

基于场景感知的端边云多数据流协同优化传输方法

国 琪^{1),3)} 杨哲铭¹⁾ 赵 昌^{1),3)} 纪 雯^{1),2)}

¹⁾(处理器芯片全国重点实验室 中国科学院计算技术研究所 北京 100190)

²⁾(中国科学院工业人工智能研究所 南京 211135)

³⁾(中国科学院大学 北京 100049)

摘 要 随着智慧城市监控和智能交通等场景下视频数据的爆发式增长,大量帧级冗余信息给传输系统带来了巨大压力。传统集中式云计算架构在应对多路并发视频上传时,受限于带宽不足、带宽波动以及核心网拥塞,难以保证高效传输。现有自适应码率与压缩算法虽能降低码率,但仍依赖整帧传输,难以在兼顾低时延的同时保障关键区域质量。同时,多数研究聚焦于单视频流优化,忽视了人机视觉差异化需求及结构化数据在低带宽下的潜力。针对上述问题,本文提出一种基于场景感知的端边云多数据流协同优化传输方法。该方法首先构建了一种新颖的端边云多流协同传输架构。随后,将多流自适应传输建模为带约束的非凸组合优化问题,并引入场景感知状态划分模型,通过感知可用带宽及其波动,将复杂连续的网络环境映射为常规、干扰和恶劣三类典型场景,从而生成候选流集合与压缩率范围,有效降低优化维度。在此基础上,提出权重驱动的多流调度与压缩率联合优化方法,通过计算各流偏好得分与调度权重,自适应激活最优流类型,并在候选集合内动态配置传输策略与压缩率。同时结合拉格朗日对偶方法,实现高效求解,以最小化边云传输时延并保障视频质量。实验结果表明,本文所提方法在高压缩率下能够有效保持视频焦点区域结构相似性指标(Structural Similarity Index, SSIM)的图像质量,在直播场景中焦点区域平均 SSIM 达到 0.882,显著优于其他基线方法。在 4G/LTE 真实带宽数据集上,本文方法较各类基线方法传输时延降低 44.0%~85.4%;在 5G 带宽数据集构建的直播场景下,可以将传输时延降低 31.82%~63.57%。综上,本方法能够在复杂多变的网络条件下兼顾视频质量与低时延,确保端边云系统能够高效地处理大规模视频任务。

关键词 视频传输;场景感知;多数据流;端边云系统;多流传输

中图法分类号 TP37

DOI号 10.11897/SP.J.1016.2026.02252

A Scene-Aware Collaborative Optimization Method for Multi-Stream Transmission in End-Edge-Cloud Systems

GUO Qi^{1),3)} YANG Zhe-Ming¹⁾ ZHAO Chang^{1),3)} JI Wen^{1),2)}

¹⁾(State Key Lab of Processors, Institute of Computing Technology, CAS, Beijing 100190)

²⁾(Institute of AI for Industries, Chinese Academy of Sciences, Nanjing 211135)

³⁾(University of Chinese Academy of Sciences, Beijing 100049)

Abstract The explosive growth of video data in smart city surveillance and intelligent transportation has placed heavy pressure on transmission systems due to massive frame-level redundancy. Traditional centralized cloud architectures struggle to support concurrent video uploads because of limited bandwidth, fluctuations, and core network congestion. Existing adaptive bitrate and com-

收稿日期:2025-09-30;在线发布日期:2026-05-13。本课题得到国家重点研发计划项目(No. 2023YFB4502805)、中国博士后科学基金项目(No. GZC20251091)以及江苏省工业人工智能重点实验室资助。国 琪,博士研究生,中国计算机学会(CCF)学生会会员,主要研究领域为多媒体计算、端边云系统、视频传输。E-mail:guoqi23p@ict.ac.cn。杨哲铭(通信作者),博士,助理研究员,中国计算机学会(CCF)会员,主要研究领域为多媒体系统、分布式推理、边缘智能。E-mail:yangzheming@ict.ac.cn。赵 昌,硕士研究生,主要研究领域为大语言模型推理加速、模型压缩。纪 雯(通信作者),博士,研究员,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为智能芯片设计、多媒体系统和人工智能系统,包括工业人工智能、高性能视觉处理器、AI 处理器、多媒体端边云计算系统、视觉编码与传输、低碳计算。E-mail:jiwen@ict.ac.cn。

pression methods can reduce bitrate but still rely on full-frame transmission, making it difficult to ensure both low latency and critical region quality. Furthermore, most studies focus on single-stream optimization, overlooking the differentiated needs of human-machine vision and the potential of structured data under constrained bandwidth. To address these issues, this paper proposes a scene-aware collaborative optimization method for multi-stream transmission in End-Edge-Cloud systems. First, a novel End-Edge-Cloud multi-stream collaborative transmission architecture is constructed. The problem is then modeled as a constrained non-convex combinatorial optimization problem, where a scene-aware state partitioning model maps dynamic network conditions into normal, disturbed, and harsh scenarios. This reduces optimization complexity by generating candidate stream sets and compression ranges. A weight-driven joint scheduling and compression strategy is then introduced, which adaptively activates optimal stream types and dynamically configures transmission policies. The Lagrangian dual method is applied for efficient solving, minimizing latency while maintaining video quality. Experimental results show that the proposed method can effectively preserve the Structural Similarity Index (SSIM) quality of video focus regions under high compression ratios. In the live-streaming scenario, the average SSIM of focus regions reaches 0.882, significantly outperforming other baseline methods. On a real 4G/LTE bandwidth dataset, the proposed method reduces transmission latency by 44.0%–85.4% compared with various baselines. In the live-streaming scenario constructed using a 5G bandwidth dataset, it further reduces transmission latency by 31.82%–63.57%. Overall, the proposed method achieves a favorable balance between video quality preservation and low-latency transmission under complex and dynamic network conditions, providing effective support for edge-cloud systems to efficiently process large-scale video tasks.

Keywords video transmission; bandwidth awareness; multi-stream data; end-edge-cloud systems; multi-stream transmission

1 引 言

随着物联网(Internet of Things, IoT)与人工智能(Artificial Intelligence, AI)技术的快速发展,全球数据规模进入指数级增长阶段。据 IDC 预测,到 2028 年,全球每秒数据生成量将达到 12.5PB^[1]。在这一趋势中,视频数据作为核心数据载体,其规模增长尤为突出,并在全球数据流量中占据主导地位。GIV 报告显示,到 2025 年,全球人均通信数据量将增长 10 倍,其中视频数据占比高达 89%^[2]。大规模视频数据的持续生成与高占比,正成为未来网络与计算基础设施的关键压力源,对核心网(Core Network, CN)的传输能力提出了前所未有的挑战。当前主流数据处理模式主要依赖集中式云计算架构,将大量原始视频数据上传至云端统一分析与处理。IDC 预测 2028 年全球将有超过 60% 的数据最终都需依托云端进行存储与处理^[1]。这不仅加剧了

网络拥塞风险,也对带宽资源、云端计算能力和任务实时响应方面提出了严峻挑战。具体而言,当前视频传输体系主要面临三大核心挑战:

第一,集中式端云传输导致时延累积与关键内容受阻。主流模式将视频流压缩后统一上传云端处理,数据需跨接入网/骨干网长距离传输,往返时延(Round-Trip Time, RTT)随端云距离增大而上升,并叠加云端排队与资源调度开销,使多路并发或带宽受限时段云时延显著放大。这种无差别的传输策略不仅使大量低价值、重复信息占用网络与云端资源,进而挤压关键帧、事件片段与语义信息的传输时机,难以满足实时监控与决策的时延要求。

第二,海量视频的集中传输导致带宽资源紧张。在大规模并发上传条件下,上行链路容量往往不足以承载全量视频流,以单路约 4Mbps 计算,十万路并发超过 400Gbps,远超多数区域网络与核心链路承载能力^[3]。尽管压缩与自适应码率(Adaptive Bitrate, ABR)^[4]可降低平均码率,但主流方案仍以整

帧传输为主,难以剔除冗余信息或针对不同区域进行差异化编码,从而难以根本缓解由带宽资源紧张引发的高时延问题。

第三,动态网络带宽波动引发传输不确定性。受信道干扰、用户竞争与网络切换影响,上传链路常出现突发带宽下降或短时飙升,使既有调度与重传机制难以及时匹配;带宽骤降易触发缓冲耗尽与丢帧,带宽回升又受切换滞后影响难以快速恢复,最终表现为卡顿、时延升高甚至中断。不同区域基础设施差异进一步加剧该不稳定性,持续削弱传输连续性与可达性。

针对上述三大挑战带来的高时延问题,本文提出一种基于场景感知的多数据流协同优化传输方法:首先,提出端边云多流协同传输架构,在边缘侧生成特征压缩流与结构化语义流,以减少上行冗余;其次,将数据流类型选择与压缩率配置统一建模为带质量约束的组合优化问题,并提出场景感知状态划分模型,通过感知当前可用带宽并对其均值与波动幅度进行建模,将复杂连续的网络条件离散为有限典型场景,从而有效缩减搜索空间并形成候选约束集;随后,在场景判决结果上,提出基于偏好打分与调度权重的自适应多数据流协同优化传输方法,并在可行压缩率集合内求取近似最优解,实现质量约束下的时延最小化。本文主要贡献总结如下:

(1)针对传统集中式云架构在视频传输中结构性低效与高时延等问题,提出一种新颖的端边云多数据流协同传输架构,该架构在终端侧执行轻量化采集与预处理,在边缘侧生成多类型上行数据流并完成场景感知划分与多流协同调度优化,在云端执行高层语义分析与全局决策。

(2)针对视频传输中因带宽受限与大幅波动引发的传输瓶颈问题,提出一种基于场景感知的多数据流协同优化传输方法,将原始非凸组合优化问题分解为场景状态划分与多流调度-压缩率联合优化两阶段,通过权重驱动机制实现多数据流的自适应选择与传输控制,从而在保证视觉质量约束的同时最小化传输时延。

(3)本文在 Cityscapes^[5]数据集、4G/LTE^[6]数据集以及 5G Traffic Datasets^[7]中 Zoom1 数据集上进行了多维度实验验证。结果表明,相较于 H.265/HEVC^[8]和 rIROI^[9]等方法,本文提出的特征压缩方法在相同或更高压缩下能够更好地保持视觉焦点区域质量:在智能交通场景中,焦点区域 SSIM 始终稳定在 0.6 以上;在更复杂的直播场景中,焦点

区域平均 SSIM 达到 0.882。传输时延方面,在 4G/LTE 真实带宽数据集上,所提方法较 mDASH 等主流方案降低 44.0%至 85.4%;在基于 5G 带宽数据集构建的直播场景中,较其他方法降低 31.82%至 63.57%,显著提升了传输效率。

本文第 2 节介绍集中式云计算架构与边缘计算架构下的视频传输相关研究工作;第 3 节阐述本文的研究动机分析;第 4 节介绍端边云多流协同传输系统架构与多流生成模型;第 5 节系统提出并实现一种基于场景感知的多数据流协同优化传输方法,包括问题建模、场景状态划分、多流调度与压缩率联合优化建模以及算法设计;第 6 节对所提方法在视觉质量与传输性能方面进行全面实验评估、对比分析与实验验证;第 7 节对全文进行总结与展望。

2 相关工作

2.1 基于云计算的视频优化传输方法

云计算技术的发展极大地促进了网络资源的智能调度与弹性扩展,为视频内容的高效传输提供了坚实的基础。早期研究主要集中在静态云集群上,利用固定配置的虚拟机进行大规模视频编码与处理,例如,FastTTPS 在并行计算资源上进行转码时间预测与调度以支持自适应流媒体处理^[10]。随着云计算向弹性架构演进,研究重点转向自动伸缩与资源编排,通过在线监测负载与服务质量(Quality of Service, QoS)指标并进行关键指标选择,动态调整资源规模,实现服务连续性与资源效率的协同优化^[11]。此外,为提升多终端、多网络环境下的传输效率,云端编码/转码与内容交付机制不断演进,并在系统层面持续关注成本、性能权衡与服务质量保障^[12]。同时,针对超高清视频传输中的带宽瓶颈问题, Pearl 等云端深度压缩框架通过联合压缩分辨率与色彩信息,在保持视觉质量的同时降低数据量与传输时延^[13]。尽管当前以云计算为核心的视频传输系统在调度、转码与压缩等方面取得了显著进展,但海量多媒体数据的高需求仍可能引发云端系统拥塞,并对云端算力与数据中心资源带来压力^[14]。

2.2 基于边缘计算的视频优化传输方法

为缓解传统云计算架构面临的高时延与传输瓶颈,边缘计算作为一种更贴近数据源的计算范式被提出^[15],例如,移动边缘计算在接近视频源与用户侧部署计算、缓存与转码,使视频内容在边缘侧完成缓存、预处理与按需转码,从而降低传输时延并提升

资源利用率,尤其适用于超高清、交互式视频等低时延业务场景^[16-18]。在此背景下,基于 HTTP 的自适应流体系,如动态自适应流媒体(Dynamic Adaptive Streaming over HTTP, DASH)^[19]和 HTTP 直播流(HTTP Live Streaming, HLS)^[20]仍为主流技术路径;其通过多码率切片与客户端在线决策提升用户体验(Quality of Experience, QoE),但在强波动网络中易出现频繁切换与体验不稳定。因此,围绕网络状态估计与码率决策,既有工作形成了多类典型机制:基于吞吐量估计的速率自适应(如加权移动平均估计器)用于预测短期网络能力并指导码率选择^[21]。基于缓存状态的自适应传输范式将缓存状态映射到码率选择上,以降低卡顿风险并提升播放稳定性,且已在大规模视频服务的线上实验中得到验证^[22]。除客户端侧决策外,也有研究从网络侧出发,构建基站负载与缓存状态的交互关系并联合优化缓存与带宽分配,以提升资源利用率与传输质量^[23]。近期工作进一步联合考虑边缘缓存与无线资源调度,结合视频分段与多码率特性对传输与服务策略进行协同优化^[24]。此外,网络辅助与边缘缓存增强逐渐成为重要补充,例如 EdgeDASH 利用网络侧辅助策略引导客户端请求,充分利用边缘缓存并提升分发稳定性^[25]。

在自适应决策层面,除启发式与控制模型外,经典基线 mDASH^[26]将码率自适应建模为马尔可夫决策过程,在时变网络下平衡视频质量与波动影响,因而常被作为代表方案参与实验。深度学习兴起后,ABR 进入系统化演进阶段:Oboe 通过离线训练与在线微调实现网络状态到传输参数的端到端映射^[27]。PiTree 以决策树形式提升强化学习模型的解释性与实用性^[28],Stick 融合启发式算法与深度强化学习策略,在传输效率与计算复杂度之间取得折中^[29]。面向实时直播与自适应流媒体等更苛刻场景,近期研究进一步强调高 QoE、可部署性以及多目标约束:GreenABR 将能耗约束纳入深度强化学习(Deep Reinforcement Learning, DRL)控制框架,在保证感知质量的前提下联合优化码率选择与终端能耗^[4]。GreenABR+进一步扩展了能耗、QoE、网络条件的综合建模,以增强跨场景适应性^[30]。面向 DASH 的 DRL 方法通常以带宽、缓冲区和历史质量为状态,端到端输出码率决策,在多种网络环境下提升 QoE^[31]。在多用户与边缘协同方面,QAVA 将多用户 HTTP 直播流汇聚至边缘智能代理,综合带宽、缓存与 QoE 公平性,对客户端码率与缓存策

