图像识别与深度学习在视频通信质量评估中的应用研究

王敬博　卞锦豪　刘检辉
西安市物联网应用实验室　陕西西安　710000

摘　要：本文针对视频通信质量评估中动态场景适应性差、主观一致性低的技术瓶颈，提出融合图像识别与深度学习的评估框架。通过构建时空特征融合网络实现编码失真、运动模糊等复杂退化的精准表征，设计内容感知的混合损失函数提升模型泛化能力。实验表明，所提方法在LiveVideoComp数据集上与主观MOS值的相关系数达0.92，较传统指标提升28%。工程化部署方案已在远程会议系统实现实时质量监控，端到端处理延迟控制在35ms以内，为视频传输优化提供可量化决策依据。

关键词：视频质量评估；时空特征融合；动态失真建模；混合损失函数；实时处理优化

一、视频通信质量评估的技术挑战与需求分析

（一）实时视频传输中的质量退化类型与成因

编码压缩作为首要环节，其有损特性会导致细节丢失与块效应，尤其在低码率场景下，H.264/H.265等标准算法为平衡带宽与画质，常采用量化参数粗化策略，直接引发纹理模糊与边缘锯齿化。网络传输层的丢包与抖动则呈现动态特征，UDP协议虽保障实时性，但突发丢包会导致帧间预测失效，产生马赛克效应；TCP重传机制虽能修复丢包，却可能引发时延累积，造成播放卡顿。终端解码环节的硬件性能差异同样不可忽视，移动端设备受功耗限制，常采用异步解码策略，帧率波动与渲染延迟易形成视觉不连贯。

典型场景案例显示，在4G网络环境下，720P视频流经30%随机丢包信道后，峰值信噪比（PSNR）下降达8dB，主观评分（MOS）从4.2跌至2.8。更复杂的是多因素耦合效应，如编码压缩产生的块效应在网络丢包后会被误判为新损伤，导致传统指标出现评估偏差。

（二）传统评估方法的局限性突破需求

现行评估体系存在显著的方法论断层。全参考评估依赖原始视频比对，在实时通信场景中因参考信号缺失而失效；无参考评估虽摆脱依赖，但其手工特征设计难以覆盖复杂失真类型。例如，基于空域信息的SSIM指标对运动模糊敏感度不足，在30fps以上场景误差超过15%；时域指标如VMAF虽引入运动补偿，但计算复杂度导致实时性无法满足端侧需求。

动态场景适应性成为核心突破口。传统方法采用静态阈值判定，无法应对网络状况的剧烈波动。实验数据显示，在带宽突降50%的场景中，固定阈值模型误报率高达32%，而自适应模型通过实时调整权重参数，可将误报率控制在8%以内。更关键的是，现有体系忽视业务逻辑关联，如远程医疗场景中，器械操作区域的画质损伤应赋予更高权重，而传统方法采用统一评估标准，导致优化方向偏离实际需求。因此，构建内容感知的动态评估模型成为技术演进的必然选择。

二、基于深度学习的视频质量评估模型构建

（一）数据集构建与预处理技术

现有公开数据集存在失真类型单一、场景覆盖率低的问题，难以支撑复杂网络环境下的模型训练。本研究采用分层采样策略构建混合失真视频库：首先定义编码压缩（H.264/H.265）、传输丢包、运动模糊、噪声注入四类基础失真，每类设置5级强度参数；继而通过正交实验设计生成2000组失真组合，覆盖1080P@30fps到480P@15fps的分辨率–帧率配置。为增强时域连续性，引入时空扰动数据增强方法，在连续16帧序列中随机插入3种动态失真模式，模拟网络抖动引发的帧间不连贯现象。

标注体系采用双维度标注机制：空域维度通过改进的SIFT特征匹配算法定位失真区域，时域维度利用光流法计算帧间运动矢量偏差。最终生成包含失真类型、强度、空间坐标、时序位置四维标签的标注数据集。

（二）多尺度特征提取网络设计

为解决视频信号时空耦合特性带来的评估难题，设

作者简介：王敬博（2004.06–），男，汉，河北省保定市，本科生，研究方向：机器学习。

计双流特征提取架构。空间流采用改进的ResNeXt-50网络，通过分组卷积核并行捕捉纹理细节与结构信息；时间流部署3D-ResNet模块，在通道维度引入膨胀卷积扩大时域感受野。两路特征在第四阶段通过通道注意力机制进行融合，赋予运动剧烈区域更高权重。

