arXiv:2303.09818v1 [eess.IV] 17 Mar 2023

面向 HTTP 自适应流媒体的实时盲式体验质量评估指标

Chunyi Li [1], May Lim[2] , Abdelhak Bentaleb [3], 以及 *Roger Zimmermann[2]*上海交通大学[1], 新加坡国立大学[2],康考迪亚大学[3]

摘要

如今，在互联网中，HTTP 自适应流媒体（HAS）是视频流传输的主流标准，它根据自适应比特率（ABR）算法切换视频内容的码率。有效的体验质量（QoE）评估指标可以为自适应比特率算法提供关键反馈。然而，在客户端实时预测此类体验质量颇具挑战性。体验质量预测需要与人类视觉系统（HVS）高度一致、延迟低且能够进行盲评估，这些要求难以同时实现。为应对这一挑战，我们分析了 HTTP 自适应流媒体系统的各种特性，并提出一种非均匀采样指标以降低时间复杂度。此外，我们设计了一种有效的体验质量指标，将分辨率和重新缓冲时间作为服务质量（QoS），并将深度神经网络的时空输出以及特定切换事件作为内容信息。这些奖励和惩罚特征通过支持向量回归（SVR）模型回归得到质量分数。实验结果表明，我们的指标准确度比主流盲式体验质量指标高 0.3，计算时间仅为视频播放时长的 60%，这说明所提指标能够为自适应比特率算法提供实时指导，并提升 HTTP 自适应流媒体的整体性能。

索引词— 体验质量、HTTP 自适应流媒体、盲质量评估；

1. 引言

如今，视频已成为互联网上的主导应用。根据思科的调查 [1], , 视频服务目前已占用超过 80% 的互联网容量，而且需求仍在增长。为应对海量视频数据传输带来的挑战，内容提供商通常采用 HTTP 自适应流媒体（HAS），它能够适应动态网络条件和各种设备分辨率。HAS 通过 HTTP/TCP 协议栈以小片段的形式传送媒体内容。由于领先的内容提供商纷纷采用，HAS 已成为视频点播（VoD）服务的主导传输方式。

在 HTTP 自适应流媒体客户端（即媒体播放器）中，自适应比特率算法（自适应码率）会为每个视频片段的下载选择合适的码率级别。自适应比特率算法 [2] 的目标是最大限度地提升用户的体验质量（QoE）。因此，有效且客观的 QoE 指标对 HTTP 自适应流媒体系统至关重要。它既可在视频会话结束后用于评估自适应比特率算法的性能，也可在播放过程中用于指导该算法为下一个视频片段选择最合适的码率。

由于以下三项要求，为码率指导设计有效的客观体验质量指标颇具挑战性。(*i*) **与人类视觉系统高度一致**：体验质量指标需要与人类视觉系统保持一致，才能准确预测用户对视频播放质量的主观感受。(*ii*)**低时延**：为了在传输过程中提供指导，体验质量指标的反馈应与视频播放同步计算，而不是像自适应码率性能评估那样在传输结束后计算。在这种情况下，体验质量指标的计算时间不能超过片段的播放时长，因此必须降低其复杂度，以确保能够实时反馈。最后，(*iii*) **盲评估**：上述实时预测在客户端执行。因此，这项任务应为无参考（NR）任务，仅使用客户端可获取的压缩/失真视频，而非采用全参考（FR）或简化参考（RR）场景；后两者需要使用原始未压缩的源视频，或需要从服务器单独获取的部分参考特征。

2. 相关工作与贡献

一般而言，为 HTTP 自适应流媒体传输提供指导的现有体验质量指标可分为三类 [3]：基于服务质量的指标、基于信号保真度／内容的指标，以及混合指标。

基于服务质量的 指标 [4, 5]复杂度最低，因为其计算通常只需简单组合网络和／或客户端数据，例如视频比特率、启动时延和再缓冲时长。然而，由于计算时未考虑视频内容，而视频内容对感知质量有很大影响，因此这类指标往往与人类视觉系统的一致性最低。

基于内容的指标 [6, 7] 利用图像或视频质量评估（I/VQA）指标分析视频内容及其信号保真度，以预测视频的失真程度 [8]。图像质量评估指标以选定的视频帧为输入，根据这些帧的质量构建时间序列模型 [9]，并输出该视频的质量分数。由于图像质量评估指标需要针对每一帧反复计算，因此可能难以扩展。现有的视频质量评估指标的运行时间也往往较长。只有