略进行集中调度^[32]。低时延直播场景下,ALVS (Adaptive Live Video Streaming)将码率选择与播放速率控制纳入统一的 DRL 决策框架,依据带宽与缓冲区状态对各视频分片端到端决策,在真实带宽轨迹上性能优于启发式及学习型 ABR 算法^[33]。需要指出的是,现有方法主要围绕单视频流的 ABR 展开。而 Wang 团队提供了从单流 ABR 到双流 ABR 协同传输的新思路,在降低数据量提升传输效率方面取得了显著的效果^[34-35]。但其流类型与优化目标仍主要服务于人类感知任务。与此同时,研究视角进一步扩展到端边云多层协同架构,相关工作联合边缘转码与客户端增强,在成本、时延与视频质量之间实现更细粒度的权衡^[36]。面向视觉智能任务,Visual E2C 通过动态调整数据规模与模型版本实现推理任务的自适应调度^[37],并进一步探索 AI 驱动的视觉端边云协作架构以提升复杂业务场景下的视频处理与传输效率^[38]。

在真实时变网络环境中,不同信息载体在失真容忍度、时延敏感性与带宽收益上差异显著,传输策略需要随场景/网络状态动态切换并进行多流联合优化。尤其在机器主导的场景下,接收端往往只需要结构化语义信息即可完成感知与决策,继续传输完整视频流不仅带来显著带宽与时延开销,且对任务收益有限。因而,现有单视频流 ABR 范式将决策空间主要限定在码率调节,难以适配不同场景下视频流与结构化语义信息的取舍与协同传输,相关机制仍存在进一步拓展空间^[16-18]。

3 研究动机

本节将介绍所提方法的研究动机,分析端云单流与端边云多流传输架构在带宽利用、传输时延和信息冗余方面的表现差异,揭示现有架构的局限性与改进的必要性。

在传统端云单视频流传输架构下,前端摄像头将像素级视频以 H. 264/H. 265^[8]压缩后直接上行至云端集中处理,如图 1(a)所示。该模式对人眼观看友好,但在机器视觉^[39]主导的场景中暴露出带宽与时延瓶颈:一方面,带宽开销大且随接入规模线性增长,即便在高效编码下仍难承载密集接入。以 720P 监控视频为例,H. 264/H. 265 编码下码率通常约为 1~3 Mbps^[40]。多路并发时整体带宽需求将呈线性增长,在弱网及大规模场景中极易触发排队与重传,导致端云传输时延急剧上升。另一方面,单视频

流模式缺乏对网络场景状态的动态适应能力,无法根据实际链路条件灵活调整传输内容,即便仅需机器判

读,也不得不无差别上传大量与决策无关的冗余背景信息,进一步加重带宽浪费与云端算力挤占。

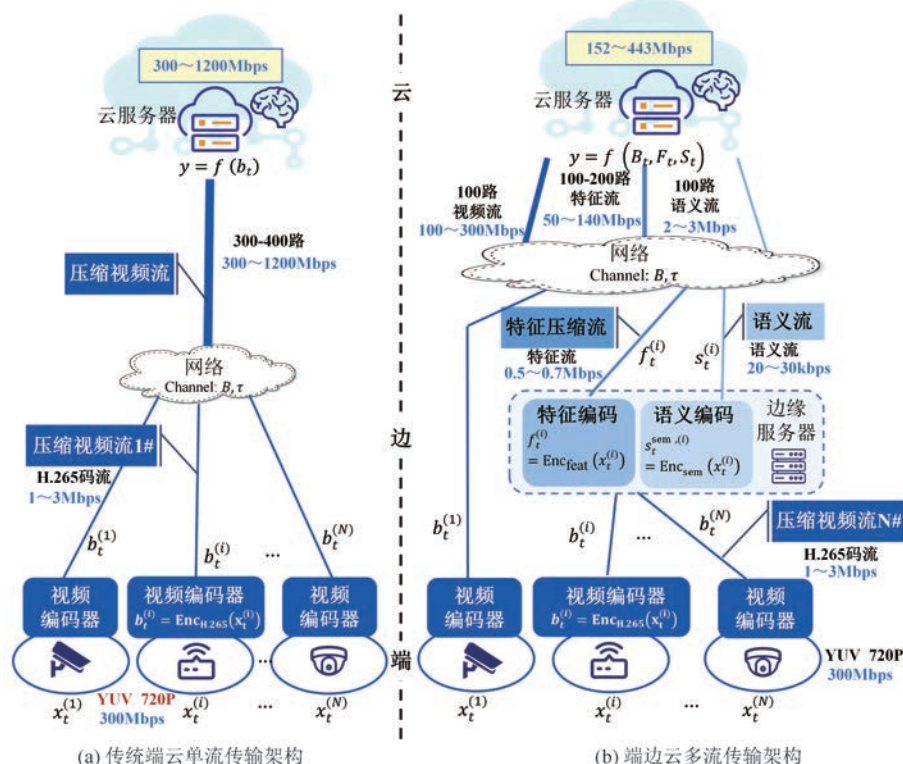


图1 传统单视频流传输与多数据流传输对比

然而,在大量机器视觉主导的场景中,任务通常只依赖目标级语义或中间特征,无需完整像素级视频,如图2所示。

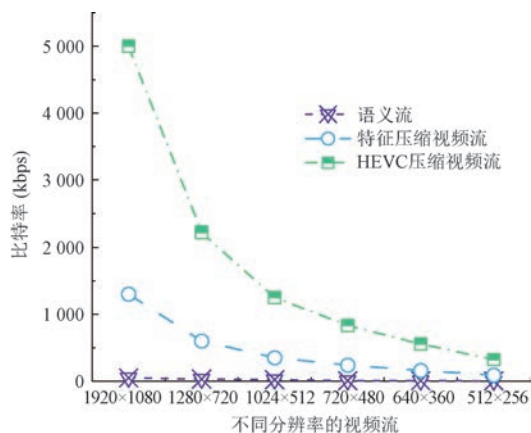


图2 三种数据流码率对比图

在统一的720p基线与相同编码设置下,HEVC单路码率约2.222 Mbps,特征压缩视频流码率约0.6 Mbps,约为HEVC的27%。在保留可视化/回溯能力的同时可节省约73%的上行带宽^[34-35]。语义流约30 kbps,对分辨率近乎不敏感,仅为HEVC的约1.35%。而且已有研究表明,仅传输特征压缩流或结构化数据即可满足大多数智能分析需求,其

数据量通常不超过原始视频的30%^[39]。

鉴于上述差异,本文提出端边云多流协同传输架构(图1(b)),利用边缘侧的地理与计算优势,先行完成去冗余与结构化处理,并根据链路状态自适应选择上行数据形态,从源头削减无关信息。该架构一方面显著降低汇聚链路占用与云端核心网压力,避免多路接入时带宽迅速饱和。另一方面缩短数据传输路径以降低长距离传输造成的高时延,能够在对实时性要求极高的场景^[41]下实现快速响应。并在弱网条件下减少无差别上传的冗余像素,抑制时延突增。因此,在动态网络场景中,如何在该架构下联合决定“传什么(流型)”与“传多少(压缩率)”,在满足视频质量约束下最小化传输时延是本文关注的核心问题。

4 系统模型

本节进一步阐述端边云多流协同传输整体架构设计,并介绍作为后续优化输入的多流生成模型。该模型由基于时空特征提取的语义分割模块和面向视觉焦点区域的特征压缩模块组成。

4.1 端边云多流协同传输架构

为解决传统端云单视频流架构在带宽受限及多样化应用场景下存在的无差异传输、时延高、带宽资

源浪费等问题,本文提出一种端边云多数据流协同传输架构,本节对其进行形式化描述,明确各层输入、输出与运行关系。(1)终端侧:负责视频数据的实时采集与轻量预处理。鉴于其计算资源有限,仅将原始或粗略压缩后的视频流上传至就近边缘节点。(2)边缘侧作为处理中枢:获取来自终端侧的压缩视频数据,并对这些数据进行处理,生成特征压缩视频流和语义流。同时实时感知当前场景的网络状况,并据此自适应激活上行数据流进行差异化传输。(3)云侧:接收并处理来自边缘的上行数据,完成事件检测、行为识别等高层任务,执行全局管理与智能决策。其中边缘侧包含以下三大核心模块:多流生成模块,以终端视频为输入对上行视频流进行结构化与非结构化分层编码,输出三种可供传输的数据形态;场景感知状态划分模块,负责实时感知当前可用带宽及其波动幅度,将复杂连续的网络环境映射为有限数量的典型场景;多流协同优化传输模块,将场景感知划分结果及三类数据流作为输入,在视频质量约束下联合决定输出的激活流型、压缩率与上传策略,目标是最小化传输时延。

4.2 多数据流生成模型

为了适应复杂异构视觉场景下多任务驱动的数据传输需求,本文多流生成模块在边缘侧将同源视频序列 $V = \{I_t\}_{t=1}^T$ 映射为不同的上行数据形态 $\mathcal{S} = \{s_1, s_2, s_3\}$, 分别对应普通压缩视频流 s_1 、基于焦点

区域的特征压缩视频流 s_2 以及语义流 s_3 , 形式上,可记为 $s_k = Enc_k(V; r_{s_k}), k \in \{1, 2, 3\}$, 其中, $Enc_k(\cdot)$ 表示第 k 类数据流的编码算子, r_{s_k} 为相应的码率或压缩率配置。三类数据流在数据量大小、语义承载能力与处理效率等方面具有差异性,可灵活适配不同网络条件。其中,普通压缩视频流 s_1 , 由标准视频编码器直接作用于原始帧序列得到: $b_t^{s_1} = Enc_{HEVC}(I_t; r_{s_1}), x_{s_1} = \{b_t^{s_1}\}_{t=1}^T$, 其中 $Enc_{HEVC}(\cdot)$ 代表 H. 264/AVC、H. 265/HEVC 或 MPEG 等标准编码流程。该类编码主要依赖块匹配的运动补偿与变换/熵编码,仅压缩低层冗余(如灰度、边缘),缺乏场景语义建模能力。因此,其压缩效率受限于视频质量保障需求,存在理论上限^[3]。尽管该数据流在可视回放方面表现良好,但在以机器视觉为主导的应用场景中,其包含的大量无关背景像素不仅增加了冗余信息,还导致较高的带宽开销。为此,本文构建如下模型加以解决。

4.2.1 时空特征提取语义分割模型

为从源头降低上行数据量并面向无需人眼观看的机器视觉任务,本文构建时空特征提取的语义分割模型生成语义流 s_3 。设输入视频帧序列为 $\{I_t\}_{t=1}^T$ 。该模型沿用 DeepLabV3+^[42] 的编解码框架,将其编码端 ResNet-101 主干网络替换为基于 C3D(Convolutional 3D Network)的 2+1D 时空特征提取器,如图 3 所示。

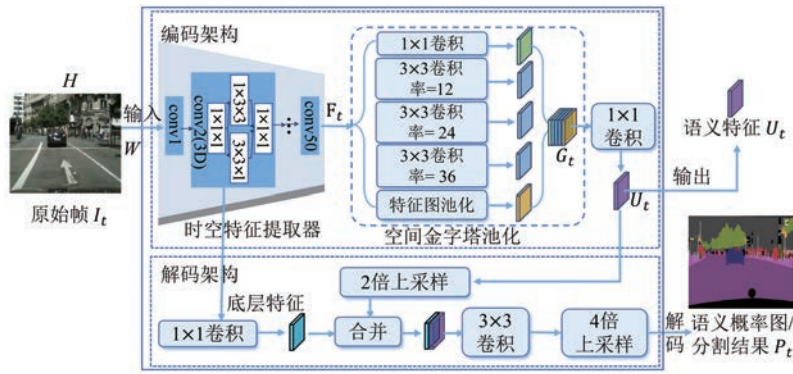


图 3 时空特征提取语义分割模型

将 $3 \times 3 \times 3$ 三维卷积分解为时间卷积 $3 \times 1 \times 1$ 与空间卷积 $1 \times 3 \times 3$, 在保留时序建模能力的同时降低参数规模和计算开销。编码器输出时空特征 F_t 后,经空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)进行多尺度上下文聚合得到 G_t 。再通过 1×1 卷积压缩为低维语义特征 U_t 。解码分支以 U_t 恢复语义概率图 P_t 。而上行仅传输 $\{U_t\}$ 作为语义流 s_3 。传输阶段对 $\{U_t\}$ 进行标量量

化与上下文熵编码,得到语义比特流 x_{s_3} 及其码率 r_{s_3} 。整体流程见算法 1。

算法 1. 时空特征提取语义分割与语义流生成

输入: 视频帧序列 $\{I_t\}_{t=1}^T$

输出: 语义概率图 $\{P_t\}$, 低维语义特征序列 $\{U_t\}$, 语义比特流 x_{s_3} , 码率 r_{s_3}

1. 初始化 语义比特流 $x_{s_3} \leftarrow \emptyset$
2. FOR $t = 1, 2, \dots, T$ DO

3. 构造时序输入片段 $X_t = \{I_{t-L}, \dots, I_t, \dots, I_{t+L}\}$
4. 将 X_t 输入 $2+1D$ 时空卷积骨干网络得到:

$$F_t \leftarrow \text{Backbone}_{2+1D}(X_t)$$
5. 将 F_t 输入 ASPP 和解码器, 得到:

$$P_t \leftarrow \text{Softmax}(g(F_t))$$
6. 通过 1×1 卷积映射到低维语义特征 $U_t \leftarrow h(F_t)$
7. 对 U_t 进行量化与熵编码

$$\hat{U}_t = Q(U_t; \Delta), b_t^{s_3} = AE(\hat{U}_t)$$
8. 将当前帧 $b_t^{s_3}$ 追加到语义比特流 $x_{s_3} \leftarrow [x_{s_3}, b_t^{s_3}]$
9. END FOR
10. RETURN $r_{s_3} = |x_{s_3}| \cdot \text{fps}/T$

其中, Backbone_{2+1D} 表示采用 $1 \times 3 \times 3$ 空间卷积与 $3 \times 1 \times 1$ 时间卷积分解 $3 \times 3 \times 3$ 卷积核的 $2+1D$ 时空卷积网络。 $g(\cdot)$ 与 $h(\cdot)$ 分别对应 DeepLabV3+ 中的语义分割头 (ASPP 与 Decoder) 和语义特征映射头。 $Q(U_t; \Delta)$ 和 $AE(\hat{U}_t)$ 分别表示标量量化算子与上下文自适应熵编码器。

