

# 端到端优化的视频编解码技术研究

吴胤鹏

(南京熊猫通信科技有限公司, 南京 210000)

**摘要:** 文章深入探讨了端到端优化的视频编解码技术,通过整合视频采集、编码、传输、解码等环节,该技术可以实现对视频数据的高效压缩与高质量传输。随着深度学习技术的发展,基于深度学习的端到端视频编解码方法逐渐兴起,依托神经网络模型强大的学习能力,实现了对传统视频编解码算法的超越。文章首先介绍了端到端视频编解码技术的背景与意义,随后详细阐述了其关键技术,包括基于深度学习的视频压缩算法、多尺度特征的运动补偿、多参考帧的辅助预测帧生成等。通过对比实验,文章展示了端到端优化视频编解码技术在提高压缩效率、减少传输延迟、提升视频质量等方面的显著优势,同时总结了当前研究的成果,并对未来研究方向进行了展望。

**关键词:** 端到端优化;视频编解码;深度学习;多尺度特征;多参考帧;压缩效率;传输延迟;视频质量  
**中图分类号:** TN919 **文献标识码:** A

## Research on end-to-end optimized video encoding and decoding technology

WU Yinpeng

(Nanjing Panda Communications Technology Co., Ltd., Nanjing 210000, China)

**Abstract:** This paper deeply discusses the end-to-end optimized video codec technology, which can realize efficient compression and high-quality transmission of video data by integrating video capture, encoding, transmission and decoding. With the development of deep learning technology, end-to-end video codec methods based on deep learning have gradually emerged. Relying on the powerful learning ability of neural network models, they have surpassed traditional video codec algorithms. This paper first introduces the background and significance of end-to-end video codec technology, and then elaborates its key technologies, including video compression algorithm based on deep learning, motion compensation of multi-scale features, and auxiliary prediction frame generation of multi-reference frames. Through comparative experiments, this paper demonstrates the significant advantages of end-to-end optimized video codec technology in improving compression efficiency, reducing transmission delay, and improving video quality. At the same time, the current research results are summarized, and the future research direction is prospected.

**Key words:** end-to-end optimization, video codec, deep learning, multi-scale feature, multi-reference frame, compression efficiency, transmission delay, video quality

### 1 引言

在信息技术高速发展的今天,视频数据已成为日常生活必不可少的信息载体,但海量的视频数据却难以得到存储与传输。传统视频编解码技术(如H.264/AVC和H.265/HEVC)在一定程度上缓解了这个矛盾,但其在压缩效率、传输延迟及视频质量仍有待改进。近年来,深度学习的崛起给视频编解码领域提供了一种全新的解决思路。基于深度学习网络,通过在视频压缩和传输各环节应用端到端的视频编解码优化技术,可以对视频数据进行高效压缩与高质量传输,不仅提高了压缩效率,还

显著减少了传输延迟,为视频通信、在线直播和远程医疗提供了强有力的支撑。

### 2 端到端优化的视频编解码技术特点

随着互联网及多媒体技术的快速发展,信息呈爆炸式增长。在实时互动场景中,视频画质是影响用户体验的关键因素,从流畅、标清、高清到超高清技术的递进演进给视频制作带来新的挑战。视频压缩技术主要通过减少信号中存在的冗余数据和视觉冗余来实现。近年来,深度学习技术在计算机视觉领域取得了显著成果,而传统混合视频压缩算法存在分立、无法联合优化

等缺陷。因此,可以利用深度学习技术的强大学习和非线性能力将其用于视频压缩算法中。

端到端的视频编解码技术是当代视频处理领域中的革新性成就,具有鲜明而深刻的特征,对视频通信、多媒体存储和传输等领域产生了重大影响。该技术在根本上突破了视频编、解码的传统框架边界,对视频处理流程的每个环节进行了深度整合,使整个处理流程从视频捕获直至最后展示都得到了优化,大幅提高了视频数据处理效率和质量。

端到端视频编解码技术最为突出的特征是它具有高度的集成性和协同性。传统视频编解码体系通常要对视频数据进行采集、预处理、编码、传输、解码和后处理,其间常常会出现信息割裂和冗余处理。而端到端技术通过搭建统一的优化框架,使各处理环节密切耦合,从而实现数据的无缝流动和共享<sup>[1]</sup>。这一集成性不但减少了各阶段间数据转换的损耗,而且通过协同优化策略使整个视频处理流程变得更有效和紧凑。

