自适应视频流传输的转码质量预测

VigneshV Menonvignesh.menon@aau.at
Christian Doppler 实验室 ATHENA 信息技术研究所（ITEC）克拉根福阿尔卑斯-亚得里亚大学 克拉根福，奥地利

Reza Farahanireza.farahani@aau.at
Christian Doppler 实验室 ATHENA 信息技术研究所（ITEC）克拉根福阿尔卑斯-亚得里亚大学 克拉根福，奥地利

Prajit T Rajendranprajit.thazhurazhikath@cea.fr
CEA, List，F-91120 帕莱索 信息技术研究所（ITEC）巴黎-萨克雷大学 巴黎，法国

Mohammad Ghanbari
ghan@essex.ac.uk 计算机科学与电子工程学院 埃塞克斯大学 科尔切斯特，英国

Hermann Hellwagner
hermann.hellwagner@aau.at
Christian Doppler 实验室 ATHENA 信息技术研究所（ITEC）克拉根福阿尔卑斯-亚得里亚大学 克拉根福，奥地利

Christian Timmerer
christian.timmerer@aau.at
Christian Doppler 实验室 ATHENA 信息技术研究所（ITEC）克拉根福阿尔卑斯-亚得里亚大学 克拉根福，奥地利

arXiv:2304.10234v1 [cs.MM] 20 Apr 2023

摘要

近年来，视频流应用的蓬勃发展推动了对视频质量评估 （VQA）的需求。简化参考视频质量评估（RR-VQA）是视频质量评估的一类，其中会提供原始视频的某些特征（例如纹理、边缘）用于质量评估。它是一个热门研究领域，适用于社交媒体、在线游戏和视频流等多种应用。本文提出了一种简化参考**转**码**质**量**预**测**模**型（ TQPM ），用于确定可能经过多个阶段转码的视频的视觉质量分数。该模型利用视频的离散余弦变换（ DCT ）能量特征（即视频的亮度、空间纹理信息和时间活动）以及每个转码阶段的目标比特率表示来预测质量。为此，本文对问题进行了形式化，并提出了一种基于长短期记忆网络（ LSTM ）的质量预测模型。实验结果表明，对于单阶段转码，TQPM 对 PSNR、SSIM 和 VMAF 的预测得分分别为 0.83、0.85 和 0.87，平均绝对误差（MAE）分别为 1.31 dB、1.19 dB 和 3.01。此外，在两阶段转码场景中，得分分别为 0.84、0.86 和 0.91，MAE 分别为 1.32 dB、1.33 dB 和 3.25。此外，TQPM 处理 4s 片段的平均时间为 0.328s，使其成为一种适用于在线流媒体应用的实用视频质量评估方法。

CCS 概念

• **信息系统** →**多媒体流传输**。

关键词

视频质量评估；简化参考；转码；VMAF 预测；视频流传输

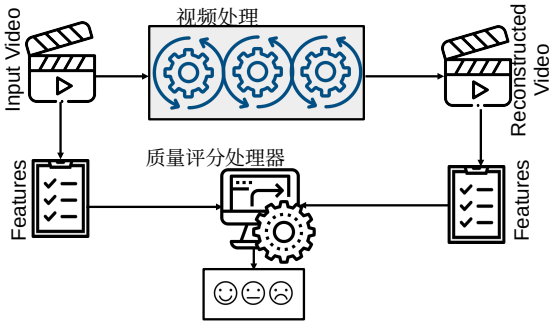
允许为个人或课堂使用而免费制作本作品的部分或全部数字副本或纸质副本，前提是不得为营利或商业利益而制作或分发副本，且副本首页须附有本声明及完整引文。必须尊重本作品中第三方组件的版权。其他用途请联系所有者/作者。*MHV '23，2023*年*5*月*7*日至*10*日，美国科罗拉多州丹佛 © 2023 版权所有归所有者/作者所有。ACM ISBN 979-8-4007-0160-3/23/05。
https://doi.org/10.1145/3588444.3591012

ACM 参考格式：
Vignesh V Menon、RezaFarahani、Prajit T Rajendran、Mohammad Ghanbari、Hermann Hellwagner 和 Christian Timmerer。2023年。《自适应视频流传输的转码质量预测》。发表于*Mile-High* 视频会议（*MHV '23*），*2023*年*5*月*7*日至*10*日，美国科罗拉多州丹佛。ACM，美国纽约州纽约市，7页。
https://doi.org/10.1145/3588444.3591012

1 引言

视频流应用对视频质量评估 （VQA）的需求日益增长。它在从采集到渲染的视频处理过程中发挥着重要作用，涵盖压缩、传输、复原和显示 [15]。在*HTTP*自适应流媒体 （HAS）中，[1], 面对各种可用的编码选项及其权衡，采用轻量且可靠的 VQA 方法至关重要。根据可用参考视频信号的信息量，VQA 方法分为全参考 （FR）、简化参考 （RR）和无参考 （NR）方法。NR-VQA 方法是“盲”的，不使用原始视频内容进行质量评估，因此 VQA 的可靠性较低 [15]。另一方面，由于 RR-VQA 方法（*i*）与基于 FR 的 VQA 方法相比所需的额外数据更少，且（*ii*）比基于 NR 的 VQA 方法更可靠，因此被用于实时场景 [6]。