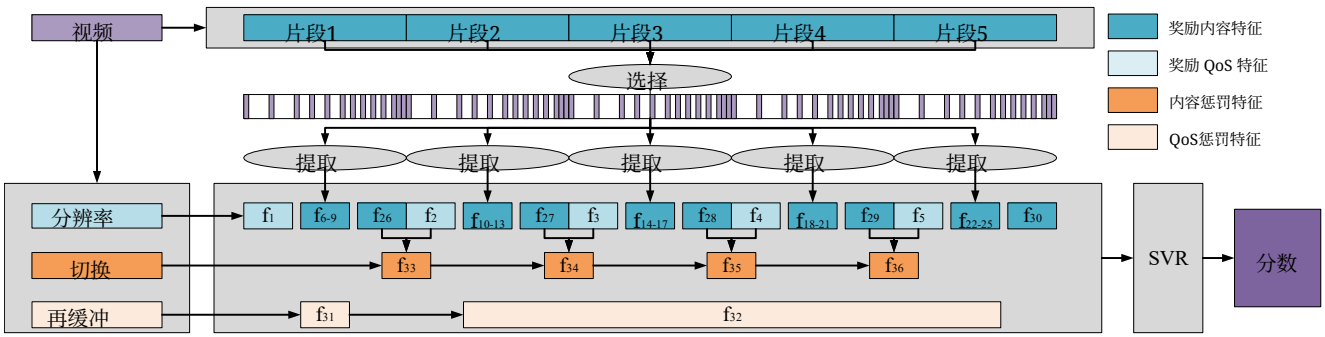


图 1。所提方法的框架。

少数经过简化的无参考视频质量评估指标 [10, 11] 能够为自适应码率提供实时反馈，而这种简化会导致其性能有所下降。

混合指标 [12, 13] 结合了服务质量和视频内容两方面的数据。其体验质量得分通常由图像质量评估/视频质量评估指标表示的奖励函数，以及基于服务质量因素的惩罚函数构成。这使指标能够在时间复杂度与人类视觉系统一致性之间取得平衡，因此有望用于码率指导。遗憾的是，现有的一些混合指标所采用的图像质量评估/视频质量评估核仅针对全参考任务设计[12], , 而通用模型 [14] 往往仅在全参考或部分参考核上与人类视觉系统高度一致，但在无参考场景下的一致性较低。因此，正如本文所示，这类混合型无参考指标的性能仍有进一步提升空间。

由于上述现有指标都无法同时满足第 1 节提出的三项要求，我们提出了一种用于 HTTP 自适应流媒体的新型体验质量评估指标，主要贡献如下。(*i*) 鉴于人类视觉系统在每个片段中的关注倾向会逐渐增强，我们采用非均匀采样方案。无需分析片段起始处过多的帧，逐步提高采样率即可降低时间复杂度。(*ii*) 我们将服务质量引入奖励特征。由于无参考函数不如全参考/部分参考函数有效，加入一些服务质量信息有助于提升模型的盲评估性能。(*iii*) 我们将内容引入惩罚特征。通过分析具体的帧变化，而非服务质量波动，可以更好地表征用户的体验质量。

3. 所提方法

为了设计出能够有效满足第 1 节所述要求的体验质量指标，我们确定了有助于实质性满足这些要求的新特征和改进方法。盲式体验质量指标的框架如图 1 所示，包含三个部分：采样、特征提取和质量回归。受第 2 节混合指标的启发，我们提取了四类特征：奖励/惩罚型服务质量特征和奖励/惩罚型内容特征。对于内容特征，首先对客户端缓冲区中的每个视频片段进行非均匀采样，以选择适合进行特征提取的图像帧子集

。服务质量/内容特征通过 ResNet-50、纹理分析和帧间差异提取，随后输入支持向量回归（SVR）模型进行回归，得到代表用户当前体验质量的质量分数。下面将详细介绍各个步骤。

