基于感知和记忆的视频动态质量评价

林丽群[1,2],暨书逸[1],何嘉晨[1],赵铁松[1,2*],陈炜玲[1,2],郭宗明[3]

(1. 福州大学物理与信息工程学院福建省媒体信息智能处理与无线传输重点实验室,福建福州 350108;2. 中国福建光电信息科学与技术创新实验室(闽都创新实验室),福建福州 350108;3. 北京大学王选计算机研究所,北京 100871)

摘 要: 由于网络环境的多变性,视频播放过程中容易出现卡顿、比特率波动等情况,严重影响了终端用户的体验质量. 为优化网络资源分配并提升用户观看体验,准确评估视频质量至关重要. 现有的视频质量评价方法主要针对短视频,普遍关注人眼视觉感知特性,较少考虑人类记忆特性对视觉信息的存储和表达能力,以及视觉感知和记忆特性之间的相互作用. 而用户观看长视频的时候,其质量评价需要动态评价,除了考虑感知要素外,还要引入记忆要素. 为了更好地衡量长视频的质量评价,本文引入深度网络模型,深入探讨了视频感知和记忆特性对用户观看体验的影响,并基于两者特性提出长视频的动态质量评价模型. 首先,本文设计主观实验,探究在不同视频播放模式下,视觉感知特性和人类记忆特性对用户体验质量的影响,构建了基于用户感知和记忆的视频质量数据库(Video Quality Database with Perception And Memory,PAM-VQD);其次,基于PAM-VQD数据库,采用深度学习的方法,结合视觉注意力机制,提取视频的深层感知特征,以精准评估感知对用户体验质量的影响;最后,将前端网络输出的感知质量分数、播放状态以及自卡顿间隔作为三个特征输入长短期记忆网络,以建立视觉感知和记忆特性之间的时间依赖关系. 实验结果表明,所提出的质量评估模型在不同视频播放模式下均能准确预测用户体验质量,且泛化性能良好.

关键词: 视觉感知特性;记忆效应;体验质量;深度学习;注意力机制

基金项目: 国家自然科学基金(No.62171134);福建省自然科学基金(No.2022J02015)

中图分类号: TP181;TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2024)11-3727-14

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230283

Research of Video Dynamic Quality Evaluation Based on Human Perception and Memory

LIN Li-qun[1,2], JI Shu-yi[1], HE Jia-chen[1], ZHAO Tie-song[1,2*], CHEN Wei-ling[1,2], GUO Chong-ming[3]

(1. *Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information*, *College of Physics and Information Engineering*, *Fuzhou University*, *Fuzhou*, *Fujian* 350108, *China*;
2. *Fujian Science & Technology Innovation Laboratory for Optoelectronic Information of China*, *Fuzhou*, *Fujian* 350108, *China*;
3. *Wangxuan Institute of Computer Technology*, *Peking University*, *Beijing* 100871, *China*)

Abstract: Due to the variability of the network environment, video playback is prone to lag and bit rate fluctuations, which seriously affects the quality of end-user experience. In order to optimize network resource allocation and enhance user viewing experience, it is crucial to accurately evaluate video quality. Existing video quality evaluation methods mainly focus on the visual perception characteristics of short videos, with less consideration of the ability of human memory characteristics to store and express visual information, and the interaction between visual perception and memory characteristics. In contrast, when users watch long videos, video quality evaluation needs dynamic evaluation, which needs to consider both perceptual and memory elements. To better measure the quality evaluation of long videos, we introduce a deep network model to deeply explore the impact of video perception and memory characteristics on users' viewing experience, and proposes a dynamic quality evaluation model for long videos based on these two characteristics. Firstly, we design subjective experiments to investigate the influence of visual perceptual features and human memory features on user experience quality under different video playback modes, and constructs a video quality database with perception and memory (PAM-VQD) based on user perception and memory. Secondly, based on the PAM-VQD database, a deep learning methodology is utilized

收稿日期:2023-03-30;修回日期:2023-08-22;责任编辑:王天慧

*通讯作者:赵铁松

to extract deep perceptual features of videos, combined with visual attention mechanism, in order to accurately evaluate the impact of perception on user experience quality. Finally, the three features of perceptual quality score, playback status and self-lag interval output from the front-end network are fed into the long short-term memory network to establish the temporal dependency between visual perception and memory features. The experimental results show that the proposed quality assessment model can accurately predict the user experience quality under different video playback modes with good generalization performance.