最先进的 RR-VQA 方法的工作流程如图 1 所示。提取经过任意视频处理过程后，原始视频和重建视频的特征（例如，像素、相对熵或熵差 [5, 34], 频域特征，如 DCT [35]）。质量评分处理器（文献中大多采用基于机器学习的实现）会结合这些特征，预测最终的视频质量 [6]。由于峰值信噪比 (PSNR)仍是视频质量评估的事实上的行业标准，因此许多 RR-VQA 方法都旨在对其进行评估 [27, 29]。此外，还有一些方法用于预测结构相似性指数（SSIM） [36, 37],时空 RR 熵差（STRRED） [38], 和空间 RR 熵差（SRRED） [32] 等指标。然而，上述指标存在局限性，例如忽略了压缩伪影的时间特性 [16]。为弥补这些不足，视频多方法评估融合（VMAF）应运而生 [14]。VMAF 被提出作为一种 FR-VQA 模型，它结合感知质量特征来预测感知质量。为此，它纳入了



Figu图 1：当前最先进 RR-VQA 方法的工作流程 方法。

通过机器学习进行人类视觉建模，并提供对视频 QoE 的可接受预测 [16]。视频多方法评估融合是不同应用中优化编码决策的准则。例如，Orduna 等人 [28] 证明，无需任何特定训练或调整，视频多方法评估融合即可用于评估用户感知到的 360 度虚拟现实（VR）序列质量。Zadtootaghaj 等人 [40] 使用视频多方法评估融合分析在线视频游戏服务的视频质量，并计算最低编码码率，以显著降低不同流媒体游戏所需的带宽。Sakaushi等人 [30] 展示了一种视频监控系统，该系统使用视频多方法评估融合衡量不同码率下视频质量的下降程度。在 [19], 中，为码率阶梯选择使视频多方法评估融合达到最大值的优化码率-分辨率组合。在 [20, 25], 中，为码率阶梯选择能够最大化视觉质量和压缩效率、且经过感知优化的码率-分辨率组合。此外，在 [18], 中，为阶梯中的每个目标码率选择使视频多方法评估融合值最高的优化帧率。因此，视觉质量预测使服务器能够为码率阶梯选择优化的编码参数 [24]。

贡献：本文提出了一种面向视频流应用的简化参考转码质量预测模型（TQPM）。据我们所知，这是首项旨在预测多阶段转码的 *VMAF*的研究，尤其针对视频流应用：在视频被转码为目标比特率表示之前，视频片段需经过多个转码阶段。为此，首先从输入视频片段中提取基于DCT能量的特征，并将转码流程的信息（即每个阶段编码器的目标比特率表示）用作 VMAF 预测的简化参考。随后，仅针对输入视频片段执行特征提取。该方法不同于最先进的 RR-VQA 方法，后者需要从转码系统的输入和输出视频片段中提取特征。所提出模型的预测性能通过使用*AppleHTTPLive Streaming* （HLS）码率阶梯进行验证

[1] 使用 x2652 高效视频编码 [33] 开源编码器进行转码。

论文大纲：第 2 节介绍本文构建的 M阶段转码模型，第 3 节则讨论其架构

[1]https://developer.apple.com/documentation/http_live_streaming/ hls_authoring_specification_for_apple_devices，最后访问日期：2023 年 4 月 2 日。[2] https://www.videolan.org/developers/x265.html，最后访问日期：2023 年 4 月 2 日。

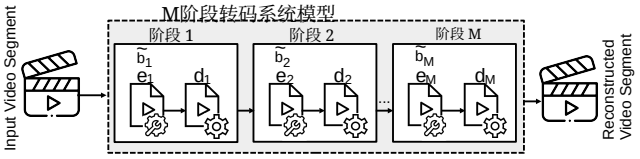


图2：本文所考虑的M阶段转码模型。其中， e_i 和 d_i 分别表示转码的 i^{th}阶段中的编码和解码，而 b_i 表示 e_i的目标比特率，其中 $i \in [1, M]$。

。第 4 节展示对 TQPM 性能的评估。最后，第 5 节总结全文。

2 M阶段转码模型

最近，视频转码已被视为一种常见解决方案，可在对延迟敏感的视频流应用中，于网络内服务器（部署在云端或边缘）重建视频序列 [7–10]。因此，本文提出了如图 2 所示的通用 HTTP自适应流媒体 M阶段转码模型，旨在应对以下场景：

(1) 单阶段转码：这是指用户接收由源服务器生成的比特流的场景。如图 3 所示，客户端 A 和 B 接收源站服务器生成的比特率表示。在此场景中，可以在源站服务器上进行视频质量评估，因为源视频片段和重建视频片段都可在源站服务器上获取。然而，使用最先进的视频质量评估方法时，必须完成编码过程才能确定视觉质量分数。此外，对输入和重建视频片段进行特征提取所需的时间（τ_f）也会增加延迟。

(2) 双阶段转码： 在这些应用中，边缘服务器上已有较高码率表示，并将其转码为用户请求的较低码率表示。如图 3 所示，客户端 C、D 和 E 分别接收 11.6 Mbps、5.8 Mbps 和 2.4 Mbps 的表示。边缘服务器将 16.8 Mbps 表示的视频片段转码为用户请求的表示。通过这种方式，源站服务器与边缘服务器之间的响应延迟和回传流量预计将会减少 [7]。在这种场景下，最先进的视频质量评估方法无法使用，因为*(i)* 目的地（客户端）没有可用作参考的原始输入视频片段，且*(ii)* 源站服务器上没有最终重建的视频片段。假设存在一种假想场景，原始视频片段和重建视频片段可以同时在源站服务器或目的地获取，那么总处理时间将包括两次编码和解码，以及对原始视频片段和重建视频片段进行特征提取。