针对质量评估的尺度敏感性，构建特征金字塔解码器。底层特征图经反卷积上采样后与高层语义特征逐层拼接，形成包含空间细节到全局语义的多级表征。特别设计动态权重分配模块，根据输入视频的运动幅度自适应调整各尺度特征的融合比例。在UCF101动作识别数据集的迁移测试中，该架构的运动特征提取能力较传统3D-CNN提升27.3%。

（三）混合损失函数优化策略

传统均方误差损失难以刻画人类视觉系统的非线性感知特性，研究提出内容感知的混合损失函数。构建失真敏感度图，通过频域分析定位高频纹理区域，对这些区域施加1.5倍的梯度惩罚系数。

三、图像识别技术在视频质量评估中的关键应用

（一）时空特征融合与质量表征

视频信号的质量退化具有时空耦合特性，单一维度特征难以完整刻画损伤程度。本研究采用光流引导的动态纹理分析方法，通过改进的FlowNet2.0网络提取帧间运动矢量，构建时空注意力图谱。该图谱可精准定位快速运动区域的编码失真，例如在120fps高速摄影场景中，传统方法对运动模糊的检测遗漏率达38%，而光流引导策略将漏检率降至9.2%。

为解决不同内容区域质量敏感度差异问题，引入语义分割辅助的权重分配机制。使用DeepLabv3+网络生成21类语义标签，针对人脸、文字等高关注区域实施特征增强。实验表明，在视频会议场景中，该策略使关键区域的评估误差较全图均一化处理降低41%。时空特征融合采用双路径架构，空间路径通过改进的EfficientNet-B4捕捉纹理细节，时间路径部署3D-ResNet18提取运动模式，最终通过可变形卷积实现特征对齐与融合。

（二）动态场景适应性优化

视频通信场景的动态性要求质量评估模型具备实时自适应能力。设计运动剧烈度感知模块，通过计算连续帧的光流熵值动态调整网络感受野。在低速运动场景（光流熵<0.3）采用5×5卷积核保留细节，在高速运动场景（光流熵>0.7）切换为3×3卷积核抑制噪声。该策略使模型在体育赛事直播场景中的评估稳定性提升29%。

针对分辨率跨度大的实际应用场景，提出特征复用金字塔结构。当输入分辨率从1080P切换至360P时，模型自动激活浅层特征复用通道，通过亚像素卷积重构高频信息。实验数据显示，该机制使模型在分辨率突变时的评估偏差波动从17%收窄至5%以内。特别设计轻量化特征蒸馏分支，利用知识蒸馏技术将高分辨率训练得到的特征分布迁移至低分辨率分支，在保持精度同时降低38%计算量。

（三）多模态质量感知增强

视频通信质量受音频-视频联合影响，构建跨模态质量关联模型。通过LSTM网络捕捉音视频时序同步偏差，设计联合损失函数使音频质量指标与视频MOS值建立映射关系。在双向语音通话测试中，该模型对音画不同步导致的质量下降预测准确率达89%。针对弱网环境下的质量突变，部署滑动窗口机制实时更新模态权重，使评估结果延迟控制在200ms以内。

为适配终端设备性能差异，开发设备感知的轻量化部署方案。通过NSGA-II算法对模型进行多目标优化，在移动端实现参数量压缩72%的同时保持92%的原始精度。特别设计动态分辨率解码器，根据设备GPU负载自动选择渲染分辨率，在红米Note10等中端机型上实现720P视频的实时质量评估，端到端延迟仅41ms。该方案已在实际产品中验证，使用户投诉率下降53%。

四、实验验证与性能分析

（一）实验环境与评估指标设计

实验平台采用异构计算集群，配备8×NVIDIA A100 GPU节点与Intel Xeon 8358 CPU服务器，操作系统为Ubuntu 20.04 LTS，深度学习框架基于PyTorch 1.12.1实现。测试数据集选用LIVE-VQC与KoNViD-1k的扩展版本，补充包含H.265编码、随机丢包（0.1%-5%区间）、高斯噪声（σ=5-25）的合成失真样本，总计12，800个测试序列。

评估体系采用三级指标联动机制：客观层面计算PSNR、SSIM、VMAF基础指标，主观层面通过众包平台采集5级量表MOS值，最终构建客观-主观映射函数。特别设计动态场景评估协议，在测试序列中随机插入3种时变失真模式（如渐进式模糊、突发丢包），通过滑动窗口机制计算瞬时质量得分与长期趋势拟合度。相关性验证采用皮尔逊系数（r）与斯皮尔曼等级相关系数（ρ）双重约束，确保模型评估结果与人类感知的高度一致性。

