

随着 5G 技术的逐步普及，一个拥有庞大用户数量的高速网络时代已经到来，网络视频占据了互联网中大部分的流量。同时，随着高清晰度摄像设备的广泛使用和观看终端屏幕分辨率的提升，网络流媒体视频数据的信息密度和复杂程度不断增加，其数据量呈现出爆炸式增长。这无疑给视频处理和编码压缩带来了巨大挑战。视频编码前处理技术具备提升视频编码传输系统率失真性能的巨大潜力，具有很高的实际应用和学术研究价值。近年来，随着深度学习技术的快速发展，涌现出了一批数据驱动的视频编码优化方法。然而，对基于深度学习的端到端率失真优化导向的视频编码前处理方法的研究还极为欠缺。一方面，在端到端深度学习训练过程中无法准确描述标准编码器的率失真特性，使得端到端系统优化难以进行；另一方面，不同应用场景之间的视频数据分布存在显著差异，使得编码前处理方法在不同场景下的适配性面临严峻挑战。

针对上述问题，本文构建了高效的微编码模型，并基于该模型设计了端到端的基于深度学习的视频前处理及编码训练测试框架，为后续研究提供了基础。之后依次研究了在面向人眼感知和机器视觉分析任务中的率失真导向的视频编码前处理方法，在实验中均取得了领先的性能。本文的创新贡献如下：

第一，提出了一种基于微编码模型的端到端视频前处理与编码框架，使具备端到端率失真性能的深度前处理模型的训练成为可能。在设计微编码模型时，采用了自下而上的设计思路，并将传统编码标准中的通用核心操作抽象出来，以高并行度的微方式进行设计，以使其能够与深度学习优化框架高度融合，且具有高度的可扩展性，为后续的研究提供了基础。

第二，提出了一种面向专业生成内容（PGC）的视频编码前处理方法，在端到端视频前处理与编码框架中引入具备时域和空域处理能力的前处理模型，以及系统率失真优化函数，训练了具有系统率失真性能的深度前处理器。其中，前处理模型考虑了 PGC 视频的特点及编码特性，在时间维度使用了时域补偿的帧间融合方式，在空间维度使用了扩张卷积的残差连接网络进行构建。实验结果显示，在大规模应用的真实编码器及 PGC 视频数据上，训练后的前处理模型能够有效提升整体系统的率失真性能，在多个主流评价指标上表现突出，在高清编码数据集上的主观和客观指标均有超过 8% 的率失真性能提升。

第三，提出了一种面向用户生成内容（UGC）的视频编码前处理方法。该方法通过视频预分析可以针对 UGC 视频的复杂率失真特性自适应地选取最合适的前处理训练与测试策略，以达到整体系统的最优性能。实验结果表明，该方法能够大幅提升在复杂数据分布情况下的视频编码系统性能，特别是对于一些在基线方法上产生负向收益的视频类型，使用所提出的方法正向收益率显著提升，平均正向收益占比提升达 10%。

第四，提出了一种面向视频分析任务的视频编码前处理方法，使用双分支的前处理网络和混合指标的损失函数进行训练。实验表明在保证下游机器视觉任务精度和不影响人眼感知视觉损失度的情况下，在动作识别及目标跟踪的典型任务上，所提出的方法可以大幅节省视频传输所需码率达 15% 以上。

综上所述，本文通过提出端到端的视频前处理及编码框架，解决了优化深度前处理器的前置问题。在此基础上，构建了编码率失真导向的深度前处理器训练测试范式，并证明了其性能的优越性和广阔的实际应用前景。进一步地，本文针对人眼感知，对专业生成视频和用户生成视频的场景适配问题提出了完善的编码前处理方法；同时，对下游机器视觉任务中的编码前处理方法进行了研究，证明了所提出前处理框架的有效性和可拓展性。所提出的方法大幅提升了视频通信系统的效率，为后续研究和应用奠定了基础。