Key words: visual perceptual properties; memory effect; quality of experience (QoE); deep learning; attention mechanism

Foundation Item(s): National Natural Science Foundation of China (No.62171134); Natural Science Foundations of Fujian Province (No.2022J02015)

1 引言

随着网络服务的快速发展以及智能设备数量的显著增长,视频业务已经广泛应用于人类生产和生活的各个方面,如网络直播、远程视频会议和短视频等[1].然而,在实际应用中,由于网络带宽、自适应流媒体技术等限制,视频播放过程中会产生初始缓冲、卡顿和比特率波动等多种播放事件.这些播放事件降低了视频播放的流畅性和清晰度,严重影响了用户的视频体验质量(Quality of Experience,QoE)[2].为了提供符合用户感知需求的视频服务,在视频播放过程中准确评估用户QoE已成为视频质量评价(Video Quality Assessment,VQA)领域的研究热点.因此,许多国内外专家学者致力于与QoE相关的视频质量评价研究,并取得了许多创新性的成果.对现有的视频质量评价方法进行总结,可分为以下两类:

(1)主观质量评价

主观质量评价是一种直接可靠的评价方法,能够准确地评估在不同视频场景和设置下的用户QoE[3].文献[4]研究证明了初始缓冲和卡顿事件与QoE的下降之间存在线性关系,且高质量视频产生的卡顿事件对QoE的影响更大.文献[5,6]的实验结果表明,卡顿事件会对用户QoE产生显著影响,用户对视频的不满意程度也会相应增加.为了测试VQA算法的通用性,并促进QoE指标的实际应用,文献[7]采用不同的网络条件和自适应算法构建了大规模QoE视频数据库,并通过主观实验对用户QoE进行评估和对比.实验结果表明,流媒体视频的QoE评估需要采用比传统VQA模型更复杂的建模方式.

主观质量评价方法虽然提供了真实可靠的QoE评估结果,但对实验环境和设备的要求较高,无法在应用系统中大规模部署.因此,客观质量评价方法是学者们重点研究的方向.

(2)客观质量评价

在过去,衡量视频质量的指标体系侧重于服务质量(Quality of Service,QoS),其可以评估网络满足用户服务需求的能力.Sawabe等人[8]提出了一种网络视频服务的QoS分析模型,该模型使网络运营商能够估计所需的QoS,以提供用户所需的视频质量,并降低网络稳定运行的维护成本.Wahab等人[9]将延迟和抖动作为网络QoS指标,建立了QoS与用户感知质量之间的映射关系.由于QoS指标是从网络传输层的角度客观反映用户对服务的满意程度,与用户的主观感受仍有一定差距.因此,以用户主观感受为核心的QoE逐渐成为VQA的研究重点.

综上,通过对现有的视频质量评价研究工作进行总结,本文发现主要存在以下不足:

(1)大多数现有的QoE数据库在时长设置和视频播放类型方面相对有限,并且公开数据库较少.目前,QoE数据库大多使用10 s到1 min的视频序列进行主观实验,这些视频序列的时长较短,无法有效分析长期记忆效应对用户QoE的影响.此外,即使存在持续时间较长的视频序列,但其考虑的视频播放类型也相对单一,无法充分反映不同播放事件对用户QoE的影响.

(2)现有的基于记忆特性的VQA算法性能仍待提升.目前QoE数据库中的视频序列时长较短,无法充分反映用户的长期记忆特性.因此,在此基础上构建的基于记忆特性的VQA算法无法有效建立长期记忆效应与用户QoE之间的映射关系.

为此,本文重点研究基于视觉感知特性和人类记忆特性的视频质量评价方法,通过构建基于用户体验质量的视频主观数据库探究视觉感知和记忆特性对用户QoE的影响,并构建基于感知和记忆的视频质量评价模型,以提高QoE的预测精度.

2 构建基于用户体验质量的视频主观数据库PAM-VQD

为了研究视觉感知特性和记忆特性对用户QoE的影响,首先,本文选取场景分布广泛的高质量视频作为实验源视频;其次,模拟网络带宽的波动,设计具有不同播放模式的测试序列,用于观测记忆效应对用户QoE的影响;最后,通过主观实验获取用户在不同播放模式

