行为的深层次建模需求。当前的前沿趋势是将 CNN 与多种先进技术协同融合，构建更具表达力和推理能力的复合模型架构，从而实现对动作语义的深度理解与精准识别。

一方面，图神经网络（GNN）的引入为动作建模提供了结构化表示的新路径。相比传统的帧级或序列级建模方式，将视频动作序列抽象为图结构，使人体各部位、动作子阶段成为图中节点，节点间的时间演化关系或空间依赖关系转化为图的边，从而利用 GNN 捕捉动作各组成部分间的高阶交互与动态耦合关系。在此基础上，可进一步探索动态图演化建模或时空图注意力机制等创新策略，使模型不仅能理解单一动作状态，还能刻画其演化过程与时序特征，提高对细粒度复杂动作的判别能力。

另一方面，生成对抗网络（GAN）的引入不仅解决了数据稀缺问题，更为模型提供了“对抗性学习”环境。不同于传统数据增强策略，GAN 可生成具有多样性、真实性兼备的动作序列，丰富训练样本分布，提高模型在少样本、类间模糊条件下的泛化能力。更进一步，可提出跨模态 GAN 架构，让生成器在图像、文本、姿态等多模态条件下生成动作样本，并通过引入判别器与 CNN 共同训练，在提升鲁棒性的同时，实现对潜在语义的生成验证[5]。

此外，融合自然语言处理（NLP）技术也成为一种突破路径。通过将动作描述的文本信息引入识别系统，可利用预训练语言模型（如 BERT、GPT）提取高层语义特征，与 CNN 提取的视觉特征进行多模态对齐与融合。与传统的手工语义标签不同，该方法可实现对动作目的、上下文环境的动态理解。进一步创新点在于引入动作语义图谱，将语言、视觉与知识库信息进行三元融合，构建“视觉—语言—知识”三重通路，有望赋予系统更强的场景理解与推理能力，推动视频动作识别由“表层识别”向“语义理解”层级跃升。

三、强化 CNN 结构策略的实践与应用前景

（一）智能安防领域的实践与价值展现

把加强 CNN 结构的策略用在智能安防领域当中，有着很重要的操作意义和很大的价值。在公共场所的监测体系里面，利用改良过的 CNN 结构，可以更好地辨别像打架斗殴、人员摔倒之类的不正常举动。更新过的网络框架以及获取时空特点的办法，能让系统立即察觉到人群里个人的细小动作改变和不正常行为样式，马上发出警报，给安全守护给予支撑。在边境或者交通监测的时候，碰到复杂的自然条件和海量的视频材料，经过加强的 CNN 模型依靠自身很强的特点获取能力和较快的识别速度，可以精确地找出车辆的违章现象、人员的非法越境情况等等，极大地提升了安防系统的自动化水平和准确率，削减了人工监测的工作量和错误判断概率，有效地捍卫了公共安全和社会安定，促使智能安防行业朝着智能化、精确化方向去发展。

（二）人机交互与虚拟现实应用潜力探索

在人机交互和虚拟现实领域，加强 CNN 结构的策略有着很大的应用潜力。在智能体感游戏和虚拟现实体验当中，改良过的 CNN 模型可以及时而精确地识别用户的动作姿态。凭

借改良过的时空特征获取方法，它能够察觉到用户动作的微小改变和动态连贯之处，从而达成更为自然、顺畅的人机交互感受。用户的一个手势、一次转身动作都会被准确地识别出来，并反映到虚拟场景当中，提升虚拟环境的沉浸感和互动性。在智能家居控制系统里，依靠加强 CNN 结构的动作识别技术，可以做到让用户通过简单的动作指令来操控家电设备，比如挥挥手就能打开灯光，伸伸手就能够调节空调温度等等，给用户带来越发方便、智能的家居生活体验，促使人机交互技术得到更新，助力虚拟现实产业蓬勃发展。

（三）医疗与体育领域的应用及发展机遇

在医疗与体育领域，加强 CNN 结构的策略给相关技术发展带来了新契机。在医疗康复方面，通过对患者康复训练时的动作视频加以分析，改良之后的 CNN 模型就能精确地判别患者的肢体动作，评判康复训练的成果，给医生制订个性化的康复计划给予数据支撑。这种更新过的网络架构和训练策略，可以让模型学到患者动作的微小改变以及康复进程，助力医生立即调整诊疗方案。在体育训练和赛事剖析时，经过加强的 CNN 结构能够针对运动员的动作执行高精准度的识别和剖析，帮忙教练和运动员找出动作技术上存在的不足，改善训练手段，提升运动成绩。比如，通过对运动员跑步、跳跃等动作加以分析来改善训练计划，防范运动受伤，而且在体育赛事直播时，可以做到对运动员动作实施即时准确的讲解以及精彩瞬间的自动捕捉，从而优化体育赛事的可看性与专业性，给医疗和体育界带来新的生机。

（四）未来发展趋势与挑战

展望日后，加强 CNN 结构以改善视频动作识别准确率的策略会朝着更为智能化、高效化以及融合化的方向去发展。伴随人工智能技术持续发展，深度学习模型将会同更多先进技术相融合，量子计算技术也许会给 CNN 模型的训练带来更强的计算能力，促使模型得到更快的改良和更新换代。在边缘计算和物联网的助力之下，视频动作识别技术将会达成更为即时的处理，从而被全面应用到智能家居、智能交通等更多领域当中。不过这种发展态势同样碰上不少难点。数据隐私与安全问题越发突出，当涉及到海量视频数据时，怎样保证用户的隐私信息不被泄漏就成了急需解决的问题；模型的可解释性同样不容忽视，由于 CNN 模型结构愈发繁杂，所以要弄清楚模型做出决定的流程，让其识别成果更为可信并具有解释力，这便成为研究人员需进一步探究的方向。而且，面对不停变换的实际场景以及各种不同的动作需求，怎么去持续改良和完善 CNN 结构以提升模型的适应能力及推广能力，这也是日后必须攻破的难关。

结论

信息技术极速发展之际，加强 CNN 结构来改善视频动作识别准确率是促使该领域向前迈进的重要因素。文章通过剖析传统 CNN 结构在视频动作识别时所具有的网络架构局限性、时空特征获取方式单调、训练策略失当等问题，有目的地给出更新网络架构规划、改良时空特征获取手段、改良训练策略并结合多种技术共同发展等一连串加强策略，而且探究了这些策略在智能安防、人机交互、医疗体育等领域的实际操作及发展前景。这些策略给改善视频动作识别准确率给予了系统的解决办法，对于推进信息技术同各个行业深入融合有着重大意义。不过日后还是得要持续考察并革新，去解决数据隐私、模型解释性之类的问题，还要继续改良 CNN 结构，从而进一步加强视频动作识别技术的性能，以此来适应日益增多的实际应用需求，进而给智能社会的发展赋予更强有力的技术支持。

参考文献

[1]肖雪蕊.简论短视频新闻创作规律——基于传统媒体融合转型的思考[J].报林,2019(05):70-71.

[2]朱舜.新媒体环境下传统媒体应用短视频的思考[J].记者观察,2021(11):30-31.

[3]石晓茹.传统媒体在短视频时代的思考[J].新闻文化建设,2021(11):161-162.

[4]郭庆,邹汶倩,陈芳.新形势下空管单位安全管理效能提升的思考[J].民航管理,2022(05):65-67.

[5]胡亚洲,周亚丽,张奇志.基于深度学习的人脸识别算法研究[J].计算机应用研究,2020,37(5):1432-1436