3.1. 采样

为实时体验质量评估，基于内容的/混合指标仅分析部分帧，以降低复杂度。传统指标会对每个片段进行均匀采样，但人们普遍认为，片段中的各帧对体验质量的贡献可能并不相同。为研究各自的贡献，我们在 Waterloo sQoE III 视频数据集 [15]上运行广泛使用的无参考视频质量评估指标 Brisque [6], 。由于每个片段时长为两秒，且通常每秒出现一个帧内编码帧 [16], , 因此我们将片段分为前后两半（开始/结束）进行非均匀采样，并将其 QoE 表示为：$QoE = w_s QoE_s + w_e QoE_e$，其中 w_s、w_e分别是片段开始/结束部分的体验质量权重参数 QoE_s 和 QoE_e。采用 Spearman 秩相关系数（SRoCC）作为相关函数 $\mathcal{S}$，用于计算 QoE 与根据该数据集的主观评估获得的单刺激平均意见分（平均意见分） M之间的相关性：

$$\mathcal{S}\ QoE\ \ M\ = w_s \mathcal{S}\ QoE_s^{\,fr_s}\ \ M\ + (,)\ (,)\ w_e \mathcal{S}\ QoE_e^{\,fr_e}\ \ M$$

fr_s fr_e (,)，其中 和 分别是从片段开始/结束部分采样的帧。

为了更好地理解采样率与相关性表现之间的关系，我们首先使用八种不同的采样率和三种常见的图像质量评估指标（Niqe [17],Piqe [18], Brisque [6]）来预测体验质量。预测的体验质量与主观评分之间的归一化 SRoCC 表明，相关因子与采样率呈对数关系，如图 2 所示。因此，采样方案可以转化为以下优化问题：

$$\begin{cases} fr = fr_s + fr_e \\ \text{maximize}(w_s \log(fr_s) + w_e \log(fr_e)) \end{cases} \tag{1}$$

其中，fr 是所需采样帧的总数。通过拉格朗日乘子法，对式（1）中的对数函数求导可得，为获得最佳采样方案，fr 与 w 成正比：

$$\frac{fr_s}{fr_e} = \frac{w_s}{w_e} = \frac{\mathcal{S}(\text{brisque}(seg_s), M)}{\mathcal{S}(\text{brisque}(seg_e), M)} \tag{2}$$

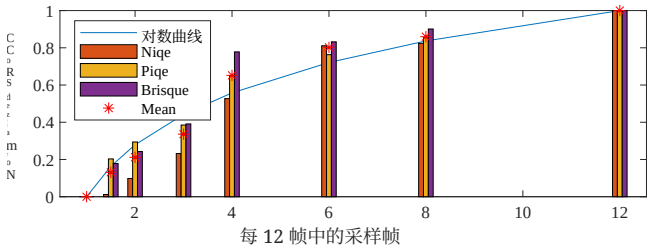


图 2。采样率与归一化 SRoCC 之间近似呈凸对数关系。

其中， seg_s 和 seg_e 分别是片段的起点和终点，体验质量由 brisque(·) 预测。详细的比例关系和推导见补充材料。

3.2. 特征提取

下面讨论如何提取这四类特征。

奖励服务质量特征。服务质量是指视频流会话期间网络或客户端侧的基本指标。其中，视频比特率、量化参数（QP）、帧率和分辨率（高度/宽度）对用户的体验质量有积极影响。在研究这些因素的相关性表现后，我们发现，视频高度作为奖励特征表现良好 r_1，可用于表征感知质量。因此，r_1 = 高度（seg），其中 seg 是 HTTP 自适应流媒体中的一个视频片段。

奖励内容特征。对第 3.1 节中的帧进行采样后，我们分析了这一系列图像，其中包含三类可提取的特征：结构特征、时间特征和色度特征。图像中的结构特征能够很好地反映感知质量 [19]，并广泛用于体验质量预测。时间特征既可以独立计算，也可以通过时空融合从结构特征中计算。色度特征与结构特征类似，可从图像中独立计算。然而，由于人类视觉系统通过三个通道处理视觉信号，计算色度特征可能会显著增加复杂度。为满足实时要求，我们采用结构特征，并通过时空融合整合时间特征，同时舍弃色度特征。

结构特征可分为空间特征和纹理特征。空间特征指图像中不同元素的相对空间位置或方向。图 3 展示了我们所采用的空间特征提取方法概览。该模块以 ResNet-50 为骨干网络，能够表示像素之间的空间相关性，并已证明其具有质量感知能力 [20]。对于片段中的第 i 个采样帧，我们将其转换为灰度图 $g(i)$，作为 ResNet-50 的输入，由其提取四个特征 $L_i = l_{1\sim4}(i)$，如下所示：

$$\begin{cases} P(i) = \text{ResNet50}(g(i)) \\ [l_1(i), l_2(i)] = [\max(P(i)), \min(P(i))] \\ [l_3(i), l_4(i)] = [\text{avg(P(i))}, \text{std(P(i))}] \end{cases} \tag{3}$$