涉及两台边缘服务器的三阶段转码场景也会出现。如图 2 所示，面向 HTTP自适应流媒体 的通用 M 阶段转码模型由一系列串联的 M 个编码器和 M 个解码器组成。M=1 转码对应单阶段转码，而 M=2 转码对应双阶段

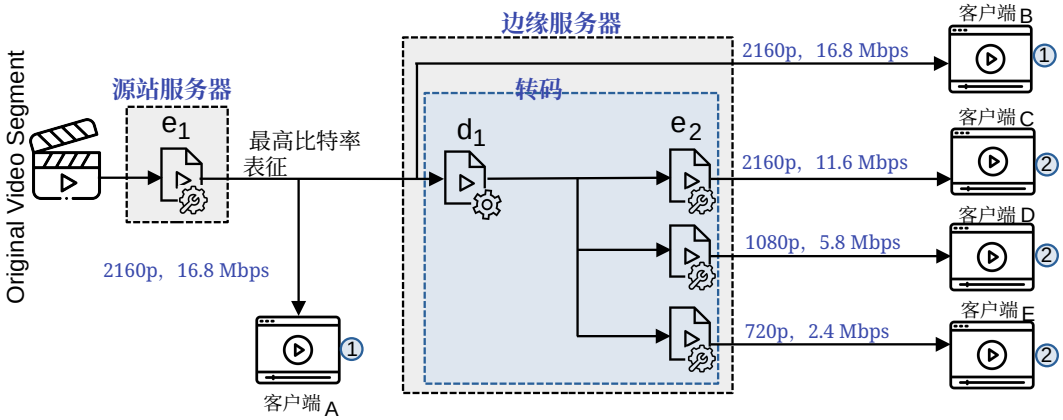


图3：自适应流媒体应用中的 VQA 示例场景。客户端 A 和 B 接收在源站服务器编码的码率阶梯中最高的码率表示（单阶段转码），而客户端 C、D 和 E 接收在边缘服务器转码的较低码率表示（双阶段转码）。

转码。如前所述，RR-VQA 在多阶段转码应用中的部署面临诸多问题。首先，利用输入视频片段和最终重建视频片段计算视频质量（τ_T）所需的总转码延迟非常高。这是因为 M 阶段转码过程中（M 次编码和 M 次解码）的编码和解码耗时，以及对输入和重建视频片段进行特征提取（τ_f）所需的时间，都会增加延迟。总转码延迟如公式（1）所示，其中 τ_{e_i} 和 τ_{d_i} 分别表示在 i^{th} 转码阶段进行编码和解码所需的时间。

$$\tau_T = \sum_{i=1}^{M}(\tau_{e_i} + \tau_{d_i}) + 2 \cdot \tau_f \quad (1)$$

其次，在大多数视频流应用中，确定视频多方法评估融合（VMAF）较为繁琐，因为（*i*）在目的端无法获得作为参考的原始输入视频片段；（*ii*）在源端无法获得最终重建的视频片段；（*iii*）缓慢的 VMAF 判定不适用于对延迟敏感的在线服务。通过输入视频片段特征和转码系统特征预测 VMAF，在源端进行视频质量评估（VQA），可以解决上述问题。

3 TQPM 架构

TQPM 架构如图4所示，包含三个步骤：

（1）输入视频片段特征描述（第 3.1 节）（2）转码模型特征描述（第 3.2 节）（3）视频质量预测（第 3.2 节）

选择低复杂度特征来表征输入视频片段，对于利用轻量级预测模型进行质量预测至关重要。高复杂度特征需要更庞大的模型（在模型大小和推理时间方面），从而增加预测延迟。提取最先进的空间信息（SI）和时间信息（TI）特征是计算密集型任务，并且与转码后的视频相关性不高。

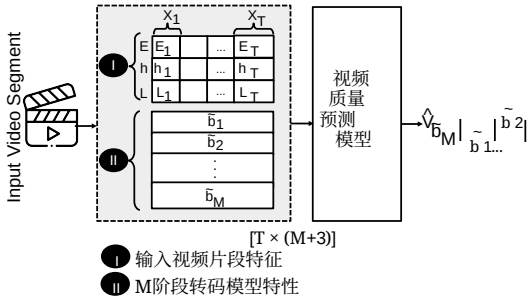


图 4：TQPM 架构

质量 [23]。本文采用一种轻量级、低延迟的方法，从输入视频片段中提取特征，具体如第 3.1 节所述。提取的特征以及转码过程中各阶段的编码目标比特率表示形式，即 $\tilde{b}_1, \tilde{b}_2, .., \tilde{b}_M$，用于预测视觉质量，预测指标包括峰值信噪比、结构相似性和视频多方法评估融合，具体讨论见第 3.2 节。