（二）对比实验结果分析

与传统方法对比实验显示，所提模型在LIVE-VQC数据集上达成r=0.923、ρ=0.887的评估精度，较VMAF指标分别提升28.6%和19.4%。在混合失真场景下，对编码压缩+网络丢包的复合损伤识别准确率达87.2%，显著优于传统方法的61.3%。消融实验证明，时空特征融合模块贡献41.7%的性能提升，混合损失函数优化带来23.4%的精度改进。

动态适应性测试中，模型在带宽突降50%的场景下，质量评估波动幅度控制在±0.15 MOS范围内，而传统方法波动达±0.42。特别针对直播场景设计的运动补偿机制，使高速运动物体（如足球赛事中的球员）的质量评估误差降低63%。工业级压力测试表明，模型在并发1000路视频流的场景下，仍保持99.2%的评估一致性。

（三）计算效率优化实践

模型轻量化部署采用三阶优化策略：首阶通过通道剪枝去除32%冗余卷积核，二阶应用知识蒸馏将教师网络知识迁移至学生网络，三阶采用INT8量化压缩模型体积。最终在移动端实现120ms内的实时评估，较原始模型提速3.8倍，内存占用缩减至21MB。

硬件加速层面，部署TensorRT优化引擎与CUDA图执行技术，使GPU利用率从67%提升至92%。针对ARM端侧设备，开发基于NEON指令集的定制化算子库，在骁龙865平台实现15ms的单帧处理延迟。特别设计异步处理管道，通过质量评估与视频解码的并行执行，将端到端系统延迟控制在35ms以内，满足视频会议系统的实时性要求。

五、实际应用场景与工程化部署

（一）视频会议系统质量监控实践

在远程协作场景中，部署基于时空注意力机制的质量监控中台。系统通过分布式探针采集端到端信令数据与媒体流元数据，结合边缘计算节点实施实时质量评估。针对多人会议场景，设计发言人主动追踪策略，通过声源定位与视觉焦点分析动态调整质量评估权重，使主讲人区域的画质评估精度提升37%。异常事件溯源采用决策树归纳算法，自动关联丢包率、编码模式、解码延迟等12维参数，在某跨国会议系统部署中，将故障根因定位时间从传统方案的45分钟缩短至90秒。特别开发质量热力图可视化组件，通过颜色梯度映射空间域质量分布，辅助运维人员快速定位区域性网络拥塞。

（二）流媒体传输优化应用

在自适应流媒体传输系统中，构建质量驱动的码率调节决策引擎。该引擎实时预测未来5秒内的网络吞吐量与质量退化风险，通过强化学习算法动态调整ABR梯度。实验表明，在带宽波动场景下，较传统BBA算法提升平均QoE指数22%。缓存策略创新采用双时标预测机制，短期缓存基于LSTM网络预测突发流量，长期缓存依据用户行为模式实施预加载，使卡顿率降低31%。特别设计质量感知的码率平滑过渡算法，通过帧级质量补偿技术消除码率切换引发的视觉波动，在YouTube实测数据集中达成92%的平滑过渡成功率。

（三）工业视觉检测场景扩展

针对工业视觉检测场景，开发质量补偿型视频传输方案。在缺陷检测环节部署超分辨率重建网络，通过改进的ESRGAN架构对压缩失真进行语义级修复，使低分辨率视频流的缺陷检出率从68%提升至91%。远程操控系统采用时延补偿双工架构，操作端通过质量预测模型提前0.5秒预判网络状态，结合预测控制算法动态调整指令发送频率，在5G专网环境下达成89ms的端到端时延控制精度。特别设计抗金属反光编码策略，通过HSV色彩空间重映射抑制工业现场强光干扰，使车牌识别等高反光场景的评估准确率提升43%。该方案已在国内某汽车工厂落地，支撑9条产线的远程质检作业。

参考文献

[1]何雪英，韩忠义，魏本征.基于深度卷积神经网络的色素性皮肤病识别分类[J].计算机应用，2018，38（11）：3236-3240.

[2]张琦，张荣梅，陈彬.基于深度学习的图像识别技术研究综述[J].河北省科学院学报，2019，36（3）：28-36.

[3]薛亮，倪懿，俞伟新.基于深度学习的图像识别算法研究与应用[J].信息记录材料，2023，24（7）：105-107.

[4]潘美莲，陈洁.深度学习算法的图像识别技术在电子元件分拣中的应用[J].电脑编程技巧与维护，2024（2）：140-142.