其中，$P(n)$ 是 ResNet-50 的输出，i 是帧索引。在片段级别，我们不再将四个空

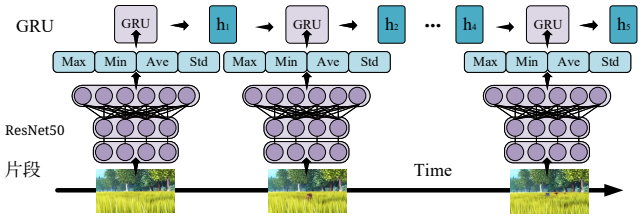


图 3。空间特征提取方法。

间特征计算为全局平均值，而是引入门控循环单元（GRU）[21] 来捕捉空间特征之间的时间关系。时空奖励 $r_{2\sim5}$ 为：

$$\begin{cases} h_i = GRU\left(W \cdot L_i + b, h_{i-1}\right) \\ r_{2\sim5} = h_{fr} \end{cases} \tag{4}$$

其中，W 和 b 是门控循环单元（GRU）[21], 中的权重和偏置参数，而 h_i 是一个类似于 L_i 的四单元元数组，用于在时域中保留人类视觉系统的记忆遗忘机制。因此，这四个特征同时结合了空间信息和时间信息。

纹理特征指图像中物体的表面特性。尽管 Sobel 算子、拉普拉斯算子 [22]等算子常用于纹理提取，但考虑到 ResNet-50 已为上述空间特征引入了较高的复杂度，此处需要一种更简单的方案。类似于视频压缩 [23], 中将宏块（MB）用作基本处理单元的做法，我们首先根据灰度图计算行和列的平均强度值 Ra_y, Ca_x。然后将灰度图 $g(i)$划分为多个 16×16宏块（MB）。对于每个宏块，我们分别计算灰度图数值与对应方向（水平、垂直和对角线）上的上述平均值之间的差值。我们选择上述最小差值作为该宏块的纹理，并将它们组合成纹理特征 r_6，如下所示：

$$\begin{cases} Hor_j = \sum |Ra_y(i) - g_{x,y}(i)| \\ Ver_j = \sum |Ca_x(i) - g_{x,y}(i)| \\ Dia_j = \sum |0.5(Ca_x(i) + Ra_y(i)) - g_{x,y}(i)| \\ r_6 = \sum \min(Hor_j, Ver_j, Dia_j) \end{cases} \tag{5}$$

其中，Hor、Ver 和 Dia 是分别使用行平均值 Ra_y 和列平均值 Ca_x 在三个方向上计算得到的纹理信息，而 j 是宏块索引。

QoS惩罚特征。在各种 QoS 指标中，再缓冲对 QoE 的负面影响最大之一 [24]。重缓冲时长和再缓冲事件的次数是体验质量模型中常用的惩罚特征 [25, 12]。此外，播放器往往愿意承受初始缓冲（少于两秒），以换取播放期间较少的再缓冲[26]。我们还注意到，QoE 的负面影响与重缓冲时长和再缓冲事件的次数均呈线性关系，因此我们只考虑其中一项。因此，我们将初始缓冲时长和平均重缓冲时长作为惩罚特征 p_1 和 p_2：

$$\begin{cases} p_1 = D_1 \\ p_2 = \frac{1}{T-1} \sum_{t=2}^{T} D_t \end{cases} \tag{6}$$

表 1。特征组及其定义列表。

特征组	组件	来源	描述
奖励服务质量	$f_{1\sim5}$	r_1	分辨率
奖励内容	$f_{6\sim25}$	$r_{2\sim5}$	时空
	$f_{26\sim30}$	r_6	纹理
惩罚服务质量	$f_{31\sim32}$	$p_{1\sim2}$	再缓冲
内容惩罚	$f_{33\sim36}$	p_3	切换、再缓冲（内容感知）