3.1 输入视频片段特征描述

使用视频复杂度分析器（VCA）[23] 开源软件提取的三个基于DCT能量的特征即（*i*）平均纹理能量 E、（*ii*）平均时间能量 h 和（*iii*）平均亮度 L，用作每个视频片段的简化参考。这些特征基于视频片段的亮度通道。由于大多数最先进编码器的码率控制不会考虑色度通道，因此所提出的方案不考虑色度通道。此外，视频质量评估指标（如视频多方法评估融合）对亮度通道的重视程度高于色度通道。这些特征基于我们之前的工作 [23]，此处将其纳入以使本文内容完整。首先，使用公式（2）计算每帧 p 中每个不重叠块 k 的纹理：

$$H_{p,k} = \sum_{i=0}^{w-1}\sum_{j=0}^{w-1} e^{|(\frac{ij}{w^2})^2 - 1|}|D(i,j)| \quad (2)$$

其中， $w \times w$像素是块的大小， D（i， j）是当 $i+j>0$时的（i， j）thDCT分量，否则为0 [11]。对纹理取平均值，以确定每帧的空间能量特征即E_p，如公式（3）所示：

$$E_p = \sum_{k=0}^{K-1} \frac{H_{p,k}}{K \cdot w^2} \tag{3}$$

其中， K 表示帧 p [22]中的块数。此外，计算每帧纹理能量与其前一帧纹理能量之间逐块绝对差之和（SAD），然后对每帧取平均，以获得每帧的时间能量特征（即， h_p），如式（4）所示：

$$h_p = \sum_{k=0}^{K-1} \frac{\mid H_{p,k} - H_{p-1,k} \mid}{K \cdot w^2} \tag{4}$$

每帧 p中不重叠块 k 的亮度定义如下：

$$L_{p,k} = \sqrt{DCT(0,0)} \tag{5}$$

其中， DCT(0,0)是 DCT 计算中的 DC分量。此外，每帧的块级亮度取平均值，记为 L_p，如式（6）所示。

$$L_p = \sum_{k=0}^{K-1} \frac{L_{p,k}}{K \cdot w^2} \tag{6}$$

将视频片段划分为 T个块，每块包含固定数量的帧（即， f_c）。计算每个块中 E、 h和 L特征的平均值，以获得输入视频片段的简化参考表示，表示为：

$$X = \{x_1, x_2, .., x_T\} \tag{7}$$

其中， x_i是每个 i^{th} 分块的特征集，表示为：

$$x_i = [E_i, h_i, L_i] \quad \forall i \in [1, T] \tag{8}$$

3.2 视频质量预测

为简单起见，假设 M阶段转码过程中编码器的设置（目标码率-分辨率对除外）均相同 [21]。码率阶梯中与目标码率对应的分辨率也假设为固定值。因此，转码模型可表征如下：

$$\tilde{B} = [\tilde{b}_1, \tilde{b}_2, .., \tilde{b}_M] \tag{9}$$

其中， $\tilde{b}_i$ 表示 e_i 编码器的目标比特率（参见图2）。˜请注意，将 B 追加到 x_i中，而 x_i是在输入视频片段表征阶段确定的，从而得到：

$$\tilde{x_i} = [x_i|\tilde{B}]^T \quad \forall \tilde{x_i} \in \tilde{X}, \quad i \in [1, T] \tag{10}$$

预测质量 $\hat{}v_{\tilde{b}_M|..|\tilde{b}_1}$ 可表示为：

$$\hat{v}_{\tilde{b}_M|..|\tilde{b}_1} = f(\tilde{X}) \tag{11}$$

长短期记忆网络模型通常用于时间序列预测应用，能够缓解长期预测中的关键问题，例如梯度消失或梯度爆炸 [39]。因此，本研究采用基于长短期记忆网络的预测模型 [12]。所述特征以维度为 [T×（M+3）的向量形式输入模型 [12]]，，其中 T表示视频片段中的块数。具体来说，序列 X 中的特征序列被输入长短期记忆网络模型，以预测相应输入视频片段以及转码过程中编码器链的视觉质量。

与真实质量之间可接受偏差的上限被认为是一个恰可察觉差异 (JND)，

$$|\hat{v} - v_G| < 1\text{JND} \tag{12}$$

其中，$\hat{}v$ 和 v_G 分别为预测质量和真实质量。本文根据当前行业惯例，将平均目标恰可察觉差异设为六个视频多方法评估融合分值[3]。

4 评估

本节首先介绍评估设置，然后展示实验结果。

4.1 评估设置

本文使用来自联合视频专家组 [2], MCML [4],SJTU [31], Berlin [3], UVG [26], BVI [17] 数据集的视频序列。使用 x265 v3.5[2]，并采用超快 预设，在配备 Intel Xeon Gold 5218R（80 个核心，频率为 2.10 GHz）的双处理器服务器上，以视频缓冲验证器 （VBV）速率控制模式按 30 fps 对序列进行编码。片段长度设为四秒。所考虑的五百个视频中，80% 用作训练数据集，其余 20% 用作测试数据集。实验中考虑的码率表示（$b_j \; \forall \; j \in [1, 12]$）用作每个转码阶段（$b_i \; \forall \; i \in [1, M]$）编码的目标˜码率，具体规定见 Apple HLS 创作规范[1]。 E、 h 和L 特征使用开源视频复杂度分析器 VCA v2.04 提取 [23]，该分析器以八个 CPU 线程运行；其中 w（参见 公式 2）设为 32。 f_c设为 15，即将视频片段划分为八个区块（T=8）。