该语义流不再关注可视化回放, 而是以结构化特征形式直接服务于云端决策等机器视觉任务, 可以在保持较强语义判别能力的同时具有极低码率开销^[39]。在实际交通场景下, 典型 5 Mbps 的输入视频经该模型处理后可压缩至约 50 Kbps^[3], 在显著降低传输负担的同时, 仍能有效支持机器任务的识别需求。

4.2.2 面向视觉焦点区域的特征压缩模型

为在语义信息与压缩效率间取得更优平衡, 本文基于上节输出的语义图 $\{P_t\}_{t=1}^T$ 构建面向视觉焦点的特征压缩模型, 生成特征流 s_2 , 如图 4 所示。特征流形式上表示为: $s_2 = \text{Enc}_2(V; r_{s_2}) = \text{Enc}_{\text{Feat}}(\mathcal{G}_{\text{focus}}(V; \{P_t\}, \text{ratio}); r_{s_2})$, 其中, $\mathcal{G}_{\text{focus}}(\cdot)$ 表示由式(1)–(3)定义的焦点重采样算子, $\text{Enc}_{\text{Feat}}(\cdot)$ 为特征压缩编码器。具体实现上, 首先利用 DeepLabV3+ 对关键目标 (如车辆、行人等) 进行检测与区域提取并结合类别及空间特征构建语义引导能量矩阵, 量化区域显著性。随后, 将第 t 帧划分为 $M \times N$ 的规则网格, 记行高向量: $\mathbf{h} = [h_1, h_2, \dots, h_M]^T$ 与列宽向量: $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$ 。根据语义分割结果, 对特征图中每个像素赋予类别标签, 并进一步执行基于类别的二次标记:

$$L^{(2)}(x, y) = \begin{cases} 1, & (x, y) \in \Omega_{\text{focus}} \\ 0, & (x, y) \in \Omega_{\text{background}} \end{cases} \quad (1)$$

其中, Ω_{focus} 表示视觉焦点及关键目标区域。每个网格块 (m, n) 内对 $L^{(2)}$ 求平均, 得到焦点重要性分数

$\alpha_{mn}^t \in [0, 1]$, α_{mn}^t 越大表示该网格越重要。由此构造语义驱动的空间权重场 $\mathbf{A}_t = [\alpha_{mn}^t]_{M \times N}$ 。

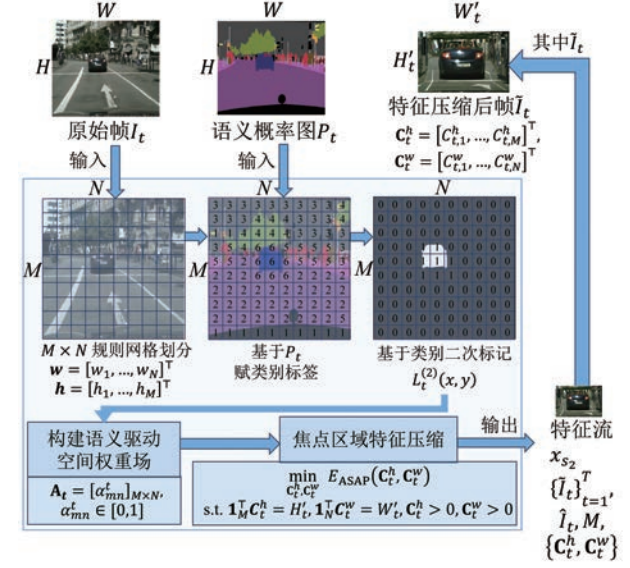


图 4 面向视觉焦点区域的特征压缩模型

记 $\mathbf{C}_t^h = [C_{t,1}^h, \dots, C_{t,M}^h]^T$, $\mathbf{C}_t^w = [C_{t,1}^w, \dots, C_{t,N}^w]^T$ 分别为压缩后第 t 帧在行、列方向上的网格尺寸向量。原始分辨率为 $H = \mathbf{h}^T \mathbf{1}_N$, $W = \mathbf{1}_M \mathbf{w}^T$, 在压缩倍率 ratio 下目标尺寸为 $H'_t = H/\text{ratio}$, $W'_t = W/\text{ratio}$ 。

为平衡压缩强度与视觉保真度, 本文采用 As-Similar-As-Possible (ASAP) 能量作为几何与语义一致性正则, 在焦点区域抑制各向异性变形, 在非焦点区域允许更强下采样。通过动态调整网格尺寸控制不同区域分辨率大小^[34], 定义为

$$E_{\text{ASAP}} \|\mathbf{A}_t \odot (\mathbf{C}_t^h - \mathbf{C}_t^w)\|_F^2 \quad (2)$$

其中, $\|\cdot\|_F$ 为 Frobenius 范数, $\odot$ 为按元素相乘, α_{mn}^t 越大, 对应行高 $C_{t,m}^h$ 、列宽 $C_{t,n}^w$ 越趋向于保持均匀。在给定目标尺寸约束下, $\mathbf{C}_t^h$ 和 $\mathbf{C}_t^w$ 可通过以下凸二次规划问题求解:

$$\begin{aligned} \min_{\mathbf{C}_t^h, \mathbf{C}_t^w} E_{\text{ASAP}}(\mathbf{C}_t^h, \mathbf{C}_t^w) \\ \text{s.t.} \quad \mathbf{1}_M^T \mathbf{C}_t^h = H'_t, \mathbf{1}_N^T \mathbf{C}_t^w = W'_t, \\ \mathbf{C}_t^h > 0, \mathbf{C}_t^w > 0 \end{aligned} \quad (3)$$

其中, $\mathbf{1}_M, \mathbf{1}_N$ 为全 1 向量, 该问题属于带线性等式和边界约束的凸二次规划。本文采用增广拉格朗日法 (Augmented Lagrangian Method) 进行数值求解。在目标函数中对等式约束加入二次罚项并结合拉格朗日乘子迭代, 从而加快收敛^[43]。通过求解上述问题, 原视频 $W \times H$ 在压缩倍率 ratio 下被映射为分辨率约为 $W'_t \times H'_t$ 的重采样帧, 规则网格被调整为高 $\mathbf{C}_t^h$ 、宽 $\mathbf{C}_t^w$ 的不等网格, 非焦点区域施加更高压缩率。基于此生成特征压缩视频流 s_2 。解码

端根据记录的网格参数 $\{C_t^h, C_t^w\}$ 执行反向轴对齐重采样以恢复空间布局并保持跨帧一致性, 计算 SSIM 以评估重采样帧与原帧的相似度。整体流程见算法 2。

算法 2. 面向视觉焦点区域的特征压缩算法

输入: 视频帧序列 $\{I_t\}_{t=1}^T$, 对应语义概率图 $\{P_t\}_{t=1}^T$, 网格规模 M, N , 压缩倍率 $ratio$

输出: 非均匀网格参数 $\{C_t^h, C_t^w\}$ 、焦点 SSIM 序列 $\mathbf{M}$ 、压缩后帧 $\{\tilde{I}_t\}$ 、特征流 x_{s_2}

1. 初始化特征流 $x_{s_2} \leftarrow \emptyset$, 焦点 SSIM 序列 $\mathbf{M} \leftarrow \emptyset$
2. FOR $t = 1, 2, \dots, T$ DO
3. 由 P_t 构建二次标记 $L_t^{(2)}(x, y)$, 得到尺寸为 $W \times H$ 的车辆焦点掩码 F_t
4. 按 $M \times N$ 规则网格对 $L_t^{(2)}$ 求块平均, 得到

$$\mathbf{A}_t = [\alpha_{mn}^t]_{M \times N}$$

5. 设目标压缩后分辨率为 $H'_t = H/ratio, W'_t = W/ratio$, 以 $\mathbf{A}_t$ 为权重, 按照式(2)和(3)求解凸二次规划, 得到第 t 帧网格参数向量 (C_t^h, C_t^w) , 更新 H'_t, W'_t
6. 根据 (C_t^h, C_t^w) 对原始帧 I_t 进行轴对齐非均匀压缩与拼接, 得到压缩后帧 $\tilde{I}_t \in \mathbb{R}^{W'_t \times H'_t \times 3}$, 对压缩帧序列 $\{\tilde{I}_t\}$ 提取特征 $U_t^{s_2}$, 并对其进行特征量化与熵编码得到:

$$\hat{U}_t^{s_2} = Q(U_t^{s_2}; \Delta), b_t^{s_2} = AE(\hat{U}_t^{s_2}), x_{s_2} \leftarrow [x_{s_2}, b_t^{s_2}]$$

7. 在解码侧, 根据记录的 (C_t^h, C_t^w) 将 $\tilde{I}_t$ 反向重采样至原分辨率, 得到重建图像 $\hat{I}_t$, 并在车辆掩码 $F_t = 1$ 的区域内计算: $SSIM_t^{\text{focus}} = SSIM(I_t, \hat{I}_t | F_t = 1)$, $\mathbf{M} \leftarrow [\mathbf{M}, SSIM_t^{\text{focus}}]$
8. END FOR
9. RETURN $r_{s_2} = |x_{s_2}| \cdot \text{fps}/T, \{C_t^h, C_t^w\}, \{\tilde{I}_t\}, x_{s_2}, \mathbf{M}$

该模型为端边云多流协同传输系统提供了兼顾可视质量与传输效率的有效方案。

5 基于场景感知的多数据流协同优化传输方法

5.1 问题表述

本文旨在满足人类视觉质量约束的前提下, 最小化传输时延, 实现对不同数据流的自适应选择与压缩率配置。前述三类上行数据流 $\mathcal{S} = \{s_1, s_2, s_3\}$ 在数据规模、语义承载能力与可视回溯能力等方面各具差异, 可根据网络场景状态灵活选择。然而, 在实际端边云系统中, 复杂异构网络导致带宽受限与显著波动, 对多流传输提出挑战。设感知可用带宽为 B , 带宽波动幅度为 ΔB , 则场景状态变量为 $\theta =$

$(\bar{B}, \Delta B)$, 在该状态下, 依据得分函数计算得到权重 $w_s(\theta)$ 最高的流为 s^* (具体规则见 5.3 部分), 其压缩率为 $r_{s^*} \in [r_{\min}, r_{\max}]$ 。设原始视频的数据量为 D , 当前场景状态 θ 的传输时延可表示为 $T(\hat{\mathbf{z}}, r_{s^*}; \theta) = \mathbf{1}_3^T (\text{diag}(B)^{-1} \text{diag}(r_{s^*})^{-1} \text{diag}(\hat{\mathbf{z}}) D)$, 当系统在连续 K 个场景下运行时, 优化问题可形式化表述为

$$\text{P1: } \min_{\{\mathbf{z}_k, r_{s^*,k}\}} \sum_{k=1}^K T_k \quad (4a)$$

$$\text{s. t. } Q_{\text{SSIM}}(D, r_{s^*} D) \geq C_B, \forall \mathcal{S} \in \{s_1, s_2\}, \forall k \quad (4b)$$

$$r_{\min} \leq r_{s^*,k} \leq r_{\max}, \forall s^* \in \{s_1, s_2, s_3\}, \forall k \quad (4c)$$

$$\hat{\mathbf{z}}_k = \mathbf{e}_{s^*} \in \{0, 1\}^3, \mathbf{1}_3^T \hat{\mathbf{z}}_k = 1, \forall k \quad (4d)$$

$$r_{s_2} \in \mathcal{R}, \mathcal{R} = \{r^{(i)} \mid r^{(i)} > 1, i = 1, 2, \dots, L\} \quad (4e)$$

约束(4b)将 SSIM 作为视频质量约束条件, C_B 为依据网络吞吐量 B 查表得到的云端可接受的视频质量下限, 该约束仅对 s_1, s_2 生效, s_3 无此约束, 其作用在于确保压缩后视频具备可接受的视觉质量。约束(4c)限定了压缩率始终处于可行范围内。约束(4d)中, $\mathbf{e}_{s^*}$ 表示与被选数据流 s^* 对应的三维标准基向量, 即 $\mathbf{e}_{s_1} = [1, 0, 0]^T, \mathbf{e}_{s_2} = [0, 1, 0]^T, \mathbf{e}_{s_3} = [0, 0, 1]^T$ 。因此, $\hat{\mathbf{z}}_k = \mathbf{e}_{s^*} \in \{0, 1\}^3$ 定义了 One-Hot 选择向量, 保证每个场景下仅有被选中的流 s^* 的时延项进入目标函数。约束(4e)要求流 s_2 取自上节中焦点区域特征压缩模型输出的候选压缩率集合 $\mathcal{R}$, 且所有候选压缩率大于 1。由上述建模可知, P1 是一个非凸组合优化问题。其中, 变量 $\hat{\mathbf{z}}_k$ 和 $r_{s^*,k}$ 均为离散取值, 且存在耦合关系, 导致整体搜索空间随场景数 K 呈指数级增长, 构成典型的组合优化问题, 难以通过穷举或解析方法直接求解。此外, 约束的非凸性进一步增加了求解复杂度。尤其是约束(4b)中 SSIM 与压缩率 r_s 之间缺乏明确解析函数关系, 无法直接代入优化模型。需要指出的是, P1 以连续 K 个场景的平均时延为优化目标, 刻画系统整体传输效率; 而在视频会议、直播与短视频等时延敏感业务中, 少量极端时延也可能显著影响用户体验与任务可靠性。为评估最坏情形下的个体体验, 本文在 P1 基础上引入 min-max 变体 MinMaxP1:

$$\text{MinMaxP1: } \min_{\{\hat{\mathbf{z}}_k, r_{s^*,k}\}} \kappa \quad (5a)$$

$$\text{s. t. } T_k(\hat{\mathbf{z}}_k, r_{s^*,k}; \theta_k) \leq \kappa, \forall k \in \{1, 2, \dots, K\}, \quad (5b)$$

该变体沿用式(4b)~(4e)的约束及原有流选择与压缩率决策空间,以最小化跨场景最大时延为目标,更强调最差场景用户体验。相对地,P1与MinMaxP1分别对应系统平均性能与最坏性能。两者原有决策变量与约束保持一致,主要区别在于优化目标不同。因此离散变量耦合与非凸约束等计算难点在MinMaxP1中同样存在。鉴于本文重点关注

整体传输效率,后续方法主要以P1的平均时延最小化为主;MinMaxP1仅作为公平性变体,用于分析平均性能与最坏时延之间的权衡。为应对P1的组合搜索复杂度与非凸约束,本文提出面向场景感知的多数据流协同优化框架,如图5所示,将原问题分解为场景状态划分子问题P2与多流协同传输优化子问题P3。