端到端的视频编解码技术充分利用了深度学习等先进技术的强大功能。卷积神经网络(CNN)和循环神经网络(RNN)等深度学习模型在图像识别、特征抽取及时间序列分析等方面发挥着至关重要的作用。在视频编解码过程中,子带编码主要有2种方式:(1)将图像数据变换到频域后按频域分带,然后用不同的量化器进行量化,从而达到最优组合;(2)分步渐进编码,在初始时对某一频带的信号进行解码,然后逐渐扩展到所有频带,随着解码数据的增加,解码图像将逐渐清晰。针对远程图像模糊查询与检索问题,相较于传统规则编解码算法,子带编码在压缩效率和编码质量上都有了更明显的改进。

### 3 端到端的优化视频编解码技术的实现

#### 3.1 深度学习视频压缩算法

在端对端的优化视频编解码技术中,以深度学习为核心的视频压缩算法逐渐占据主流地位。卷积神经网络和循环神经网络等深度学习模型为视频压缩技术带来了创新性突破。H.264与H.265等传统视频压缩技术是基于手动设计特性和编码的方式,但它们对于复杂视频内容的处理通常具有一定局限性。而以深度学习为基础的视频压缩技术可以自动识别视频的特性和模式,从而达到更高的压缩效率<sup>[2]</sup>。

以基于卷积神经网络的视频压缩技术为例,该技术首先从输入的视频帧中提取特征,能够在分辨率达到1 920×1 080的视频帧中抽取出数千个特征图像,其中含有边缘、纹理、颜色等等级信息。编码阶段模型采用循环神经网络对特征图量化编码,该模型经过训练后学会了码率进行最优分配的方法,在比特率受限的情况下能尽量多保留一些重要信息。实验数据显示,使用该

基于深度学习的压缩算法对一段时长10 min、帧率30 fps的高清视频进行压缩,在相同的码率下(如5 Mbps),相较于H.265,其峰值信噪比(PSNR)平均提高了约2 dB,这意味着视频质量有了显著的提升。此外,该深度学习模型可以对各种视频内容进行自适应处理。例如,针对动态场景较丰富的体育赛事视频,该模型能够更加关注运动物体准确编码;针对静态场景较丰富的纪录片视频,该模型能够对背景区域进行重压缩,以达到更加有效的码率分配。

#### 3.2 多尺度特征进行运动补偿

多尺度特征运动补偿在端到端的优化视频编解码技术中具有重要意义,视频序列中对象的移动是多样复杂的,传统运动补偿方法一般根据单一尺度上的特性进行预测,难以精确捕捉到不同尺度上的运动信息,而多尺度特征运动补偿方法可以有效解决该问题<sup>[3]</sup>。以分辨率1 280×720、帧率25 fps的视频为例,首先需要提取各帧多尺度特征,采用3个尺度:原始尺度、1/2尺度和1/4尺度,通过特征提取算法(如方向梯度直方图(HOG)或卷积神经网络)获取每个尺度的相应特征。同时,分别计算不同规模下的运动。在原始尺度下,小幅度的运动可准确被估算;在1/2尺度下,能够较好捕捉中等幅度的运动;在1/4尺度下,能够较好处理大幅度的运动。实验结果显示,在处理具有多种运动幅度的视频序列时,相较于传统单一尺度的运动补偿方法,多尺度特征的运动补偿技术可以将平均编码的效率提升20%。

在运动补偿时,需要对不同规模的运动估计结果进行融合与优化。假定一个含有快速运动物体与缓慢运动背景的视频场景,利用多尺度运动补偿能够给快速运动物体配置更多的编码资源,并更加有效地压缩缓慢运动的背景。根据数据显示,该优化方式可以在确保视频的高质量的同时平均降低15%的码率。此外,对多尺度特征进行运动补偿可以增强视频编码的鲁棒性,若视频中存在噪声、遮挡及复杂运动模式,则可以对运动轨迹进行较为精确的预测,以降低预测误差。研究发现,在苛刻的视频录制环境中,相较于传统技术,利用多尺度特征进行运动补偿可以显著提高视频解码的质量,使峰值信噪比(PSNR)平均增加超过1.5 dB。在科技日益发展的今天,多尺度特征运动补偿会进一步和深度学习等先进技术相结合,从而给视频编解码工作带来更加出色的表现。

#### 3.3 多参考帧辅助预测帧

多参考帧的辅助预测帧在端对端的优化视频编解码技术发挥着关键作用。传统视频编码通常仅以相邻数帧为参考帧进行预测编码。但在应对复杂视频内容及快速运动场景时,其常存在预测不准的情况,从而导致编码效率下降及视频质量差。

(下转第190页)