其中， T 是片段总数， t 是片段索引， D_t 是该片段的缓冲时长。

内容惩罚特征。视频QoS分析 [15] 表明，码率切换和重新缓冲事件会给用户带来糟糕的体验，其影响取决于若干因素，包括：(*i*) 切换模式：通常认为，从高码率降至低码率比从低码率升至高码率更令人不快，而较长的重新缓冲时间[27] 会进一步加剧此类切换事件带来的负面影响。(*ii*) 视频内容：对于物体移动缓慢且场景变化极少的视频，码率切换或重新缓冲事件的影响有限；而在动作片中，如果切换/重新缓冲发生在剧烈运动期间，或刚好发生在场景变化之后，通常会导致用户体验质量显著下降。传统上，每次切换（或重新缓冲）事件的负面影响是根据片段间码率变化（或重新缓冲时长）计算的，而不考虑视频内容。相反，我们采用高效的 FastSSIM [28] 指标来表示片段间视频内容的帧间差异。内容惩罚特征 p_3 综合了上述三个因素，即补充材料中所述的重新缓冲时间、切换幅度和内容，如式（7）所示。

$$\begin{cases} swh = \mathrm{ReLU}(bitrate(seg_t) - bitrate(seg_{t+1})) \\ p_3 = (1 + \frac{D_t}{C_1})(1 + \frac{swh}{C_2})/\mathrm{ssim}(seg_t, seg_{t+1}) \end{cases} \tag{7}$$

其中， swh 是经过 ReLU 函数映射的片段间码率差，因为它能够表征上述切换模式中的码率下降； C_1 和 C_2 是用于归一化的常数。FastSSIM 函数 ssim(·) 作用于片段 seg_t 中的最后一个采样帧和片段 seg_{t+1}中的第一个采样帧。

3.3. 整体体验质量回归

经过上述操作，我们最终得到两个奖励特征组 r_1 和 $r_{2\sim6}$，以及两个惩罚特征组 $p_{1,\ 2}$和 p_3，分别表示对用户体验质量的正面和负面影响。如表 1 所示，这些特征组映射到特征$f_{1\sim36}$ 。利用从服务质量和视频内容信息中提取的这些特征，通过支持向量回归（SVR）构建质量预测模型，以生成总体体验质量得分，如图 1 所示。如先前研究 [15], 所建议，我们使用播放过程中最近的五个片段，通过 SVR 模型进行预测。该 SVR 模型使用 LIBSVM [29]实现，并采用径向基函数（RBF）核进行特征融合[30]。

4. 性能评估

4.1. 实验设置

所提指标在 Waterloo sQoE III[15] 和 LIVE Netflix II [35] 数据集上进行了验证。这些数据集包含多种内容类型和视频编解码器的主观评分视频，并通过不同网络条件和自适应码率算法进行流传输。主观研究采用双任务单刺激（SS）实验，用户评分作为真实值。数据集按 80/20 的比例随机划分为训练集和测试集，同时确保相同的视频内容被划分到同一集合中 [10]。为进行公平比较，划分和评估过程重复 1,000 次，并将平均结果作为最终性能。

我们通过以下方式评估该指标：首先，为评估其与人类视觉系统的一致性，我们采用三种常见的相关性函数，即 SRoCC、肯德尔秩相关系数（KRoCC）和皮尔逊线性相关系数（PLCC），衡量该指标与主观评分的相关程度。其次，为评估其时延，我们测量了该指标在搭载 i7-8750H CPU 的笔记本电脑上的计算时间；这是 HAS 服务中常见的客户端配置 [1]。第三，由于这是一个盲评估场景，我们将该指标与 18 种主流盲QoE评估指标进行比较，并将它们作为基线。模型训练时，我们保持 ResNet-50 的参数不变（该网络已在 ImageNet [36] 上针对图像分类进行了预训练），并使用 Adam 优化器 [37]更新其他参数。为确保公平比较，其他基于学习的方法也采用类似机制进行训练。在 IQA 模型中，我们评估了三种不同的采样率，分别是每隔 5、10 和 20 帧均匀采样一次；对于混合模型，由于其原始 SSIM 特征无法用于无参考任务，我们每隔 20 帧均匀采样一次，并在其基于内容的特征中采用广泛使用的 Brisque [6]指标。

4.2. 实验结果与讨论

表2和图4展示了基线方法与所提方法的性能结果，其中相关因子越大，表示与人类视觉系统的一致性越高。计算时间以视频时长的比值表示；比值越小，复杂度越低，且必须小于1才能满足实时性要求。在基于内容的指标中，V-BLIINDS [31]的预测性能最佳，但其计算时间超过视频时长的 40× 倍，因此无法为码率指导提供实时反馈。基于服务质量的指标计算速度快得多，计算时间不到视频时长的1%，但其预测的体验质量与主观评分的相关性相对较差。大多数混合模型在一致性与时延之间实现了更好的平衡。然而，这些模型通常针对全参考特征而设计，但切换至