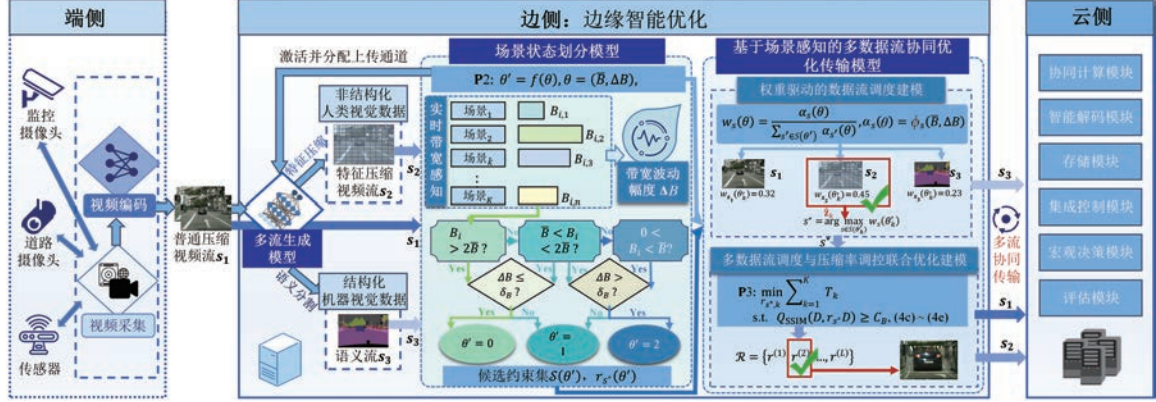


图5 基于场景感知的多数据流协同优化传输方法

5.2 场景感知状态划分模型

为降低问题P1的搜索复杂度,本文提出场景感知状态划分模型,作为第二个子问题P2,其核心作用在于:通过对网络场景状态进行建模与分类,将复杂连续状态空间映射为有限数量的典型场景,进一步生成候选约束集,为后续优化传输缩减搜索空间。具体而言,系统通过实时感知当前可用带宽大小 B_i 与波动情况,提取可用带宽均值 $\bar{B}$ 与波动幅度 ΔB ,并据此完成状态变量的场景映射。设传输路径为 $\mathcal{P} = \{\ell_0, \ell_1, \dots, \ell_M\}$,其中第 i 条链路的带宽容量为 B_i ,瞬时利用率 $u_i \in [0, 1]$,则瓶颈容量和可用带宽^[44]分别为

$$B_{\text{bott}} = \min_{0 \leq i \leq M} B_i, A_P = \min_{0 \leq i \leq M} [B_i(1 - u_i)] \quad (6)$$

然而,瓶颈容量通常不可先验获得。鉴于边缘节点与视频源距离近,其RTT相较边侧到云端的传输可近似忽略^[39],而边到云的通信易受公网与并发业务影响并往往成为性能瓶颈,本文将带宽感知与传输策略设计聚焦于边云链路。设链路 i 的探测任务量为 V_i ,起止时间为 $t_i^{(s)}$ 与 $t_i^{(e)}$,路径可用带宽采用自导拥塞思想的主动探测:

$$\hat{B}_i = \frac{V_i}{t_i^{(e)} - t_i^{(s)}}, i = 0, 1, \dots, M, \quad (7)$$

但可用带宽受到最小链路带宽限制:

$$\hat{B}_{\mathcal{P}} = \min_{0 \leq i \leq M} \hat{B}_i = \min_{0 \leq i \leq M} \frac{V_i}{t_i^{(e)} - t_i^{(s)}} \quad (8)$$

在长度为 $T_w = 10$ s的滑动时间窗内,统计 $\hat{B}_{\mathcal{P}}$ 可用带宽的均值与波动幅度:

$$\bar{B} = \frac{1}{K} \sum_{k=1}^K \hat{B}_{\mathcal{P}}^{(k)}, \Delta B = B_{\max} - B_{\min} \quad (9)$$

据此定义联合场景状态变量: $\theta = (\bar{B}, \Delta B)$ 。并将其作为核心的场景感知状态划分模型的输入。从建模角度,端边云多数据流协同传输的场景状态 θ' 由可用带宽及其波动等连续链路变量决定;在连续状态空间上直接优化不仅求解与在线判决复杂度高,也不利于在线调度与实现。为此,本文采用阈值离散化,将连续链路状态映射为有限个典型场景,在保留关键链路差异的同时形成可部署的有限状态决策。已有工作在不同场景中采用三状态{Good, Medium, Bad}表征来支撑在线决策;例如,多跳无线网络中利用三状态链路质量驱动时变调度^[45]、LEO卫星数据采集中基于统计构建分段信道状态用于在线资源分配^[46]、水声链路中将链路状态建模为三状态随机过程(如马尔可夫链)以支持评估与策略设计^[47]。受上述工作启发,本文将 θ' 离散为{常规,干扰,恶劣}三类:分别对应高带宽且低波动的正常区间、均值下降且波动增强的轻度拥塞区间,以及持续低带宽并伴随高时延的退化区间。

基于此,本文首先依据经验阈值 δ_B 将场景粗分为稳定与波动两类,分别对应 $\sigma = 0$ 与 $\sigma = 1$:

$$\sigma = \begin{cases} 0, & \Delta B \leq \delta_B \\ 1, & \Delta B > \delta_B \end{cases} \quad (10)$$

随后结合 $\hat{B}_P$ 与 θ 进一步将复杂的网络环境映射为有限数量的典型场景,并形式化为子问题 P2:

$$P2: \theta' = f(\theta), \theta = (\bar{B}, \Delta B), \quad (11)$$

$$\theta' = \begin{cases} 0, & (\hat{B}_P > k_B \bar{B}) \wedge (\Delta B \leq \delta_B), \\ 1, & (\bar{B} < \hat{B}_P \leq k_B \bar{B}) \vee (\delta_B \leq \Delta B < k_D \delta_B), \\ 2, & (0 < \hat{B}_P \leq \bar{B}) \vee (\Delta B \geq k_D \delta_B) \end{cases} \quad (12)$$

其中, $\theta' \in \{0, 1, 2\}$ 分别表示常规、干扰、恶劣三类场景。而且 θ' 作为输入可以驱动以下三项关键机制: (1) 压缩率的搜索范围约束; (2) 候选数据流类型的筛选与优先级判定; (3) 调度权重函数 $w_s(\theta)$ 的计算输入, 详见式(14)。因此场景划分结果 θ' 不仅缩减了搜索空间, 还为后续调度提供约束输入。具体而言, 系统在不同场景下对候选数据流集合 $\mathcal{S}(\theta')$ 与被选中的压缩率集合 $r_s(\theta')$ 施加限制:

$$\begin{aligned} \mathcal{S}(\theta') &= \begin{cases} \{s_1, s_2\}, & \theta' = 0 \\ \{s_2\}, & \theta' = 1 \\ \{s_3\}, & \theta' = 2 \end{cases} \\ r_s(\theta') &\subseteq [r_{\min}, r_{\max}] \\ r_{s_2} &\in \mathcal{R} = \{r^{(i)} \mid r^{(i)} > 1, i = 1, 2, \dots, L\} \end{aligned} \quad (13)$$

在此约束下: (1) 常规场景中, 带宽充足且带宽波动幅度小, 系统优先选择人类视觉上的非结构化数据流, 包括普通压缩视频流和特征压缩视频流; (2) 恶劣场景中, 网络波动剧烈或可用带宽严重受限。系统将传输资源优先分配给面向机器视觉的关键语义流, 确保接收端在极端带宽条件下仍能获取最核心的信息内容。 (3) 干扰场景中, 可用带宽大小适中或者带宽波动幅度中等, 网络较稳定。优先考虑特征压缩视频流。系统会根据当前网络状态及视频质量约束, 从面向人类视觉的特征压缩视频流集合 $\mathcal{R} = \{r^{(1)}, r^{(2)}, \dots, r^{(L)}\}$ 中自适应选取最优的压缩率版本进行传输。并根据实时感知的可用带宽动态调整压缩率。在带宽受限情况下, 提高压缩率以降低数据量, 同时保障焦点区域图像质量。

5.3 多数据流协同优化传输模型

通过求解场景感知状态划分子问题 P2, 可获得场景状态划分结果 θ' 及其候选约束集 $\mathcal{S}(\theta')$ 。需要指出的是, $\mathcal{S}(\theta')$ 仅限定了不同场景下的候选数据流集合, 并未直接给出最终调度优先级。为区分候选流的相对重要性, 本文引入权重驱动的调度机制, 用于指导压缩率与调度策略的联合优化。在此基础上, 原问题 P1 被进一步分解为多数据流调度优化子问题 P3, 提出多数据流协同优化传输模型。其核

心目标是在已知场景状态 θ' 下, 依据调度权重 $w_s(\theta)$ 建模优先级, 并在约束集合内实现压缩率与多流传输策略的联合优化, 从而在保证视频质量的前提下, 最小化端边云整体传输时延。

5.3.1 权重驱动的数据流调度建模

为刻画不同场景下各数据流的传输优先级, 本文在 P2 的基础上定义调度权重:

$$w_s(\theta) = \frac{\alpha_s(\theta)}{\sum_{s' \in \mathcal{S}(\theta')} \alpha_{s'}(\theta)}, \alpha_s(\theta) = \phi_s(\bar{B}, \Delta B), \quad \sum_{s \in \mathcal{S}(\theta)} w_s(\theta) = 1 \quad (14)$$

其中, $\alpha_s(\theta)$ 表示第 s 类数据流在状态 θ 下的相对偏好得分, $\phi_s(\cdot)$ 用于衡量当前网络下数据流类型与传输资源的适配度, 可通过规则或数据驱动构造。该评分机制反映不同类型数据流在复杂网络环境中的适配度与重要性, 是优化模型调度决策的关键输入, 并最终决定选择变量 $\hat{z}_k$:

$$\hat{z}_k = \mathbf{e}_{s^*} \in \{0, 1\}^3, s^* = \arg \max_{s \in \mathcal{S}(\theta'_k)} w_s(\theta'_k) \quad (15)$$

其中, $\hat{z}_k$ 为 One-Hot 选择向量, 即在第 k 个场景下仅有一个元素取值为 1, 其余元素为 0, 表示只能选择单一的数据流。被选择的数据流为权重最高的流 s^* 。

5.3.2 多数据流调度与压缩率联合优化建模

在当前可用带宽 $\hat{B}_P$ 、场景状态划分结果 θ' 和流类型 s^* 被感知和确定后, 离散变量 $\hat{z}_k$ 被固定, P1 目标函数转化为单流情形下的时延最小化问题:

$$T(\hat{z}_k, r_{s^*, k}; \theta'_k) = \mathbf{1}_3^T (\text{diag}(\hat{B}_{P, k})^{-1} \text{diag}(r_{s^*, k})^{-1} \text{diag}(\hat{z}_k) D) \quad (16)$$

于是, 本文将压缩率调控与数据流调度的联合优化问题 P3 形式化表述为

$$P3: \min_{r_{s^*, k}} \sum_{k=1}^K T_k \quad \text{s. t. (4d) } \sim \text{(4e)} \quad (17)$$

由于 SSIM 综合考虑亮度、对比度和结构信息, 更贴近人类视觉感知, 相较于 PSNR 更适合作为视频传输的质量约束, 在视频压缩与传输质量评估中具有更高的建模准确性。本文在优化模型中采用其简化后的 SSIM 近似表达式:

$$Q_{\text{SSIM}}(x, y) = \left[\frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right]^\alpha \cdot \left[\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right]^\beta \left[\frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3} \right]^\gamma \quad (\alpha > 0, \beta > 0, \text{ and } \gamma > 0) \quad (18)$$

其中, 协方差 σ_{xy} 的计算公式为

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (19)$$

由公式中可以看出, SSIM 的值与图像压缩率 r_s^* 之间并不存在一种可以直接套用的显式函数关系。为探究二者的关联性, 本文采用皮尔逊相关系数分析方法对其进行度量。计算得到 SSIM 与压缩率 r_s^* 的相关系数 $\rho = -0.9963$, 可见二者之间存在显著的高度负相关关系^[35]。为直观展示该趋势, 本文在 Cityscapes 数据集上统计了不同压缩率 r_s^* 下焦点区域 SSIM 的变化规律, 如图 6 所示。

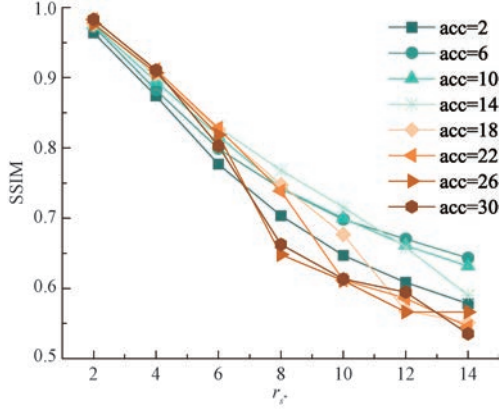


图 6 不同压缩率 r_s^* 下焦点区域 SSIM 的变化规律

在此基础上, 采用最小二乘法对两者关系进行二次回归拟合, 得到

$$Q_{SSIM} = -0.001 \cdot r_s^{*2} - 0.0558 \cdot r_s^* + 1.0506 \quad (20)$$

将视频质量下限记 C_B , 对式(20)求得

$$\frac{dQ_{SSIM}}{dr} = -0.002r - 0.0558 < 0, \forall r > 0 \quad (21)$$

由此可知, Q_{SSIM} 在可行域 $r_s^* > 0$ 上严格单调递减。对任意 $r_1 < r_2$ (且 $r_1, r_2 \in \mathcal{R}$), 有

$$Q_{SSIM}(r_2) - Q_{SSIM}(r_1) = \int_{r_1}^{r_2} \frac{dQ_{SSIM}(r)}{dr} dr < 0 \quad (22)$$

因此, 保证质量 $Q_{SSIM}(r) \geq C_B$ 等价于对压缩率施加上界约束, 可用于在优化过程中快速筛选候选压缩率, 从而显著降低搜索和求解开销。至此, P3 子问题被转化为带约束的非线性规划问题 P4。

$$\begin{aligned} \text{P4: } & \min \sum_{r_s^*, k=1}^K T_k \\ \text{s. t. } & -0.001r_s^{*2} - 0.0558r_s^* + 1.0506 \geq C_B \\ & 1 - r_s^* < 0, \forall \mathcal{S} \in \{s_1, s_2\}, \forall k \end{aligned} \quad (23)$$

在本文的建模框架下, 考虑到约束(4b)压缩率 r_s^* 与质量约束 Q_{SSIM} 之间的强非线性关系, 直接解

析求解难度较大。因此, 本文利用拉格朗日对偶方法^[48]对问题 P4 进行处理。具体而言, 构造如下拉格朗日函数:

$$\begin{aligned} \mathcal{L}(r_s^*, \hat{z}, \lambda, \mu) = & \sum_{k=1}^K T_k - 0.001\lambda r_s^{*2} - 0.0558\lambda r_s^* \\ & + 1.0506\lambda + \mu(1 - r_s^*) \end{aligned} \quad (24)$$