对长短期记忆网络模型进行超参数调优，以获得最佳预测性能 [39]。长短期记忆网络单元数设为 50，使用 Adam 优化器 [13]，以 10^{-3} 的学习率训练模型 100 个周期。训练长短期记忆网络模型所用的损失函数为平均绝对误差（MAE）。针对每个测试序列，比较 M=1 （单阶段）和 M=2 （两阶段）转码的实际质量与预测质量，包括峰值信噪比、结构相似性和视频多方法评估融合 [14]。由于假定内容以最高分辨率显示（即2160p），因此将转码后的内容缩放（双三次插值）至 2160p 分辨率，以确定视觉质量。

4.2 实验结果

在第一项实验中，将 TQPM 的处理时间（即，τ_p）与当前最先进的 RR-VQA 方法中的总转码延迟 τ_T（参见Eq.1）进行比较。观察到 M=1 和 M=2 的平均 τ_T 分别为 1.92s 和 3.78s。对一个 4s 视频片段进行特征提取的平均耗时（$^{\tau}f$）为 0.323s。此外，长短期记忆网络模型的平均推理时间为 5 ms。因此，TQPM 对一个 4s 视频片段的平均处理时间为 0.328s。因此，TQPM

[3]https://streaminglearningcenter.com/codecs/finding-the-just-noticeable-difference-with-netflix-vmaf.html，最后访问时间：2023 年 4 月 2 日。[4]https://vca.itec.aau.at，最后访问时间：2023 年 4 月 2 日。

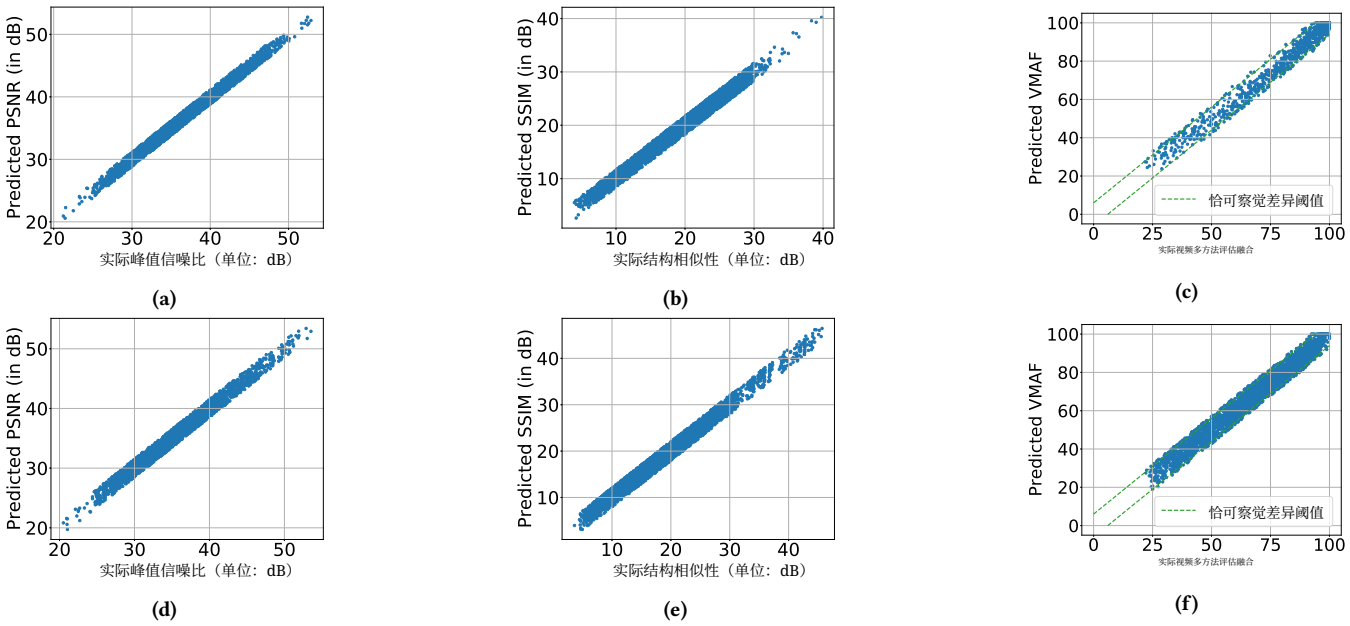


图5：M=1 （分别为(a)峰值信噪比、(b)结构相似性和(c)视频多方法评估融合）与M=2转码（分别为(d)峰值信噪比、(e)结构相似性和(f)视频多方法评估融合）的实际质量与预测质量散点图。