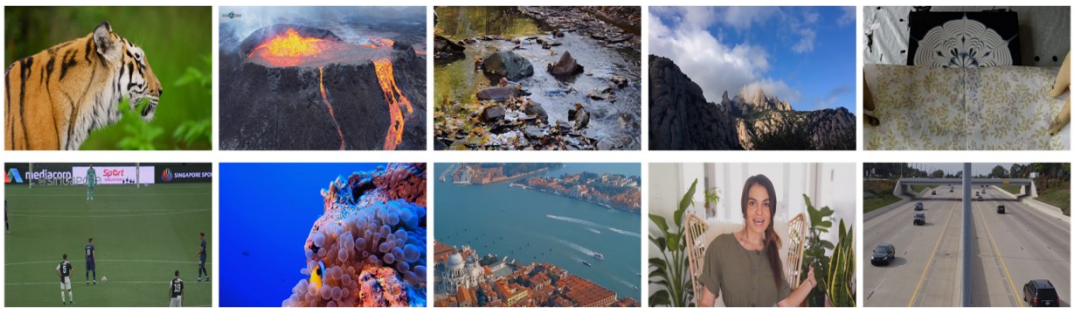


图1　源视频内容示例图

下的视频主观质量评分,从而建立基于感知和记忆的视频主观数据库(Video Quality Database with Perception And Memory,PAM-VQD).

2.1　源视频选择

目前,许多成功的研究使用了10~15 s的短视频序列[10~14],如LIVE(Laboratory for Image and Video Engineering) Video Quality Database[10]、LIVE Mobile VQA Database[11]和CSIQ(Categorical Subjective Image Quality) Video Quality Database[12]数据集. 然而,这些研究并不能反映典型的视频流情况,即用户观看的视频可能长达几分钟. 因此,为了进一步研究视频播放过程中用户记忆效应对QoE的影响,本次实验将采用10个时长为2 min的源视频,其视频内容示例如图1所示. 以上10个视频内容均来自用户经常访问的在线视频网站YouTube,空间分辨率为1 920×1 080,帧率为24~50 fps,视频内容包括:动物(Animal)、火山(Volcano)、溪水(Brook)、风景(Scene)、绘画(Painting)、足球(Football)、海洋(Ocean)、建筑(Architecture)、人类(Human)以及交通(Traffic).

为保证视频内容的时空复杂性,本文以空间信息(Spatial Information,SI)和时间信息(Temporal Information,TI)作为视频序列特征值[13],计算方式如下:

$$\mathrm{SI}=\max_{\mathrm{time}}\{\mathrm{std}_{\mathrm{space}}[\mathrm{sobel}(F_n)]\} \tag{1}$$

$$\mathrm{TI}=\max_{\mathrm{time}}\{\mathrm{std}_{\mathrm{space}}[F_n(i,j)-F_{n-1}(i,j)]\} \tag{2}$$

其中,$F_n(i,j)$为第n帧的第i行和第j列的像素,$\mathrm{std}_{\mathrm{space}}$为经Sobel滤波器滤除后的帧中像素的标准差,$\max_{\mathrm{time}}$为时间序列的最大值. 一般而言,视频的SI值越高,空间场景越复杂;TI值越高,场景时域变化越明显. 如图2所示,本次实验选取的源视频具有不同的时空复杂度,且广泛跨越SI-TI空间,因此具有一定代表性.

2.2　测试序列构建

本实验主要通过设计不同的视频播放模式,以分析不同的记忆特性对用户QoE的影响. 测试序列的构建步骤如下.

(1)源视频处理

为模拟视频播放过程中出现的比特率切换事件,

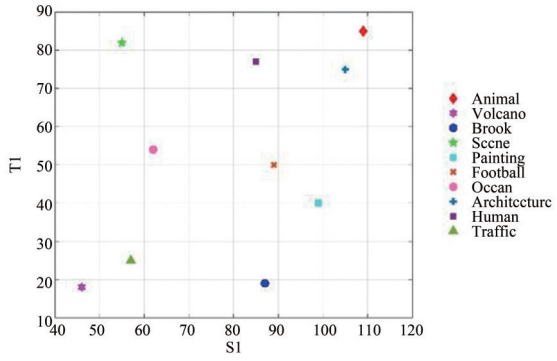


图2　视频的空间和时间信息

本文将2 min的源视频按照表1中显示的分辨率和视频比特率进行编码. 本次实验设置了3种常见的视频播放比特率,分别为一般网络条件下的最佳视频比特率(6 000 kbps)以及2种较低的比特率(1 000 kbps和750 kbps),用于模拟可用带宽不足时产生的比特率适应事件.

表1　测试序列的分辨率及比特率设置

视频名称	360P	480P	1080P
分辨率	640×360	854×480	1 920×1 080
视频比特率/kbps	750	1 000	6 000

在实际观看过程中,不同分辨率的视频经过播放器的渲染之后会自动适应于屏幕分辨率的高低. 因此,本文将不同空间分辨率的视频统一利用双线性插值算法上采样至1 920×1 080进行播放,以还原用户的真实观看体验. 为模拟视频播放过程中的初始缓冲和卡顿事件,本文采用Adobe Premiere Pro软件制作卡顿效果,随后覆盖到指定的视频帧,并设置所需的卡顿时长,最终生成的卡顿效果如图3所示.