其中, $\lambda \geq 0, \mu \geq 0$ 分别为质量约束与下界约束的拉格朗日乘子。通过求解对偶问题:

$$\max_{\lambda, \mu \geq 0} \min_{r_s^*, \hat{z}} \mathcal{L}(r_s^*, \hat{z}, \lambda, \mu) \quad (25)$$

即可将原问题转化为无约束问题求解, 并通过迭代优化获得近似最优解。该方法的优势在于: 一方面避免了直接在非凸约束空间中进行全局搜索, 显著降低了计算复杂度; 另一方面, 可通过对偶变量的收敛性保证解的可行性与近似最优性。

5.4 算法设计

在完成 P1 的问题建模与 P2、P3、P4 的子问题分解后, 本节进一步给出算法实现流程, 用于在复杂动态网络场景下求解多数据流调度与压缩率联合优化问题。

算法 3 的目的是在保证视频质量约束的前提下, 最小化总传输时延。本文将多数据流协同优化传输问题形式化为带约束非线性规划问题 P4。整体流程分为两个核心阶段: 一是场景感知与流集合筛选阶段; 二是权重驱动的数据流与压缩率联合优化阶段。

算法 3. 基于场景感知的多流协同传输算法

输入: 可用带宽 $\hat{B}_P$, 带宽波动阈值 δ_B , 原始视频数据量 D , 质量约束 C_B , 场景数 K , 候选压缩率集合 $\mathcal{R}$, 数据流集合为 $\mathcal{S} = \{s_1, s_2, s_3\}$

输出: 各场景下的流选择与压缩率 $\{\hat{z}_k, r_{s^*, k}\}, k = \{1, 2, \dots, K\}$, 总传输时延 T_{total}

1. 初始化: $T_{\text{total}} \leftarrow 0$
2. FOR $k = 1, 2, \dots, K$ DO
3. //阶段一: 场景感知状态划分
4. 计算带宽均值 $\bar{B}$ 与波动幅度 ΔB , 构造状态变量 $\theta = (\bar{B}, \Delta B)$
5. 根据式(10)–(12)求解 P2, 输出场景划分结果 $\theta' \in \{0, 1, 2\}$ 、候选流集合 $\mathcal{S}(\theta')$ 与压缩率集合 $r_s(\theta')$
6. //阶段二: 多流调度与压缩率优化
7. 对 $\forall s \in \mathcal{S}(\theta'_k)$ 计算偏好得分 $\alpha_s(\theta)$, 依据式(14)计算调度权重 $w_s(\theta)$
8. $\hat{z}_k = e_{s^*} \in \{0, 1\}^3, s^* = \arg\max w_s(\theta'_k)$
9. IF $s^* = s_3$ THEN
10. 以 r_{s_3} 计算传输时延 T_k , 并更新总时延 $T_{\text{total}} \leftarrow T_{\text{total}} +$

T_k

11. ELSE $s^* \in \{s_1, s_2\}$

12. 根据式(23)构建优化问题 P4

13. $\mathcal{R}(\theta') = \{r \in \mathcal{R} \mid r_{\min}(\theta', \hat{B}_p) \leq r \leq r_{\max}(\theta', \hat{B}_p)\}, \mathcal{R} = \{r^{(1)}, r^{(2)}, \dots, r^{(L)}\}$

14. $\forall r_{s^*} \in \mathcal{R}(\theta')$ 检验质量约束 $Q_{\text{SSIM}} \geq C_B$
采用拉格朗日对偶方法进行迭代优化
输出近似最优压缩率 $r_{s^*,k}$

15. 计算时延 T_k , 更新总时延 $T_{\text{total}} \leftarrow T_{\text{total}} + T_k$

16. END IF

17. END FOR

18. RETURN $\{\hat{\mathbf{z}}_k, r_{s^*,k}\}, T_{\text{total}}$

具体而言,首先,设定时间窗口并初始化总传输时延 T_{total} 为 0(第 1 行);第一个阶段系统将感知到的可用带宽 $\hat{B}_p$ 作为输入计算带宽均值 $\bar{B}$ 与波动幅度 ΔB ,并据此构造 $\theta = (\bar{B}, \Delta B)$ (第 4 行),根据式(10)–(12)求解 P2,将连续状态空间映射为有限典型场景 θ' ,分别对应常规、干扰和恶劣场景,最终生成候选数据流 $\mathcal{S}(\theta')$ 与压缩率范围 $r_s(\theta')$,有效缩减了优化空间(第 5 行)。第二阶段是在第一阶段基础上,算法对候选流逐一计算偏好得分 $\alpha_s(\theta)$,并依据式(14)归一化得到调度权重 $w_s(\theta)$ (第 7 行);依据最优性准则选取调度权重最大的流 s^* ,定义 One-Hot 选择向量 $\hat{\mathbf{z}}_k$,保证每个场景下仅有被选中的流 s^* 的时延项进入目标函数(第 8 行)。在流选择确定后,激活该数据流,至此算法进入压缩率优化阶段。若激活的流是语义流 s_3 (第 9 行);因其数据结构化且对分辨率不敏感,因此无需考虑视频质量约束,直接进行传输并更新总时延 T_{total} (第 10 行)。若激活的流是人类视觉的视频流 s_1, s_2 (第 11 行),则需要根据式(23)构建并求解优化子问题 P4;首先利用启发式搜索机制在非连续的可行压缩率域中快速遍历与质量筛选,从而逼近近似最优解^[49],缩减压缩率候选集合 $\mathcal{R}$ 的搜索范围(第 13 行);随后,算法通过拉格朗日对偶方法将 P4 的非凸约束问题转化为对偶问题求解,根据式(24)和(25)对 P4 进行迭代优化,最终输出在场景状态 θ' 下满足 SSIM 约束的近似最优压缩率配置 $r_{s^*,k}$ (第 14 行),更新总传输时延(第 15 行)。当所有 K 个场景均完成决策与优化后,算法最终输出每个场景下的流选择 s^* 与最优压缩率 r_{s^*} 集合,最后将总传输时延 T_{total} 返回作为算法执行结果,算法终止(第 18 行)。

算法复杂度与近似度分析。设共有 K 个场景

窗口,场景感知阶段对每个窗口仅需计算带宽均值与波动幅度并完成阈值判决,时间复杂度为 $O(K)$ 。在多流调度与压缩率优化阶段,每窗至多涉及 $\{s_1, s_2, s_3\}$,其中 s_1 与 s_3 在给定场景下码率为预设,只有特征压缩流 s_2 需要在场景相关候选压缩率集合 $\mathcal{R}(\theta')$ 上搜索。记 $|\mathcal{R}_{\max}| = \max_{\theta'} |\mathcal{R}(\theta')|$,且对偶迭代次数为常数,则优化阶段的复杂度为 $O(K |\mathcal{R}_{\max}|)$,整体复杂度为 $O(K + K |\mathcal{R}_{\max}|) \approx O(K |\mathcal{R}_{\max}|)$ 。在本文实验设置下,场景数与候选规模均为小常数,因此计算开销可控,满足在线部署需求。从优化性质来看,子问题 P4 中采用两级近似求解策略:1)依据网格压缩精度与 SSIM 约束,对原集合 $\mathcal{R}$ 进行可行性筛选,得到 $\mathcal{R}(\theta')$,在保持全局最优解的前提下有效缩减搜索空间;2)在该可行域上用拉格朗日对偶进行有限步迭代,得到满足约束的 ϵ -近似压缩率解。该近似带来的时延与质量权衡将在第 6 章实验中进一步验证。

6 实验验证

为了验证方法的可行性和有效性,本文采用 Python 与 MATLAB 相结合的方式开展多流数据生成与端边云多流协同传输优化仿真实验。实验环境为 Intel Core i5@2.0 GHz 处理器的 64 位操作系统。

6.1 实验设置

6.1.1 数据集

实验数据采用真实移动网络带宽记录数据进行仿真模拟。考虑到实际网络环境中链路时延、带宽等参数变化相对复杂、不可控,且真实部署实验成本高昂,因此本文采用公开的 4G/LTE^[6] 网络带宽数据集,该数据集记录了 2015 年 12 月 16 日至次年 2 月 4 日期间,研究人员沿比利时根特市及其周边地区多条路线采集的 4G 网络下的带宽日志信息。数据涵盖 40 条轨迹,包含乘坐汽车、电车、自行车、火车、步行和公交等多种交通方式,每种方式持续采集 5 小时带宽数据。

多流数据生成实验数据采用 Cityscapes^[5] 数据集。该数据集包含来自全球不同城市的车载摄像头。Cityscapes 包含 3574 张城市交通场景图像,分辨率为 2048×1024 ,图像具有高语义密度和典型城市街景特征,有丰富的语义标注,涵盖地面、行人、车辆、建筑物等 19 类语义类别。该数据集具有高分辨率与典型城市交通视觉特征。

6.1.2 仿真环境设置

为验证所提出方法的有效性,本文基于多流生成模型产生的三类数据流,构建端边云协同仿真并评估其在复杂网络条件下的传输性能。实验采用 Python 3.10 等仿真平台,在配备主流 64 位操作系统及通用 CPU 的计算环境下完成。系统主要包括终端设备、边缘计算节点以及云服务器三大模块:首先,终端设备模拟城市交通场景的视频采集。其次,所采集的视频流被传输至部署 Jetson Nano 的边缘计算平台。该节点作为中间计算模块,部署了经过优化的 DeepLabV3+语义分割模型,结合自定义二次分割与区域压缩策略生成 s_1, s_2, s_3 , 同时部署基于场景感知的多流协同优化传输方法及带宽感知模型,按 P4 所述的多数据流调度与压缩率联合优化传输方案,动态调整传输的数据流类型以及压缩率,以适配实时变化的网络环境。最后,云服务器负责接收与进一步分析处理所传输的数据流。

网络环境方面,本文采用真实测量的 4G/LTE 带宽数据集^[6],将 5 小时内不同交通工具和城市区域下的带宽时间序列作为边缘节点与云端之间的仿真输入,以刻画时变带宽及其波动特性。记当前时窗的带宽波动幅度为 ΔB ,历史平均带宽为 $\bar{B}$,首先依据 ΔB 将网络划分为稳定与波动两类状态,进一步结合当前带宽与 $\bar{B}$ 的关系,将场景细分为三类:常规场景,当前带宽大于 $k_B \bar{B}$ 且 $\Delta B < \delta_B$,网络稳定、支持高质量视频传输;干扰场景,当前带宽为 $[\bar{B}, k_B \bar{B}]$ 或 $\Delta B \in [\delta_B, k_D \delta_B]$,存在中等程度抖动和偶发卡顿;恶劣场景,当前带宽为 $[0, \bar{B}]$ 或 $\Delta B > k_D \delta_B$,网络极不稳定,严重影响传输质量与时延。本文取 $\delta_B = 15$ Mbps, $k_B = 2$, $k_D = 2$,其中各阈值取值及三类场景划分的合理性于第 6.4.3 节的消融实验和第 6.4.4 节的超参数选择部分进行验证。

6.2 对比方法

在实验中,本文设计了以下 5 种不同的对比方法来和本文所提方法进行对比分析:

(1)固定比特率方法:一种无关网络吞吐量的视频压缩方式,在整个视频传输过程中采用固定压缩率,不考虑网络状态变化,是最基础的视频传输方案。

(2)比特率自适应(Markov Decision Process Based DASH, mDASH)^[26]:一种适用于实时视频流的低复杂度次优贪婪算法,该方法基于马尔可夫决策过程,综合考虑码率切换频率与幅度、缓存占用等对用户体验质量有显著影响的关键因素。

(3)缓存自适应方法^[23]:一种基于 Stackelberg 模型建立基站流量负载与缓存状态之间的关联模型,优化缓存命中率与带宽分配,从而提升视频传输系统效率。

(4)面向人类视觉的双流传输方法:该方法根据网络状态在普通压缩视频流与特征压缩视频流之间自适应切换,不涉及面向机器视觉的语义流传输。带宽充足时,优先传输普通压缩视频流以保障人类视觉质量;带宽受限时,切换至特征压缩视频流以降低传输负载,从而在保留关键视觉信息的同时减少传输时延和数据量。

(5)基于深度学习的自适应传输方法(ALVS-DRL)^[33]:该方法面向低时延直播,通过深度强化学习联合控制视频码率与播放速率,以 QoE 指标为优化目标。该方法在多条真实带宽轨迹上离线训练策略网络,在线阶段依据观测带宽与缓冲状态逐时隙输出离散码率/播放速率决策,以适应网络波动。其真实轨迹驱动与联合控制特性与本文高度相关。

6.3 实验结果

本文首先采用公开的 4G/LTE 数据集中六种不同的交通方式进行可用带宽感知实验,如图 7 所示。横坐标表示时间,纵坐标表示吞吐量,即可用带宽。仿真分析表明,网络带宽的波动受交通方式的

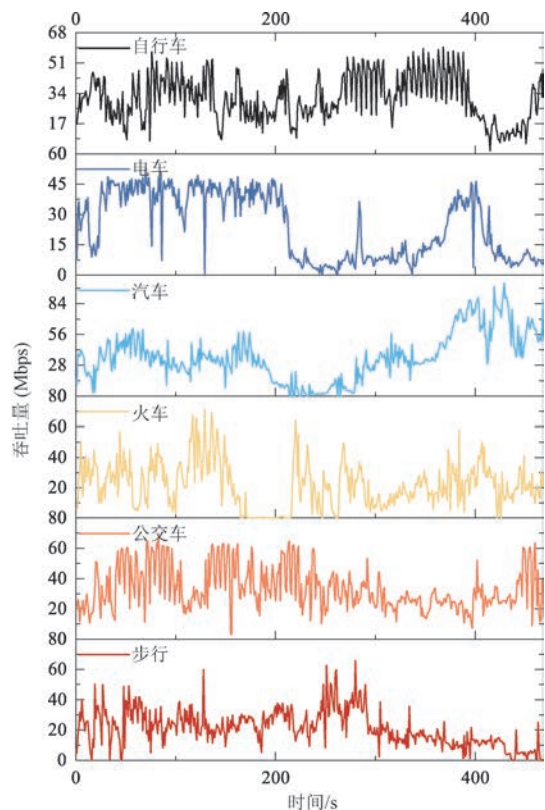


图7 六种交通方式的可用带宽感知结果

速度和途经地点影响显著。特别是火车在行驶过程中,其带宽波动幅度远超其他五种交通方式。以汽车为例,在 193 秒至 273 秒的时间段内,汽车驶入大型建筑物、隧道等基站信号覆盖不佳区域,网络环境受限,导致感知到的可用带宽不仅处于较低水平,还出现了显著的下降峰值。