表2。Waterloo sQoE III 和 LIVE Netflix II 数据集上的性能结果。

Type	子类型	方法	Waterloo sQoE III				LIVE Netflix II			
			SRoCC	KRoCC	PLCC ↑	时间 ↓	SRoCC	KRoCC	PLCC ↑	时间 ↓
内容	IQA	Brisque5[6]	0.4959	0.3468	0.4690	1.944	0.3146	0.2182	0.2009	1.614
		Brisque10[6]	0.4842	0.3394	0.4600	0.965	0.2704	0.1880	0.1344	1.276
		Brisque20[6]	0.4547	0.3146	0.4478	0.498	0.2543	0.1794	0.1517	0.455
		NIQE10[17]	0.4021	0.2732	0.4488	1.027	0.6538	0.4761	0.6684	0.529
		PIQE10[18]	0.4232	0.2807	0.4349	1.128	0.6746	0.4898	0.6871	0.541
	VQA	VIIDEO[7]	0.3946	0.2651	0.4903	8.047	0.2843	0.1885	0.3228	5.624
		V-BLIINDS[31]	0.7389	0.5456	0.7244	45.625	0.7510	0.5755	0.7653	42.755
		Resnet50[32]	0.5707	0.4113	0.5635	0.381	0.4278	0.2995	0.4317	0.307
		FAST-VQA[11]	0.7391	0.5500	0.7710	0.246	0.5137	0.3644	0.5748	0.232
QoS	客户端	FTW[4]	0.1835	0.1337	0.3229	0.001	0.0804	0.0858	0.0648	0.001
		MoK2011[25]	0.1687	0.1294	0.2156	0.001	0.0795	0.0650	0.0874	0.001
	网络	Liu2012[33]	0.2529	0.1717	0.2424	0.001	0.6633	0.4684	0.6366	0.001
		Xue2014[34]	0.3412	0.2245	0.3081	0.003	0.5830	0.4123	0.4961	0.003
		Yin2015[5]	0.1458	0.0932	0.3232	0.007	0.0804	0.0616	0.0648	0.007
混合	Mix	SQI[13]	0.1515	0.1100	0.2225	0.501	0.7347	0.5298	0.6329	0.458
		电视服务质量体验[14]	0.5068	0.3565	0.4667	0.524	0.6686	0.4136	0.5109	0.482
		Bentaleb2016[12]	0.1979	0.1387	0.3405	0.498	0.4454	0.2982	0.4530	0.456
		KSQI[3]	0.5285	0.3875	0.5268	0.505	0.7394	0.5492	0.7315	0.462
		所提方法	**0.8627**	**0.6871**	**0.8824**	0.606	**0.7739**	**0.5914**	**0.7898**	0.691

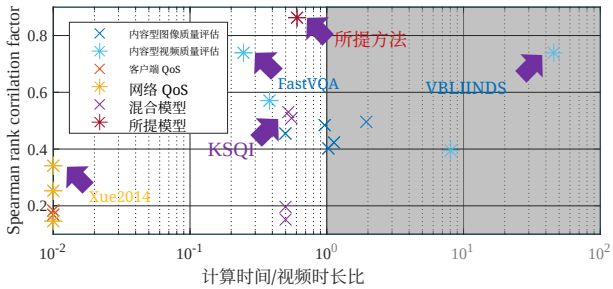


图 4。基线方法和所提方法的 SRoCC 与时间比率表现。时间比率为 < 1 的指标（白色面板）可实现实时 QoE 预测，为码率提供指导。

无参考后，其预测性能便不再理想。结果表明，在两个数据集的所有三种相关性度量上，我们的指标均优于所有基于服务质量的指标和混合指标（相较其他混合指标，增幅最高达0.35）；在两个数据集的SRoCC和PLCC方面，也优于所有基于内容的指标，同时将计算时间比控制在1以下。在全局体验质量指标的性能因素方面[38],，我们的ROC曲线下面积（AUC）值比SOTA HAS QoE 指标高出约0.09，且在正确分类方面优于所有现有指标。因此，我们的指标能够克服盲评场景中缺少参考信息的问题，仍可提供有效的体验质量预测，与人类视觉系统具有高度一致性（平均SRoCC为0.81），且时延足够低（约为视频播放时长的65 %）。

4.3. 消融研究

为验证采样方法和不同特征类型的贡献，我们还进行了消融研究，结果如表 3 所示。具体考察的因素包括：(1) 非均匀采样，(2) 奖励 QoS 特征，(3) 奖励内容特征（空间特征和纹理特征），(4) QoS惩罚特征，以及 (5) 内容惩罚特征。结果表明，移除任何单一因素都会导致性能下降，这证实了这些因素均对表2中的性能结果有所贡献。各组的耗时能够满足实时评估的要求。消融结果还表明，QoS特征组 (2)(4) 比内容特征组 (3)(5) 更有效，两者结合可获得理想的性能，而采样方法 (1) 则能以较小的额外耗时进一步提升模型性能。