(2)构建测试序列

经过上述处理后,本文可进行测试序列的构建. 如图4所示,本次实验设置了八种视频播放模式,每种模式下有10个视频,因此数据库PAM-VQD测试序列的总数为80个. 不同播放模式的设置依据如图4所示.

测试序列1:视频稳定播放. 该序列用于获取用户在不受各类播放事件干扰下的记忆情况,共有10个视频,每个视频的持续时间为2 min视频比特率设置为一

图3 存在卡顿事件的测试序列

般网络条件下的最佳视频比特率(6 000 kbps),同时并未设置初始缓冲、比特率切换和卡顿事件.

测试序列2:在视频播放开始时设置初始缓冲事件.初始缓冲是指视频开始播放前的等待时间[14].为了研究视频播放初期产生的缓冲事件对用户记忆的影响,本文构建了一组在视频播放前存在3 s缓冲时间的测试序列,共有10个视频,视频的比特率为6 000 kbps.由于插入3 s的初始缓冲事件,每个视频的持续时间为123 s.

测试序列3:在视频播放过程中设置卡顿事件.为了研究视频播放过程中的卡顿事件对用户记忆的影响,本文构建了一组视频播放过程中存在3 s卡顿时延的测试序列,共有10个视频,视频的比特率为6 000 kbps.由于插入3 s的卡顿事件,每个视频的持续时间为123 s.

测试序列4:在视频播放过程中设置两次的卡顿事件.在预实验中,本文发现视频播放过程中的卡顿事件容易被测试人员遗忘.因此,本文在视频播放过程中设置了两次的卡顿事件,以观察不同数量的卡顿事件是否会对用户QoE产生更大的影响.该测试序列共有10个视频,视频的比特率为6 000 kbps,由于插入两个时长为3 s的卡顿事件,每个视频的持续时间为126 s.

测试序列5:在视频播放过程中设置三次的卡顿事件.该测试序列共有10个视频,视频的比特率为6 000 kbps.由于插入3个时长为3 s的卡顿事件,每个视频的持续时间为129 s.

测试序列6:在视频播放结尾设置卡顿事件.为了研究视频播放结尾产生的卡顿事件对用户记忆的影响,本文构建了一组视频播放结尾存在3 s卡顿事件的测试序列,共有10个视频,视频的比特率为6 000 kbps.由于插入3 s的卡顿事件,每个视频的持续时间为123 s.

测试序列7:在视频播放结尾设置比特率向下切换事件.比特率切换是指客户端测量当前的带宽和/或缓冲区状态,并以适当的比特率请求视频的下一部分,以维持视频播放的流畅度[15].为了研究视频播放结尾产生的比特率切换事件对用户记忆的影响,本文构建了一组在视频播放结束前15 s存在比特率向下切换事件的测试序列,共有10个视频,每个视频的持续时间为2 min,测试序列前105 s的视频比特率为6 000 kbps,最后15 s的视频比特率为1 000 kbps.

测试序列8:在视频播放过程中设置连续的比特率波动事件.为了研究网络条件不稳定时的用户记忆状态,本文构建了一组视频播放过程中存在多个比特率切换事件的测试序列,共有10个视频,每个视频的持续时间为2 min.每个测试序列的前30 s将维持原始比特率(6 000 kbps)进行播放;当视频播放至30 s时,视频比特率将向下切换至750 kbps,模拟可用带宽不足时产生的比特率适应事件,持续时间为20 s;当视频播放至50 s时,恢复原始比特率进行播放,持续时间为20 s,随后比特率再次切换至750 kbps,持续时间为20 s,直至最后30 s视频才恢复原始比特率进行播放.

2.3 主观实验

本文按照ITU-R BT.500[16]标准进行室内场景的布置,并确保实验过程中无外部噪声干扰.实验中使用的液晶显示器的分辨率为1 920×1 080像素,测试序列的播放软件为potplayer.所有测试序列均以1 920×1 080的分辨率在液晶显示器上进行播放.