6.3.1 视觉焦点区域视频压缩的性能评估

评估不同压缩方法对焦点区域 SSIM 的影响。实验中,将本文的焦点区域特征压缩方法与传统的 HEVC 视频编码方法^[8]以及基于鲁棒局部兴趣区域(robust local Region of Interest, rIROI)^[9]的视频压缩算法进行了比较。分别使用本文的特征提取方法和 rIROI 提取特征像素,HEVC 编码所有像素的重要性相同。在压缩精度和压缩率变化的情况下

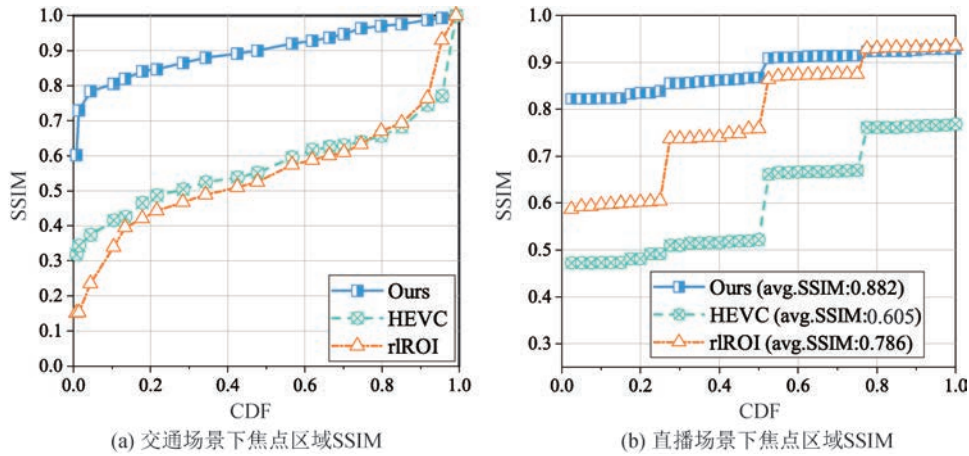


图 8 不同场景下焦点区域 SSIM 的 CDF 曲线

在 Cityscapes 城市交通场景数据集上的实验结果表明,本文提出的基于视觉焦点区域的方法在 SSIM 指标上显著优于其他方法,即使视频码率仅为原始视频的 4%,仍然保持了视频的细节。图 8 (a)可以看出,本文方法的 SSIM 最小值始终稳定在 0.6 以上,而且 SSIM 值主要集中在 0.8 至 1.0 的高质量区间,表明其在焦点区域保持了较高的图像质量。相比之下,HEVC 方法^[8]和 rIROI^[9]方法的 SSIM 值大多分布在 0.4 至 0.7 之间,质量波动较大。这表明这两种方法在处理复杂场景时,难以在低比特率条件下保持稳定的视频质量。

为进一步验证所提方法在新型业务场景中的适用性,本文自建了贴近实际业务的直播数据集:2025 年 12 月 6 日~10 日于抖音平台随机录制了 4 段直播视频(886×1920,每段约 10 s,内容涵盖服装/篮球带货等场景),并从中均匀采样 40 帧用于测试。采用与前文一致的语义分割模型对每帧进行语义分割,将人物及其身前产品区域定义为焦点,其余为背

景,生成二值掩模。实验设置与 Cityscapes 保持一致:压缩精度固定为 20×20 网格,码率约为原始视频体积的 4%,以焦点区域 SSIM 为指标,对本文方法、HEVC^[8]以及 rIROI^[9]方法进行对比,得到图 8 (b)的 CDF 曲线。结果显示,本文方法在直播场景下仍显著优于对比方案:焦点区域平均 SSIM 达 0.882,且所有样本均高于 0.82;rIROI 与 HEVC 分别为 0.786 与 0.605。进一步地,在 10%分位点处,本文方法 SSIM 仍约为 0.84,而 rIROI 与 HEVC 仅约为 0.60 和 0.48,表明本文方法在最具挑战的帧中仍能显著减弱质量退化。综上,基于焦点区域的非均匀压缩在严格码率约束下可显著提升人物区域的重建质量,更契合直播等以关键区域高保真为核心诉求的传输场景。

评估原始图像与解码图像的 SSIM。其中,压缩精度控制下采样阶段网格中的单元数量,压缩率控制视频码率,即原始视频体积除以压缩率。

在城市交通场景下,本文采用 Cityscapes^[5]数据集开展对比实验。将图像分辨率均匀划分为 20×20 的网格,压缩精度设置为 20,总计 400 个空间单元;将视频码率设置为原始视频的 4%,即压缩比率为 25。该设置旨在平衡压缩效率与图像质量,确保实验结果能够反映算法在实际应用中的性能。焦点区域由车辆、行人等语义类别组成,本文方法利用语义标签生成焦点重要性图,并通过轴对齐空间变换实现非均匀采样;对比的 rIROI 方法基于鲁棒局部显著性图确定兴趣区域。图 8 给出了不同场景不同方法在焦点区域 SSIM 上的 CDF 曲线。

为了深入评估所提出的多流自适应传输策略在实际网络环境中的性能表现,本文在六种典型的网络带宽波动条件下开展了系统性实验。实验中选取一个大

6.3.2 多流传输性能评估

小为 20 MB 的 MP4 视频作为原始输入,结合特征压缩视频流压缩率 $\mathcal{R}$,构建符合实际应用需求的视频传输模拟场景。为全面比较不同传输策略的性能,本文选取 5 种具有代表性的基线方法作为对比:固定比特率传输

方法、mDASH^[26]方法、缓存自适应传输^[23]方法、面向人类视觉的双流传输策略以及基于深度强化学习的自适应传输方法(ALVS-DRL)^[33],并与本文所提人机融合多流自适应传输策略进行对比,如图 9 所示。

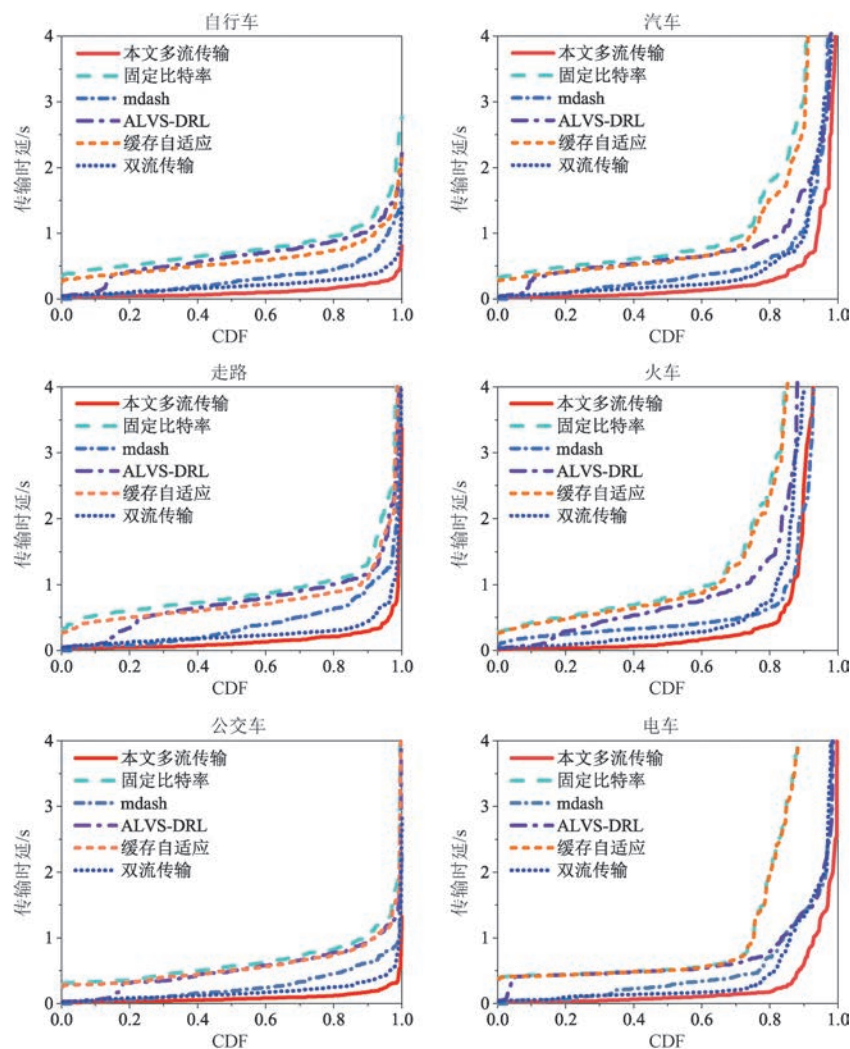


图9 本文方法与其他方法的传输时延对比图

实验结果表明,本文方法的时延分布整体处于右下方低时延区间,表明在多数场景下本文方法相比其他方法的平均传输时延最低。结合图 10 的平均传输时延结果可以看出,本文方法相较固定比特率的平均时延降低 85.4%,相较 mDASH 平均时延降低了 44%,表明所提方法在自适应流媒体场景下具有更高的传输效率。相较缓存自适应传输方法平均时延降低 83.9%,相较基于深度强化学习的 ALVS-DRL 方法平均时延降低 70.9%,相较双流传输方法平均时延降低 50.4%,表明在相同链路条件下,多流协同在降低平均时延方面均优于单流自适应与深度学习型调度,同时也体现了机器视觉结构化数据在未来视频传输领域的重要趋势。

进一步地,图 10 对六种交通模式的平均时延对比表明,本文方法在多数轨迹下取得最低或接近最低的平均时延,其中电车场景优势最显著,降低 57%至 86%;在图 10(a)的 3 种轨迹中,尽管时延差距有所缩小,但本文方法依然保持稳定的低时延水平。另外,在火车场景中,尽管由于网络波动较大和其他复杂因素的影响,时延表现稍逊于其他场景,但本文方法依然能够显著降低时延,表现出良好的性能。综上所述,本文提出的人机融合多流自适应传输方法在多种带宽波动模式下均表现出较好的传输效率和时延稳定性。与其他基线相比,联合普通压缩流、特征流与语义流的协同调度,有助于在保证人类视觉感知和机器视觉任务性能的同时,进一步降低整体传输时延。

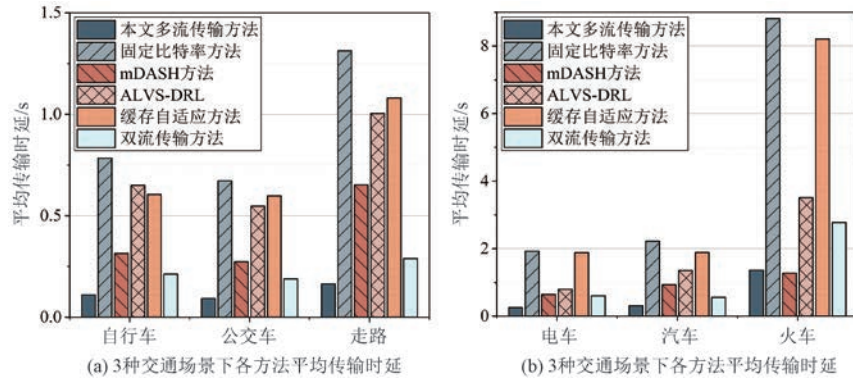


图 10 六种交通场景下不同方法平均时延对比

6.4 扩展实验与讨论

6.4.1 MinMaxP1 在 5G 视频会议直播场景下的性能分析

评估在强时延敏感业务下的最坏情形性能。在与 P1 保持一致的多流结构与 SSIM 约束下,仅将优化目标由平均时延替换为跨场景最大时延最小化,以显式抑制尾部与极端时延。实验选取 5G Traffic Datasets^[7]中 Zoom1 会议直播带宽轨迹,按 1 s 统计带宽 B ,窗口 6 s,场景划分沿用第 5 章 $\theta=(\bar{B},\Delta B)$ 。为保证可比性,三类数据流归一化数据量均设为 1,其中 s_1 以 20 MB HEVC 为基准, s_2 由 $\mathcal{R}$ 给定, s_3 取

等效压缩率 74^[34]。 δ_B 通过网格搜索确定:在场景比例约束(约 7 : 1.5 : 1.5 且各类 $\geq 5\%$)下,选择使 MinMaxP1 相对 P1 的 P95 改善最大者。实验结果如图 11 与表 1 所示, $\delta_B = 13.226$ Mbps 时 MinMaxP1 的平均和最大时延分别降至 0.0314 s 和 0.1780 s,较 P1 的平均和最大时延分别降低 86.5% 和 89.5%,且 P95 提升 9.1 倍。从分布角度看(图 11),P1 有 6.95% 的窗口时延超过 1 s,而 MinMaxP1 基本降至 0,时延主要集中在 0~0.1 s 区间。结果表明,MinMaxP1 在不改变约束与决策空间的前提下显著削弱长尾与极端时延,更适用于会议直播等尾部敏感场景。

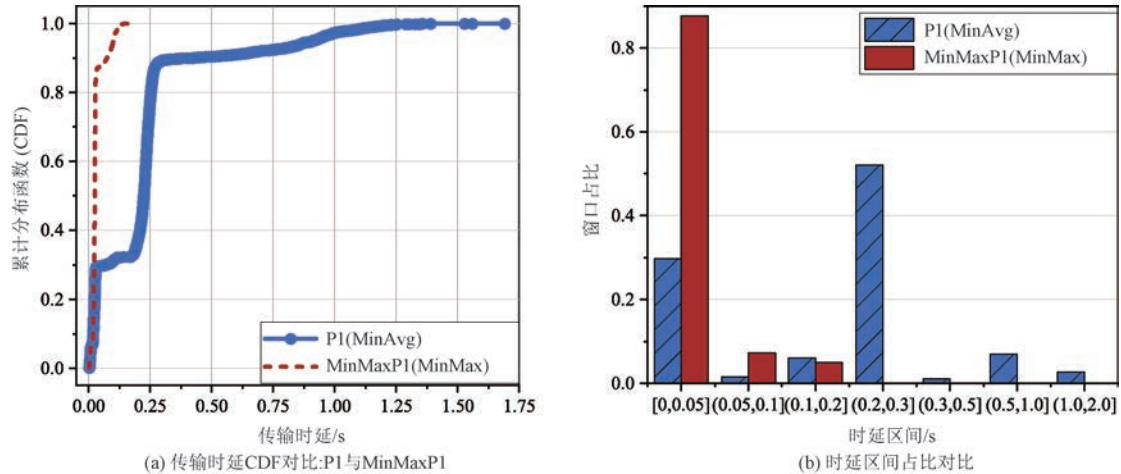


图 11 Zoom1 轨迹下 P1 与 MinMaxP1 的传输时延对比

表 1 平均、P95 及最大时延对比

优化策略	平均时延	P95 时延	最大时延
P1	0.232 s	0.910 s	1.691 s
MinMaxP1	0.031 s	0.100 s	0.178 s