表3。在滑铁卢 sQoE III 数据集上舍弃不同特征后的性能结果。

舍弃的特征	SRoCC	KRoCC	PLCC	Time
None	**0.8627**	**0.6871**	**0.8824**	0.606
(1)	0.8514	0.6716	0.8694	0.605
(2)	0.4521	0.3270	0.4761	0.606
(3) $r_2 \sim r_5$	0.7662	0.5822	0.8194	0.381
(3) r_6	0.8468	0.6680	0.8637	0.441
(3) 全部	0.7313	0.5478	0.7490	0.216
(4)	0.8319	0.6544	0.8691	0.606
(5)	0.8434	0.6767	0.8639	0.390
(2)(4)	0.3204	0.2214	0.4006	0.605
(3)(5)	0.7010	0.5215	0.7073	0.001

5. 结论

本研究旨在解决为码率指导而测量 HTTP 自适应流媒体客户端感知质量这一难题。我们提出了一种盲体验质量评估指标，无需参考信息，即可低时延地提供与人类视觉系统高度一致的体验质量反馈。具体而言，我们将服务质量和视频内容信息映射为奖励特征和惩罚特征，并引入非均匀采样机制，以识别用于分析的相关帧。实验结果表明，所提方法

指标在两个数据库的三种相关性度量中均取得最佳结果，表明其与人类视觉系统高度一致，同时满足时延和盲评估要求。该指标适用于在客户端实时评估体验质量，有助于提高 HTTP 自适应流媒体服务的整体资源利用率。