˜表1：本文考虑的使用x265高效视频编码编码器进行编码的 b_1 种表示在M=1 和M=2时TQPM的预测准确率。

			峰值信噪比预测				SSIM 预测				VMAF 预测			
			M=1		M=2		M=1		M=2		M=1		M=2	
$\tilde{b}_1$			R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
b_1	360p	0.145 Mbps	0.82	1.20 分贝	-	-	0.89	1.08 分贝	-	-	0.87	3.35	-	-
b_2	432p	0.300 Mbps	0.83	1.19 dB	0.84	1.37 dB	0.89	1.14 dB	0.87	1.34 dB	0.87	3.51	0.76	3.38
b_3	540p	0.600 Mbps	0.83	1.19 dB	0.85	1.28 dB	0.88	1.18 dB	0.85	1.21 dB	0.90	4.05	0.84	3.55
b_4	540p	0.900 Mbps	0.83	1.19 dB	0.83	1.22 dB	0.86	1.17 dB	0.86	1.11 dB	0.90	3.83	0.89	3.53
b_5	540p	1.600 Mbps	0.82	1.22 dB	0.82	1.15 dB	0.84	1.19 dB	0.85	1.38 dB	0.90	3.45	0.90	3.44
b_6	720p	2.400 Mbps	0.83	1.26 dB	0.83	1.28 dB	0.82	1.18 dB	0.83	1.57 dB	0.88	2.88	0.91	3.45
b_7	720p	3.400 Mbps	0.81	1.30 dB	0.85	1.23 dB	0.83	1.20 dB	0.82	1.35 dB	0.84	2.89	0.94	3.03
b_8	1080p	4.500 Mbps	0.84	1.28 dB	0.83	1.28 分贝	0.88	1.23 分贝	0.82	1.34 dB	0.87	2.28	0.95	3.03
b_9	1080p	5.800 Mbps	0.86	1.31 dB	0.87	1.42 dB	0.83	1.29 dB	0.86	1.30 dB	0.87	2.23	0.95	3.34
b_{10}	1440p	8.100 Mbps	0.84	1.39 分贝	0.81	1.41 dB	0.87	1.29 dB	0.87	1.32 分贝	0.85	2.73	0.96	2.96
b_{11}	2160p	11.600 Mbps	0.79	1.50 dB	0.82	1.31 dB	0.88	1.17 分贝	0.84	1.32 分贝	0.82	2.58	0.96	3.02
b_{12}	2160p	16.800 Mbps	0.84	1.49 dB	0.79	1.26 dB	0.88	1.19 dB	0.86	1.35 dB	0.86	2.38	0.96	2.99
平均值			**0.83**	**1.31 dB**	**0.84**	**1.32 dB**	**0.85**	**1.19 dB**	**0.86**	**1.33 dB**	**0.87**	**3.01**	**0.91**	**3.25**

其处理时间显著低于最先进的 RR-VQA 方法。

第二个实验评估了 M=1和 M=2 转码的预测质量分数与实际质量分数之间的相关性。如图5a、5b和5c以及图5d、5e和5f所示，预测的 PSNR、SSIM 和视频多方法评估融合分数分别与实际分数高度相关（例如，单阶段和双阶段转码的 VMAF 预测平均 R^2 分数分别为 0.87 和 0.91）。此外，预测误差小于可接受的一个恰可察觉差异（JND）阈值（即，六个 VMAF 点），这表明 TQPM 的准确度足够高。

在最后一项实验中，本文使用平均绝对误差（MAE）评估 TQPM 对˜本文所考虑的 b_1 码率表示在 M=1 和 M=2 转码中的预测性能。如表1所示，M=1 和 M=2 转码的 VMAF 预测平均 MAE 分别为 3.01 和 3.25。M=2 的结果对应于将 b_1 码率表示转码为码率阶梯中可能的较低码率表示时的平均视觉质量预测准确度˜。请注意，由于 b_1是码率阶梯中最低的码率表示，因此不存在与 $\tilde{b}_1 = b_1$ 对应的场景。观察发现，M=2的 R^2 分数会随着 $\tilde{b}_1$ 的增加而升高。这是因为存在...

随着 $\tilde{b}_1$ 增加，训练数据量也会增加（转码为较低码率表示）。

5 结论

本文提出了TQPM，一种面向视频流应用的在线转码质量预测模型。所提出的基于长短期记忆网络的模型使用基于DCT能量的特征作为简化参考来表征输入视频片段，用于预测M阶段转码过程的视觉质量。通过使用x265开源高效视频编码编码器对Apple HLS码率阶梯进行编码和转码，验证了TQPM的性能。对于单阶段转码，TQPM预测峰值信噪比、结构相似性和视频多方法评估融合的平均绝对误差分别为1.31 dB、1.19 dB和3.01。此外，对于双阶段转码，峰值信噪比、结构相似性和视频多方法评估融合的预测平均绝对误差分别为1.32 dB、1.33 dB和3.25。

在本文中，转码缩放和转码码率调整均被视为转码，即，假设码率阶梯表示所用的编码器/编解码器相同。未来将研究不同编解码器的码率阶梯表示之间的转码。另一个未来研究方向是基于所提出的模型，在端到端直播流媒体系统中定义一个决策组件。

6 致谢

谨此感谢奥地利联邦数字与经济事务部、国家研究、技术与发展基金会以及Christian Doppler研究协会提供的资金支持。Christian Doppler实验室ATHENA：https://athena.itec.aau.at/。