6.4.2 5G 会议直播场景下的适用性分析

验证在视频会议直播等新型业务下的适用性。本文选取 5G Traffic Datasets^[7]中 Zoom1 真实带宽轨迹,按 7 : 1.5 : 1.5 比例抽取得到约 9033 个带宽样本,以 6 s 窗口聚合得到窗口级带宽及波动指标;

在统一的 19.6 MB 数据量与一致的时延计算口径下,对比固定比特率、缓存自适应、mDASH、ALVS-DRL 及双流传输等基线。阈值通过网格搜索确定,将平均时延和比例偏差惩罚作为优化目标选择最优阈值:双流基线最优 $\delta_B = 1.3449$,本文方法最优 $\delta_B = 10.4265$, $k_B = 1.20$, $k_D = 2.12$ 。图 12 右图结果显示,本文方法在平均时延与 P95 时延上均最优:平均传输时延为 0.28839 s,较各基线分别降低 63.57%、43.66%、52.63%、54.29% 和 31.82%。P95 时延

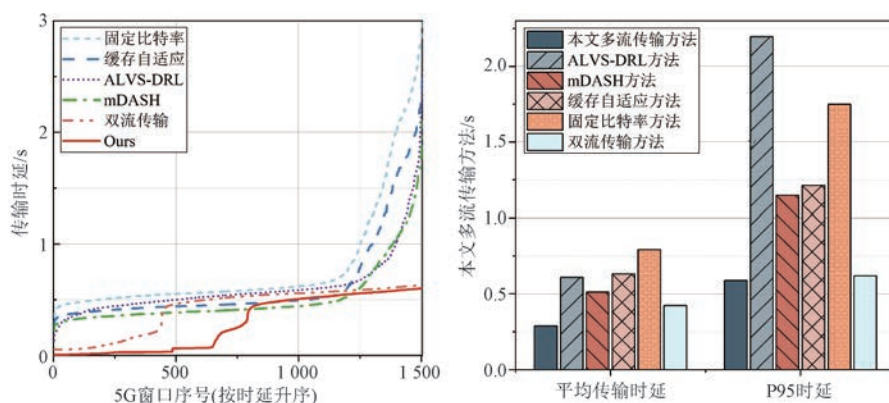


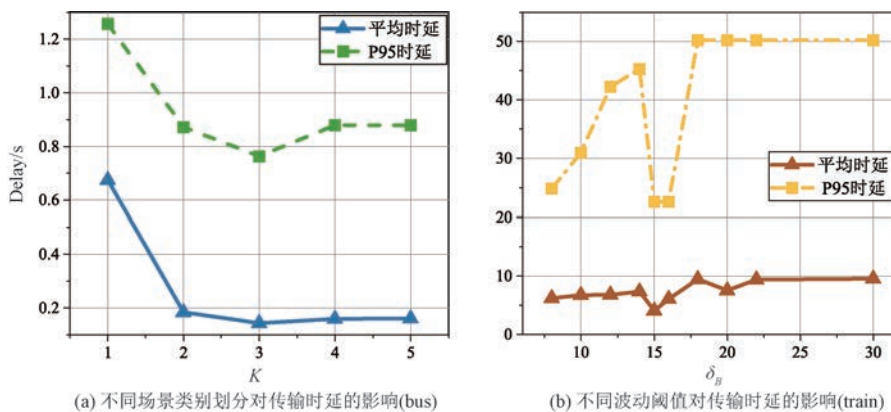
图 12 5G 场景下传输时延对比效果

为 0.5887 s,亦显著低于对应基线。左图可以直观看出,本文方法在全分位段保持更低且更平滑的时延分布,并抑制高分位尾部上翘。

6.4.3 场景离散化消融实验

为量化场景离散化粒度对传输时延的影响,在固定阈值配置($\delta_B = 15, k_B = 2, k_D = 2$)与统一压缩率候选集合条件下,本文以 bus 轨迹为代表对场景类别数 $K \in \{1, 2, 3, 4, 5\}$ 进行消融,并统计平均时延和 P95 时延。具体地, $K = 1$ 不区分场景(全 HEVC); $K = 2$ 将链路划分为常规/干扰两类(常规采用 HEVC 和特征流混合传输,干扰仅传特征流); $K = 3$ 在 $K = 2$ 的基础上显式引入恶劣场景并仅传语义流,以在极端低带宽与高波动条件下优先保证

任务可用性; $K = 4$ 和 $K = 5$ 则进一步细分干扰和常规场景并分配差异化的传输策略。结果如图 13(a)所示, $K = 3$ 在两项指标上均取得最优(平均 0.14 s、P95 0.76 s),相较 $K = 2$ (平均 0.18 s、P95 0.87 s)分别降低约 22.7% 与 12.4%。上述现象表明,三类划分在复杂度可控的前提下,通过对高拥塞区段触发仅语义流的强降载策略有效抑制队列峰值;当 $K \geq 4$ 时,平均时延维持在约 0.16 s, P95 时延相较 $K = 3$ 反而上升约 15%。这说明更细粒度的划分在当前阈值与压缩率集合下难以带来稳定增益,反而可能导致 P95 回升并引入额外策略复杂度。因此,本文将{常规,干扰,恶劣}作为默认场景离散化配置。



(a) 不同场景类别划分对传输时延的影响(bus)

(b) 不同波动阈值对传输时延的影响(train)

图 13 不同场景类别和不同波动阈值 δ_B 下的传输时延对比

6.4.4 场景判决阈值超参数选择

系统包含波动阈值 δ_B 、带宽均值阈值系数 k_B 、恶劣场景判决阈值 k_D 三个判决超参数。本文在 LTE 轨迹^[6]上进行超参数选择实验,并以平均时延与 P95 时延作为评价指标选取默认值。

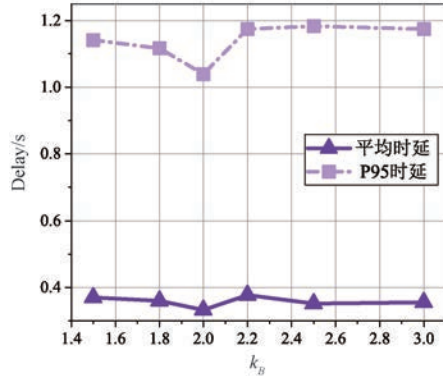
δ_B : 为避免 δ_B 纯经验设定,本文选取带宽波动幅度最强的 train 轨迹,以 10 s 滑动窗计算 ΔB ,扫描 $\delta_B \in \{8, 10, 12, 14, 15, 16, 18, 20, 22, 30\}$ 下运行多数数据流调度,统计平均 P95 时延(图 13(b))。

$\delta_B < 12$ Mbps 会导致恶劣场景过判,语义流占比过高, P95 升至约 30~45 s; $\delta_B \geq 18$ Mbps 会低估波动,在不稳定段仍频繁采用原始/低压缩特征流, P95 升至约 50 s。14~16 Mbps 区间最稳定,其中 $\delta_B = 15$ Mbps 取得最低的平均时延 4.03 s 与较低 P95 时延 22.6 s,故取 $\delta_B = 15$ Mbps。

k_B : 为避免轨迹偏置,选取 bicycle 轨迹为代表 $k_B \in \{1.5, 1.8, 2.0, 2.2, 2.5, 3.0\}$ (图 14(a))。 $k_B = 2$ 时,可同时最小化平均时延(约 0.33 s)与

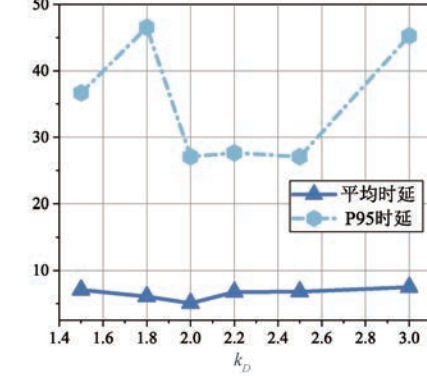
P95(约 1.04 s), k_B 偏小会增加常规场景与原始流占比, 偏大则将更多窗口推入干扰/恶劣并放大低带宽段排队, 均使 P95 回升。故取 $k_B=2$ 。

k_D : 依旧选取带宽波动幅度最强的 train 轨迹并固定 $\delta_B=15$ Mbps, $k_B=2$, 扫描 $k_D \in [1.5, 3.0]$



(a) 带宽均值阈值系数设定(bicycle)

(图 14(b))。过小使恶劣场景过判、P95 增大; 过大使低带宽窗口误判为干扰而仅传特征流, 长尾加重。 $k_D=2$ 时平均和 P95 时延同时最小(约 5.1 s、27 s), 实现了恶劣与干扰场景的较优平衡。故取 $k_D=2$ 。



(b) 恶劣场景判决阈值设定(train)

图 14 k_B, k_D 阈值设定依据

6.4.5 边缘侧高并发接入下的计算开销与排队时延

为了评估所提多流传输框架在资源受限边缘侧的计算开销与大规模并发部署下的排队时延, 本文在 Jetson Nano 上实测三类数据流的生成耗时, 窗口长度设为 6 s, 每窗选取关键帧作为代表帧, 其处理时间记为服务时间 S 。实测: s_1 编码耗时约 0.4909 s, s_3 语义推理耗时约 0.0541 s, 特征流 s_2 生成以 s_3 输出掩模为输入, 边侧开销计为 $S_{s_2} = S_{s_3} + S_{\text{codec}}$, 其中 S_{codec} 表示在掩模已给定条件下的特征流生成处理耗时(约 0.0104 s), 主要包括焦点区域网格/权重计算、非均匀下采样与重建等。故 $S_{s_2} \approx 0.0645$ s。此外, 多流决策与压缩率求解开销极小: 分类和选流的耗时仅为 0.0031 ms, 拉格朗日对偶

迭代求解 r_s 的平均/P95 时延分别为 0.2078 ms/0.2197 ms, 总计约 0.211 ms。相比生成耗时低 2~3 个数量级, 可忽略不计, 不影响排队仿真结论。

仿真中假设每路摄像头每 6 s 产生一个窗口任务, 并发路数取 $N \in \{1, 4, 8, 12, 16, 20, 24\}$, 分别构造固定比特率基线(每窗均执行 s_1 编码)与所提方法的任务序列, 统计排队等待时延 W_q 与边侧总时延 $W = W_q + S$ 。如图 15 所示, 随 N 增大, 基线在中高并发下出现显著排队累积与长尾; 本文方法因窗口级服务时间更短且可切换轻量分支而有效抑制队列拥塞: 例如 $N=16, 20$ 时基线的 P95 排队时延已达数百秒量级, 而本文方法仍为个位数秒; 在更高并发 $N=24$ 时长尾增长更缓, 表明在边缘端具备更好的并发可扩展性。

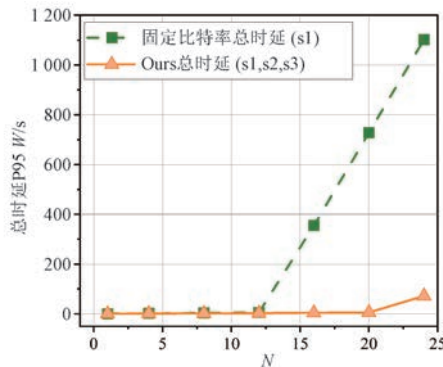
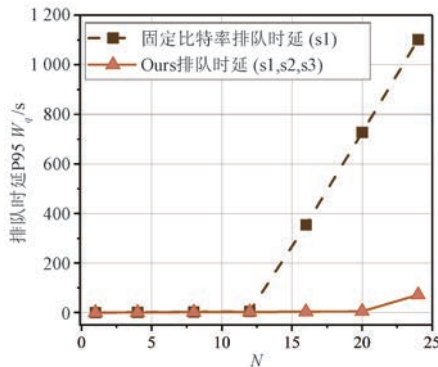


图 15 边缘节点高并发下的排队时延与总时延(P95)

6.4.6 不同拟合函数对 SSIM 建模的影响分析

本节验证 SSIM 建模的合理性, 并分析不同拟合函数对回归结果及后续优化建模的潜在影响。本

文在 Cityscapes 数据集上设置中等网格压缩精度 18(兼顾形变建模精度与边缘端可部署的求解/重建开销), 统计不同压缩率 $r_{s,*} \in \{2, 4, 6, 8, 10, 12, 14\}$

下焦点区域 SSIM,并分别采用线性、二次/三次、对数、指数、幂律与倒数等函数进行回归拟合,拟合优劣采用决定系数 R^2 、均方根误差(RMSE)以及小样本修正的赤池信息准则(AICc)综合评估,并辅以残差分布进行对比(见图 16、表 2)。结果表明,多种函数均能刻画 SSIM 随压缩率增大而下降的趋势,但不同函数在局部拟合能力上存在显著差

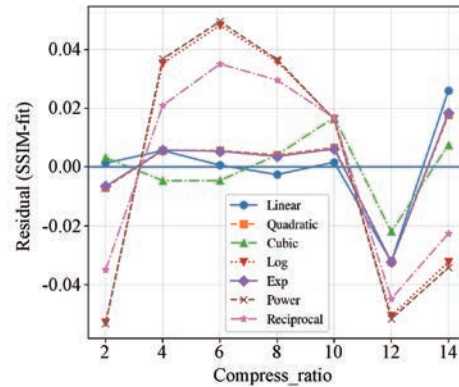
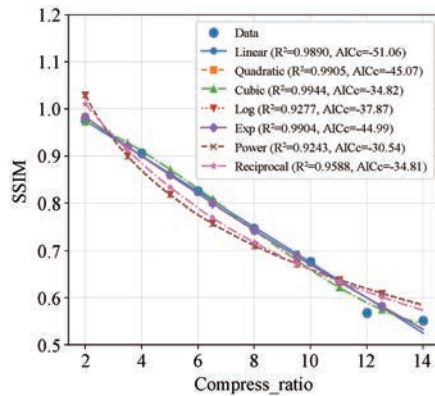


图 16 拟合曲线(左)与残差(右)对比图

表 2 整体拟合指标

拟合函数	RMSE	R^2	AICc
线性 (Linear)	0.01580	0.98901	-51.061
二次 (Quadratic)	0.01471	0.99049	-45.068
指数 (Exp)	0.01479	0.99038	-44.991
对数 (Log)	0.04057	0.9276	-37.865
三次 (Cubic)	0.01125	0.99443	-34.819
幂 (Power)	0.04151463	0.92426431	-30.543
倒数 (Reciprocal)	0.0306	0.95884	-34.813

7 结 论

针对动态复杂场景下视频传输面临的高时延、带宽受限与剧烈波动等挑战,本文提出了一种基于场景感知的端边云多数据流协同优化传输方法。与依赖集中式云架构的传统单流传输及整帧压缩方式不同,本方法在边缘侧构建多流生成机制,将原始视频数据划分为三类互补的数据形态:普通压缩流、焦点区域特征压缩流与结构化语义流,以适应多样化的网络条件与任务需求。在此基础上,本文将多流自适应选择与压缩率配置问题建模为一个非凸组合优化问题,并提出一种基于场景感知状态划分的联合优化机制,显著降低了求解维度与计算复杂度。具体而言,场景感知状态划分模型将动态网络环境划分为常规、干扰与恶劣三种典型状态,以此限定候选数据流类型与压缩率范围;在协同优化阶段,则采用权重驱动策略结合拉格朗日对偶方法,实现对数