6. 参考文献

[1] 思科，《思科视觉网络指数：2018–2023年预测与趋势》，白皮书，2020年。[2] A. Bentaleb、B. Taani、A. C. Begen、C. Timmerer 和 R. Zimmermann，《HTTP流媒体码率自适应方案综述》，*IEEE Communications Surveys & Tutorials*。[3] Z. Duanmu、W. Liu、D. Chen、Z. Li、Z. Wang 等，《面向自适应流媒体视频的知识驱动体验质量模型》，*arXiv* 预印本 *arXiv:1911.07944*，2019年。[4] T. Hoßfeld、R. Schatz、E. Biersack 和 L. Plissonneau，《YouTube 上的互联网视频传输：从流量测量到体验质量》，载于数据流量监测与分析。施普林格，2013年。[5] X. Yin、A. Jindal、V. Sekar 和 B. Sinopoli，《HTTP 动态自适应视频流的控制理论方法》，载于*ACM SIGCOMM*，2015年。[6] A. Mittal、A. K. Moorthy 和 A. C. Bovik，《空间域中的无参考图像质量评估》，*IEEE TIP*，2012年。[7] A. Mittal、M. A. Saad 和 A. C. Bovik，《完全盲视频完整性判定器》，*IEEE TIP*，2015年。[8] G. Zhai 和 X. Min，《感知图像质量评估：综述》，中国科学：信息科学，2020年。[9] X. Sui、K. Ma、Y. Yao 和 Y. Fang，《将全向图像作为移动摄像机视频进行感知质量评估》，*IEEE TVCG*，2021年。[10] Z. Tu、X. Yu、Y. Wang、N. Birkbeck、B. Adsumilli 和 A. C. Bovik，《RAPIQUE：用户生成内容的视频质量快速精准预测》，*IEEE OJSP*，2021年。[11] H. Wu、C. Chen、J. Hou、L. Liao、A. Wang、W. Sun、Q. Yan 和 W. Lin，《FAST-VQA：通过片段采样实现高效的端到端视频质量评估》，欧洲计算机视觉国际会议，2022年。[12] A. Bentaleb、A. C. Begen 和 R. Zimmermann，《SDNDASH：利用软件定义网络提升 HTTP 自适应流媒体的体验质量》，载于*ACM*多媒体会议，2016年。[13] Z. Duanmu、K. Zeng、K. Ma、A. Rehman 和 Z. Wang，《流媒体视频的体验质量指数》，*IEEE*信号处理选题期刊，2016年。[14] D. Ghadiyaram、J. Pan 和 A. C. Bovik，《学习连续时间流媒体视频体验质量模型》，*IEEE TIP*，2018年。[15] Z. Duanmu、A. Rehman 和 Z. Wang，《自适应视频流媒体体验质量数据库》，*IEEE*广播汇刊，2018年。[16] H. Amirpour、E. C. etinkaya、C. Timmerer 和 M. Ghanbari，《面向自适应 HTTP 流媒体的快速多码率编码》，载于数据压缩会议（*DCC*），2020年。[17] A. Mittal、R. Soundararajan 和 A. C. Bovik，《打造“完全盲”的图像质量分析器》，*IEEE*信号处理快报，2012年。[18] N. Venkatanath、D. Praneeth、M. C. Bh、S. S. Channappayya 和 S. S. Medasani，《使用基于感知的特征进行盲图像质量评价》，载于 IEEE NCC，2015年。[19] Q. Li、W. Lin、J. Xu 和 Y. Fang，《使用统计结构与亮度特征进行盲图像质量评估》，IEEE TMM，2016年。[20] J. Kim、H. Zeng、D. Ghadiyaram、S. Lee、L. Zhang 和 A. C. Bovik，《用于图像质量预测的深度卷积神经模型：数据驱动图像质量评估的挑战与解决方案》，IEEE Signal Process Mag，2017年。[21] J. Yan、J. Li、Y. Fang、Z. Che、X. Xia 和 Y. Liu，《自由视点视频的主观与客观体验质量》，IEEE TIP，2022年。[22] M. Sonka、V. Hlavac 和 R. Boyle，《图像处理、分析与机器视觉》，圣智学习，2014年。[23] X. Min、J. Zhou、G. Zhai、P. Le Callet、X. Yang 和 X. Guan，《用于光场重建、压缩和显示质量评估的指标》，IEEE TIP，2020年。[24] C. G. Bampis 和 A. C. Bovik，《学习预测流媒体视频体验质量：失真、再缓冲与记忆》，arXiv 预印本 arXiv:1703.00633，2017年。[25] R. K. Mok、X. Luo、E. W. Chan 和 R. K. Chang，《QDASH：感知体验质量的 DASH 系统》，载于 ACM MMSys，2012年。[26] Tisa-Selma、A. Bentaleb 和 S. Harous，《利用机器学习进行视频体验质量推断》，载于 IEEE IWCMC，2021年。[27] B. Taraghi、M. Nguyen、H. Amirpour 和 C. Timmerer，《Intense：深入研究卡顿事件和质量切换及其对 HTTP 自适应流媒体体验质量的影响》，IEEE Access，2021年。[28] M.-J. Chen 和 A. C. Bovik，《快速结构相似性指数算法》，《实时图像处理期刊》，2011年。[29] C.-C. Chang 和 C.-J. Lin，《LIBSVM：支持向量机库》，ACM TIST，2011年。[30] L. Krasula、K. Fliegel 和 P. Le Callet，《FFT MI：自然色调映射图像质量评估的特征融合》，IEEE TMM，2020年。[31] M. A. Saad、A. C. Bovik 和 C. Charrier，《自然视频质量的盲预测》，IEEE TIP，2014年。[32] K. He、X. Zhang、S. Ren 和 J. Sun，《用于图像识别的深度残差学习》，载于 IEEE CVPR，2016年。[33] X. Liu、F. Dobrian、H. Milner、J. Jiang、V. Sekar、I. Stoica 和 H. Zhang，《协同互联网视频控制平面方案》，载于 ACM SIGCOMM，2012年。[34] J. Xue、D.-Q. Zhang、H. Yu 和 C. W. Chen，《评估自适应 HTTP 视频流媒体的体验质量》，载于 IEEE ICME研讨会，2014年。[35] C. G. Bampis、Z. Li、I. Katsavounidis、T.-Y. Huang、C. Ekanadham 和 A. C. Bovik，《迈向感知优化的端到端自适应视频流媒体》，arXiv 预印本 arXiv:1808.03898，2018年。[36] O. Russakovsky、J. Deng、H. Su、J. Krause、S. Satheesh、S. Ma、Z. Huang、A. Karpathy、A. Khosla、M. Bernstein 等，《ImageNet 大规模视觉识别挑战赛》，IJCV，2015年。[37] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，arXiv 预印本 arXiv:1412.6980，2014年。[38] L. Krasula、K. Fliegel、P. Le Callet 和 M. Klíma，《客观图像与视频质量模型的准确性：一种新的性能评估方法》，载于 QoMEX，2016年。