参考文献

<escape>[1] </escape>A. Bentaleb、B. Taani、A. C. Begen、C. Timmerer 和 R. Zimmermann。2019。关于 HTTP 流媒体比特率自适应方案的综述。*IEEE*通信调查与教程 21, 1 (2019), 562–585。 <escape>[2] </escape>Jill Boyce、Karsten Suehring、Xiang Li 和 Vadim Seregin。2018。JVET-J1010：JVET 通用测试条件和软件参考配置。 <escape>[3] </escape>B. Bross、H. Kirchhoffer、C. Bartnik、M. Palkow 和 D. Marpe。2020。AHG4多格式柏林测试序列。载于 *JVET-Q0791*。 <escape>[4] </escape>Manri Cheon 和 Jong-Seok Lee。2018。面向沉浸式体验的压缩4K超高清视频的主观与客观质量评估。*IEEE*视频技术电路与系统汇刊 28, 7 (2018), 1467–1480。https://doi.org/ 10 1109 2017 2683504 。/TCSVT. 。[5] Shyamprasad Chikkerur、Vijay Sundaram、Martin Reisslein 和 Lina Karam。2011。客观视频质量评估方法：分类、综述与性能比较。*IEEE*广播汇刊 57, 2 (2011), 165–182。https://doi.org/10.1109/TBC.2011.2104671<escape>[6] </escape>Shahi Dost、Faryal Saud、Maham Shabbir、Muhammad Gufran Khan、Muhammad Shahid 和 Benny Lovstrom。2022。简化参考图像与视频质量评估：方法综述。*EURASIP*图像与视频处理期刊 (2022)。https://doi.org/10.1186/s13640-021-00578-y <escape>[7] </escape>Reza Farahani。2021。用于 HTTP自适应视频流媒体的 CDN 和 SDN 支持及播放器交互。载于 第*12*届*ACM*多媒体系统会议论文集。398–402。 <escape>[8] </escape>Reza Farahani、Hadi Amirpour、Farzad Tashtarian、Abdelhak Bentaleb、Christian Timmerer、Hermann Hellwagner 和 Roger Zimmermann。2022。RICHTER：用于低延迟实时视频流传输的混合式P2P-CDN架构。载于 首届*Mile-High* 视频会议论文集。87–88。<escape>[9] </escape>Reza Farahani、Mohammad Shojafar、Christian Timmerer、Farzad Tashtarian、Mohammad Ghanbari 和 Hermann Hellwagner。2022。ARARAT：用于 HTTP自适应视频流媒体的协作式边缘辅助框架。*IEEE*网络与服务管理汇刊 (2022)。<escape>[10] </escape>Reza Farahani、Farzad Tashtarian、Christian Timmerer、Mohammad Ghanbar 和 Hermann Hellwagner。2022。LEADER：用于 HTTP自适应视频流媒体的协作式边缘与 SDN 辅助框架。载于 *ICC 2022-IEEE*国际通信会议。IEEE，745–750。<escape>[11] </escape>NB Harikrishnan、Vignesh V Menon、Manoj S Nair 和 Gayathri Narayanan。2017。图像压缩技术的比较评估。载于 *2017*年国际新兴技术算法、方法论、模型与应用国际会议（ICAMMAET）。IEEE，1–4。[12] Sepp Hochreiter 和 Jürgen Schmidhuber。1997。长短期记忆。Neural Comput. 9, 8（1997年11月），1735–1780。https://doi.org/10.1162/neco.1997.9. 8.1735[13] Diederik P. Kingma 和 Jimmy Ba。2015。Adam：一种随机优化方法。载于《第三届国际学习表征会议》（ICLR 2015），美国加利福尼亚州圣迭戈，2015年5月7日至9日。[14] Zhi Li 等。2018。VMAF：持续探索之旅。Netflix 技术博客 25（2018）。[15] Weisi Lin 和 C.-C. Jay Kuo。2011。感知视觉质量指标：一项综述。Journal of Visual Communication and Image Representation 22, 4（2011），297–312。https://doi.org/ 10 1016 2011 01 005 。/j.jvcir. . . Tsung-Jung Liu、Yu-Chieh Lin、[16] Weisi Lin 和 C.-C. Jay Kuo。2013。视觉质量评估：最新进展、编码应用与未来趋势。APSIPA Transactions on Signal and Information Processing 2（2013），e4。https://doi.org/10. 1017/ATSIP.2013.5[17] Alex Mackin、Fan Zhang 和 David R. Bull。2015。不同帧率下主观视频质量的研究。载于《2015年IEEE国际图像处理会议》（ICIP）。 3407 3411 10 1109 2015 7351436 – 。https://doi.org/ . /[18] ICIP. . Vignesh V Menon、Hadi Amirpour、Mohammad Ghanbari 和 Christian Tim- merer。2022。CODA：一种面向高帧率视频流传输的内容感知帧丢弃算法。载于《2022年数据压缩会议》（DCC）。475–475。https://doi.org/ 10 1109 52660 2022 00086 。/DCC . . Vignesh V Menon、Hadi [19] Amirpour、Mohammad Ghanbari 和 Christian Tim- merer。2022。OPTE：面向实时视频流传输的在线逐标题编码。载于《ICASSP 2022——2022年IEEE国际声学、语音与信号处理会议》（ICASSP）。 1865 1869 10 1109 43922 2022 9746745 – 。[20] https://doi.org/ . /ICASSP . . Vignesh V Menon、Hadi Amirpour、Mohammad Ghanbari 和 Christian Tim- merer。2022。面向自适应视频流传输的感知感知型逐标题编码。载于《2022年IEEE国际多媒体与博览会会议》（ICME）。IEEE计算机学会，美国加利福尼亚州洛斯阿拉米托斯，1–6。https://doi.org/10.1109/ ICME52920.2022.9859744[21] Vignesh V Menon、Hadi Amirpour、Mohammad Ghanbari 和 Christian Tim- merer。2023。EMES：面向基于高效视频编码的自适应码率流传输的高效多编码方案。ACM Trans. Multimedia Comput. Commun. Appl. 19, 3s，文章 129 2023 20 10 1145 3575659 （3月），页数。[22] https://doi.org/ . / Vignesh V Menon、Hadi Amirpour、Christian Timmerer 和 Mohammad Ghan- bari。2021。INCEPT：面向高效视频编码的帧内CU深度预测。载于《2021年IEEE第23届多媒体信号处理国际研讨会》（MMSP）。1–6。https: // doi.org/10.1109/MMSP53017.2021.9733517[23] Vignesh V Menon、Christian Feldmann、Hadi Amirpour、Mohammad Ghanbari 和 Christian Timmerer。2022。VCA：视频复杂度分析器。载于《第13届ACM多媒体系统会议论文集》（爱尔兰阿斯隆）（MMSys '22）。计算机协会，美国纽约州纽约市，259–264。https: // doi.org/10.1145/3524273.3532896[24] Vignesh V Menon、Prajit T Rajendran、Reza Farahani、Klaus Schoeffmann 和 Christian Timmerer。2023。融合纹理信息的面向流媒体应用的视频质量评估。arXiv:2302.14465 [cs.MM][25] Vignesh V Menon、Prajit T Rajendran、Christian Feldmann、Klaus Schoeffmann、Mohammad Ghanbari 和 Christian Timmerer。2023。面向恰可察觉差异的自适应实时流媒体双遍逐标题编码方案。（2023年3月）。https: //doi.org/ 10 36227 22256704 1 。/techrxiv. .v [26] Alexandre Mercat、Marko Viitanen 和 Jarno Vanne。2020。UVG数据集：用于视频编解码器分析与开发的50/120fps 4K视频序列。计算机协会，美国纽约州纽约市，297–302。https://doi.org/10.1145/3339825. 3394937[27] Manish Narwaria、Weisi Lin、Ian Vince McLoughlin、Sabu Emmanuel 和 Liang- Tien Chia。2012。基于傅里叶变换的可扩展图像质量度量。IEEE Transactions on Image Processing 21, 8（2012），3364–3377。https://doi.org/10.1109/ TIP.2012.2197010[28] Marta Orduna、César Díaz、Lara Muñoz、Pablo Pérez、Ignacio Benito 和 Narciso García。2020。360VR内容的视频多方法评估融合（VMAF）。IEEE Transactions on Consumer Electronics 66, 1（2020），22–31。https://doi.org/10. 1109/TCE.2019.2957987[29] Margaret H. Pinson 和 Stephen Wolf。2003。一种合并多个主观数据集的客观方法。载于《视觉通信与图像处理》 *2003* 583 592 10 1117 12 509909 。– 。https://doi.org/ . / . Airi Sakaushi、[30] Kenji Kanai、Jiro Katto 和 Toshitaka Tsuda。2017。视频监控系统中不同编码码率下图像增强效果的图像质量评估。载于《2017年IEEE第六届全球消费电子会议》（GCCE）。 1 2 10 1109 2017 8229463 – 。https://doi.org/ . /GCCE. . Li [31] Song、Xun Tang、Wei Zhang、Xiaokang Yang 和 Pingjian Xia。2013。SJTU 4K视频序列数据集。第五届多媒体体验质量国际研讨会（QoMEX2013）（2013年7月）。[32] Rajiv Soundararajan 和 Alan C. Bovik。2013。基于简化参考时空熵差异的视频质量评估。I EEE Transactions on