异:三次函数虽取得更高 R^2 ,但在小样本条件下参数更多、AICc 较差,存在过拟合风险;线性函数 AICc 最优但拟合误差相对更大。综合拟合误差、模型复杂度与残差稳定性,二次函数在精度与复杂度之间提供更优折中,残差更接近零且波动更稳定。因此本文在优化约束建模中优先采用二次近似函数。

据流类型与压缩率的自适应调节,在维持视频质量的同时,有效降低传输时延。实验结果表明,本文方法在智能交通和直播等场景下均能有效保持视觉焦点区域质量,并在 4G/LTE 和 5G 等真实带宽环境中显著降低平均传输时延。总体而言,本文方法能够在复杂动态网络条件下兼顾关键区域质量保持与低时延传输,体现出良好的场景适应性和传输效率优势。未来工作将进一步拓展该方法在多场景应用中的适应性,并结合实际系统部署,不断提升其实时性与实用性。

参 考 文 献

- [1] Revelations in the Global DataSphere, 2024: Key Trends and Takeaways(IDC), <https://www.idc.com/getdoc.jsp?containerId=prCHC52667624> 2024,9
- [2] GIV 2025: Unfolding the Industry Blueprint of an Intelligent World (GIV), https://www-file.huawei.com/-/media/corp2020/pdf/giv/historyreports/whitepaper_en_2018.pdf 2018,6,21
- [3] Gao Wen, Tian Yong-Hong, Wang Jian. Digital retina: Key links in the evolution of smart city systems. Science China: Information Sciences, 2018,48(8):1076-1082(in Chinese)
(高文,田永鸿,王坚. 数字视网膜:智慧城市系统演进的关键环节. 中国科学:信息科学,2018,48(8):1076-1082)
- [4] Turkkan B O, Dai T, Raman A, et al. GreenABR: Energy-aware adaptive bitrate streaming with deep reinforcement learning//Proceedings of the 13th ACM Multimedia Systems

- Conference. Athlone, Ireland, 2022:150-162
- [5] Cordts M, Omran M, Ramos S, et al. The Cityscapes dataset for semantic urban scene understanding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016:3213-3223
- [6] Van Der Hooft J, Petrangeli S, Wauters T, et al. HTTP/2-based adaptive streaming of HEVC video over 4G/LTE networks. *IEEE Communications Letters*, 2016, 20 (11): 2177-2180
- [7] Choi Y H, Kim D, Ko M, et al. ML-based 5G traffic generation for practical simulations using open datasets. *IEEE Communications Magazine*, 2023, 61(9):130-136
- [8] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668
- [9] Tan W, Yan B, Li K, et al. Image retargeting for preserving robust local feature: application to mobile visual search. *IEEE Transactions on Multimedia*, 2015, 18(1):128-137
- [10] Agrawal P, Zabrovskiy A, Ilangoan A, Timmerer C, Prodan R. FastTTPS: fast approach for video transcoding time prediction and scheduling for HTTP adaptive streaming videos. *Cluster Computing*, 2021, 24(3):1605-1621
- [11] Pintye I, Kovács J, Lovas R. Enhancing machine learning-based autoscaling for cloud resource orchestration. *Journal of Grid Computing*, 2024, 22(4):68
- [12] Kumar T, Sharma P, Tanwar J, Alsghier H, Bhushan S, Alhumyani H, Sharma V, Alutaibi A-I. Cloud-based video streaming services: Trends, challenges, and opportunities. *CAAI Transactions on Intelligence Technology*, 2024, 9(2): 265-285
- [13] Huang S, Xie J, Muslam M M A. A cloud computing based deep compression framework for UHD video delivery. *IEEE Transactions on Cloud Computing*, 2022, 11(2):1562-1574
- [14] Moina-Rivera W, Garcia-Pineda M, Gutiérrez-Aguado J, et al. Cloud media video encoding: Review and challenges. *Multimedia Tools and Applications*, 2024, 83(34):81231-81278
- [15] Ran Li-Qiong, Weng Yue-Nan, Xu Xiao-Ling, et al. A review of cloud, edge and fog computing security. *Chinese Radio Engineering*, 2023, 53(2):247-261(in Chinese)
(冉黎琼, 翁越男, 徐小玲等. 云、边缘和雾计算安全性研究综述. *无线电工程*, 2023, 53(2):247-261)
- [16] Khan M A, Baccour E, Chkibene Z, Erbad A, Hamila R, Hamdi M, Gabbouj M. A survey on mobile edge computing for video streaming: Opportunities and challenges. *IEEE Access*, 2022, 10:120514-120550
- [17] Jedari B, Premsankar G, Illahi G, Di Francesco M, Mehrabi A, Ylä-Jääski A. Video caching, analytics, and delivery at the wireless edge: A survey and future directions. *IEEE Communications Surveys & Tutorials*, 2020, 23(1):431-471
- [18] Maia A, Boutouchent A, Kardjadjia Y, Gherari M, Soyak E G, Saqib M, Boussekar K, Cilbir I, Habibi S, Ouledsidi Ali S, Ajib W, Elbiaze H, Erçetin O, Ghamri-Doudane Y, Gliitho R. A survey on integrated computing, caching, and communication in the cloud-to-edge continuum. *Computer Communications*, 2024, 219:128-152
- [19] Stockhammer T. Dynamic adaptive streaming over HTTP-standards and design principles//Proceedings of the Second Annual ACM Conference on Multimedia Systems. San Jose, USA, 2011:133-144
- [20] Go Y, Kwon O C, Song H. An energy-efficient HTTP adaptive video streaming with networking cost constraint over heterogeneous wireless networks. *IEEE Transactions on Multimedia*, 2015, 17(9):1646-1657
- [21] Thang T C, Ho Q D, Kang J W, et al. Adaptive streaming of audiovisual content using MPEG DASH. *IEEE Transactions on Consumer Electronics*, 2012, 58(1):78-85
- [22] Huang T Y, Johari R, McKeown N, et al. A buffer-based approach to rate adaptation: evidence from a large video streaming service//Proceedings of the 2014 ACM Conference on SIGCOMM. New York, USA, 2014:187-198
- [23] Xu X, Liu J, Tao X. Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation. *IEEE Access*, 2017, 5:16406-16415
- [24] Zhang X, Tian J, Zhang J, Xiang C. Fine-grained caching and resource scheduling for adaptive bitrate videos in edge networks. *ACM Transactions on Sensor Networks*, 2023, 19(4):1-30
- [25] Bayhan S, Maghsudi S, Zubow A. EdgeDASH: Exploiting network-assisted adaptive video streaming for edge caching. *IEEE Transactions on Network and Service Management*, 2020, 18(2):1732-1745
- [26] Zhou C, Lin C W, Guo Z. mDASH: A Markov decision-based rate adaptation approach for dynamic HTTP streaming. *IEEE Transactions on Multimedia*, 2016, 18(4):738-751
- [27] Akhtar Z, Nam Y S, Govindan R, et al. Oboe: Auto-tuning video ABR algorithms to network conditions//Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. Budapest, Hungary, 2018:44-58
- [28] Meng Z, Chen J, Guo Y, et al. PiTree: Practical implementation of ABR algorithms using decision trees//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France, 2019:2431-2439
- [29] Huang T, Zhou C, Zhang R X, et al. Stick: A harmonious fusion of buffer-based and learning-based approach for adaptive streaming//Proceedings of IEEE INFOCOM 2020-IEEE Conference on Computer Communications. Virtual, 2020: 1967-1976
- [30] Turkkan B O, Dai T, Raman A, et al. GreenABR+: Generalized energy-aware adaptive bitrate streaming. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024, 20(9):1-24
- [31] Souane N, Bourenane M, Douga Y. Deep reinforcement learning-based approach for video streaming: Dynamic adap-

- tive video streaming over HTTP. *Applied Sciences*, 2023,13(21):11697
- [32] Ma X, Li Q, Zou L, et al. QAVA: QoE-aware adaptive video bitrate aggregation for HTTP live streaming based on smart edge computing. *IEEE Transactions on Broadcasting*, 2022,68(3):661-676
- [33] Ozcelik I M, Ersoy C. ALVS: Adaptive Live Video Streaming using deep reinforcement learning. *Journal of Network and Computer Applications*, 2022,205:103451
- [34] Wang Y, Xu J, Ji W. A feature-based video transmission framework for visual IoT in fog computing systems//*Proceedings of the 2019 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*. Cambridge, UK, 2019:1-8
- [35] Wang Y, Xu J, Ji W. An optimal coverage model for the deployment of iot devices in feature-based video transmission systems//*Proceedings of the 2019 IEEE Visual Communications and Image Processing (VCIP)*. Sydney, Australia, 2019:1-4
- [36] Wang S, Yang J, Bi S. Adaptive video streaming in multi-tier computing networks: Joint edge transcoding and client enhancement. *IEEE Transactions on Mobile Computing*, 2023,23(4):2657-2670
- [37] Yang Z, Hu D, Guo Q, et al. Visual E 2 C: AI-driven visual end-edge-cloud architecture for 6G in low-carbon smart cities. *IEEE Wireless Communications*, 2023,30(3):204-210
- [38] Yang Z, Ji W, Wang Z. Adaptive joint configuration optimization for collaborative inference in edge-cloud systems. *Science China Information Sciences*, 2024,67(4):149103
- [39] Ji W, Liang B, Wang Y, Qiu R, Yang Z-M. Crowd V-IoE: Visual internet of everything architecture in AI-driven fog computing. *IEEE Wireless Communications*, 2020,27(2):51-57
- [40] Hikvision, H.264(5) & H.264(5)+Recommended Bit Rate at General Resolutions, <https://www.hikvision.com/content/dam/hikvision/ca/faq-document/H.2645-%26-H.2645-Recommended-Bit-Rate-at-General-Resolutions.pdf>, 2021
- [41] Liu L, Li H, Gruteser M. Edge assisted real-time object detection for mobile augmented reality//*Proceedings of the 25th Annual International Conference on Mobile Computing and Networking*. Los Cabos, Mexico, 2019:1-16
- [42] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany, 2018:801-818
- [43] Panozzo D, Weber O, Sorkine O. Robust image retargeting via axis-aligned deformation. *Computer Graphics Forum*, 2012,31(2pt1):229-236
- [44] Yin Hong-Bo. A survey on network available bandwidth measurement methods. *Computer Knowledge and Technology*, 2010,6(6):1329-1331(in Chinese)
(尹红波. 网络可用带宽测量方法综述. *电脑知识与技术*, 2010,6(6):1329-1331)
- [45] Song Y, Zhang C, Fang Y, Lin P. Revenue maximization in time-varying multi-hop wireless networks: A dynamic pricing approach. *IEEE Journal on Selected Areas in Communications*, 2012,30(7):1237-1245
- [46] Huang H, Guo S, Liang W, Wang K, Okabe Y. Coflow-like online data acquisition from low-earth-orbit datacenters. *IEEE Transactions on Mobile Computing*, 2019,19(12):2743-2760
- [47] Toffolo N, Montanari A, Campagnaro F, Zorzi M. Modeling acoustic channel variability in underwater network simulators from real field experimental data//*Proceedings of the IEEE OCEANS 2022 Conference*. Hampton Roads, USA, 2022:1-7
- [48] Chiang M, Low S H, Calderbank A R, et al. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 2007,95(1):255-312
- [49] Michalewicz Z, Schoenauer M. Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary Computation*, 1996,4(1):1-32



GUO Qi, Ph. D. candidate. Her current research interests include multimedia computing, end-edge-cloud systems, video transmission.

YANG Zhe-Ming, Ph. D. , assistant professor. His research interests include multimedia systems, distributed inference, and edge intelligence.

ZHAO Chang, M. S. candidate. His main research in-

terests include large language model inference acceleration and model compression.

JI Wen, Ph. D. , professor. Her research interests include intelligent processor design, multimedia systems, and artificial intelligence systems, covering industrial artificial intelligence, high-performance vision processors, AI processors, multimedia end-edge-cloud computing systems, visual coding and transmission, and low-carbon computing.

Background

Video transmission is essential for uploading data from edge devices to the cloud to support intelligent perception and decision-making. However, traditional End-Cloud architec-

tures rely heavily on compressed video streams to balance quality and bandwidth, which often leads to redundant transmission and high latency under constrained bandwidth and

fluctuating networks, limiting large-scale real-time applications.

This work bridges End-Edge-Cloud collaboration, multimedia networking, and machine vision. Current mainstream methods (e. g. , MPEG-DASH and buffer/bitrate adaptation) have improved visual quality but still face two major limitations: they fail to coordinate heterogeneous human-machine data demands across video, feature, and semantic streams, and they lack adaptability to highly dynamic or constrained network environments, resulting in increased latency and degraded responsiveness. To address these challenges, we propose a scene-aware collaborative optimization method for multi-stream transmission in End-Edge-Cloud systems, which adaptively schedules video, feature, and semantic streams based on bandwidth and scenario conditions to reduce redundancy and improve timeliness. Experiments show that the proposed method maintains the SSIM of visual focus regions predominantly above 0.8 in intelligent transportation scenarios and achieves an average focus-region SSIM of 0.882 in live-streaming scenarios, significantly outperforming other baselines. On a real 4G/LTE bandwidth dataset, it reduces average transmission latency by 44.0%—85.4%; in a live-streaming scenario constructed using a 5G bandwidth dataset, it further reduces average latency by 31.82%—63.57% and achieves the best P95 latency.

This work was supported by the National Key R&D Program of China (Grant No. 2023YFB4502805), the China Postdoctoral Science Foundation (Grant No. GZC20251091), and Jiangsu Key Laboratory of AI for Industries.

The significance of these projects lies in advancing End-Edge-Cloud collaborative transmission technologies to provide strong support for key national sectors such as security monitoring, intelligent traffic, and industrial inspection. The main contribution of this study is to address the challenge of multi-stream collaborative processing and transmission within the End-Edge-Cloud architecture, as part of the broader project “Integrated End-Edge-Cloud Intelligent Computing and Storage Collaborative Scheduling and Application Demonstration.” Our team has long focused on multimedia transmission and End-Edge-Cloud collaboration, with strong expertise in End-Edge-Cloud scheduling, and resource optimization. The team has played a leading role in End-Edge-Cloud architecture research and standardization and has been recognized with two National Science and Technology Awards and two Beijing Science and Technology Awards. In addition, we have conducted extensive studies on visual IoT systems, video compression and transmission, and edge intelligence, achieving notable results in double-stream models, adaptive compression, and multi-task perception fusion.