视频技术电路与系统23, 4 (2013), 684–694. https://doi.org/10.1109/TCSVT.2012.2214933[33] G. J. Sullivan、J. R. Ohm、W. J. Han 和 T. Wiegand。2012。高效视频编码（高效视频编码）标准概述。*IEEE*视频技术电路与系统汇刊22, 12 (2012), 1649–1668. [34] Shiqi Wang、Xiang Zhang、Siwei Ma 和 Wen Gao。2013。利用基元熵进行简化参考图像质量评估。见*2013*年图像编码研讨会（*PCS*）。 193 196 10 1109 2013 6737716 – 。[35] https://doi.org/ 。 /PCS。 。 Xu Wang、Gangyi Jiang 和 Mei Yu。 。基于Contourlet域和自然图像统计的简化参考图像质量评估。见*2009*年第五届国际图像与图形会议。45–50。https://doi.org/10. 1109/ICIG.2009.44[36] Zhou Wang 和 Alan C. Bovik。2006。现代图像质量评估。图像、视频和多媒体处理综合讲座2, 1 (2006), 1–156. https://doi.org/10.2200/S00010ED1V01Y200508IVM003 2009 [37] Zhou Wang 和 Qiang Li。2011。面向感知图像质量评估的信息内容加权。*IEEE* 图像处理汇刊 20, 5 (2011), 1185 1198 10 1109 2010 2092435 – 。 https://doi.org/ 。/TIP。 。[38] Zhou Wang 和 Eero Simoncelli。 。使用小波域自然图像统计模型进行无参考图像质量评估。*SPIE* 论文集——国际光学工程学会 5666 (03 2005)。https://doi.org/ 10 1117 12 597306 。 / 。Yong Yu、Xiaosheng Si、Changhua Hu 和 Jianxun Zhang。2019。循环神经网络综述：长短期记忆网络单元与网络架构。神经计算 31, 7 (07 2019), 1235–1270。 https://doi.org/10.1162/neco_a_01199[39] Saman Zadtootaghaj、Steven Schmidt、Nabajeet Barman、Sebastian Möller 和 Maria G. Martini。2018。基于与视频编码复杂度相关的游戏特征对电子游戏进行分类。载于*2018*年第*16*届网络与系统支持游戏年度研讨会（*NetGames*）。1–6。https: //doi.org/10.1109/NetGames.2018.8463434 2005 [40] 2005 2005