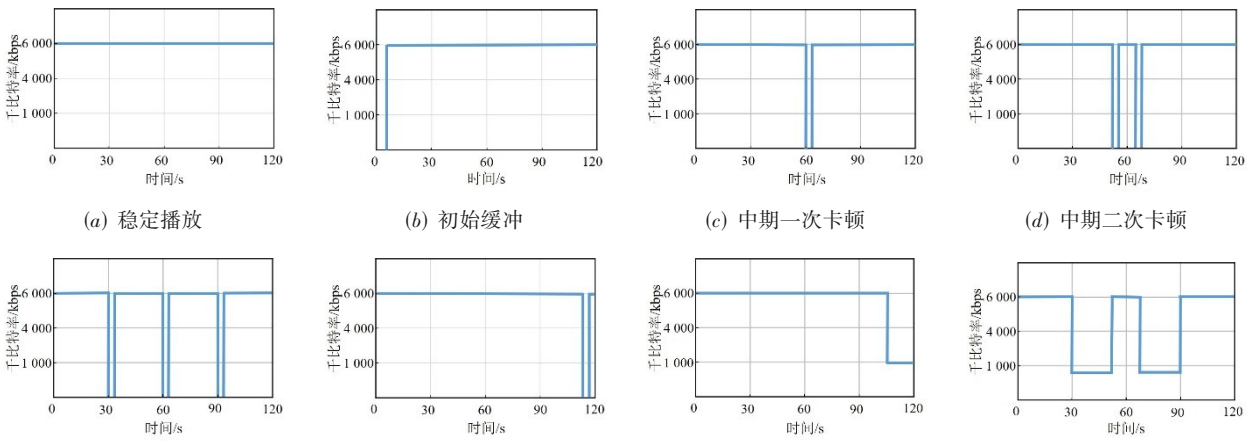
参与本次实验的测试人员共有18名,包括9名男性和9名女性,年龄在20~30岁之间,满足ITU-R BT.500标准中测试人员不少于15人的要求.所有的测试人员的视力和色觉均正常,且大部分没有视频质量评价的研究经历.本次实验采用单激励法对测试序列进行主观评分,评分范围为0~10分[17].实验分为预实验和正式实验两个阶段.

预实验阶段,为了确保正式实验时能获得相对稳定的实验结果,首先,向参与测试的人员介绍评分流程以及相关事宜;其次,向测试人员展示8个测试序列集以外的视频序列,并模拟视频评分过程,以确保测试人员熟悉整个实验流程.

正式实验阶段,首先,测试人员对所有源视频进行评分.所有源视频均单独且随机播放,每个源视频的持续时间为2 min,且相邻视频之间会出现10 s的黑屏时间供测试人员进行评分记录.源视频总数为10个,观看总时长为1 300 s;其次,参与测试的人员对所有的测试序列根据实验要求进行打分.所有测试序列将随机显示,每个测试序列的持续时间至少为2 min,且相邻视频之间有10 s的黑屏时间供测试人员进行打分.测试序列的总数为80个,每个测试人员的观看总时长约为2.5 h.为了将视觉疲劳对测试人员评分结果的影响降至最低,实验过程共分为5个实验阶段,每个实验阶段的时长限制在30 min以内,相邻阶段之间设置10 min的休息时间.

2.4 数据库构建

本文收集每名测试人员对所有测试序列的评分结



(a) 稳定播放　(b) 初始缓冲　(c) 中期一次卡顿　(d) 中期二次卡顿

(e) 中期三次卡顿　(f) 结尾卡顿　(g) 比特率向下切换　(h) 比特率波动

图4　8种视频播放模式

果,并进行相应的统计处理,即通过计算每个测试人员对于视频的评分结果与平均主观意见分(Mean Opinion Score,MOS)之间的相关系数,以验证获得的MOS值的可靠性. 图5的实验结果表明此次主观实验所得到的视频MOS值具有很高的准确度与真实性.

此外,为获得有效的实验数据,本文对所有测试人员的评分结果进行数据筛选. 筛选过程依据ITU-R BT.1788[13]标准进行,剔除了3组不可靠数据,共获得15组有效数据,满足ITU-R BT.500建议书中测试人员数量不少于15的要求. 虽然实验获取的15名测试人员的数据均有效且可靠,但仍有必要验证测试人员的数量是否符合实验需求. 数据饱和度是判断样本数量是否充足的有效方法[18]. 本文计算了实验数据之间的相关性,其结果如图6所示. 从图中可以看出,所有测试人员的相关性在测试人数达到13人时趋于稳定. 因此,本次实验选择15名测试人员具有合理性.

综上,本文所构建的数据库PAM-VQD的视频包括8种视频播放模式,每种模式下有10个视频,总的测试序列80个,加上10个源视频,即PAM-VQD数据库有90个视频. 不同模式下的视频持续时间和比特率因设置初始缓冲、比特率切换和卡顿事件不同而不同. PAM-VQD的构成如表2所示.

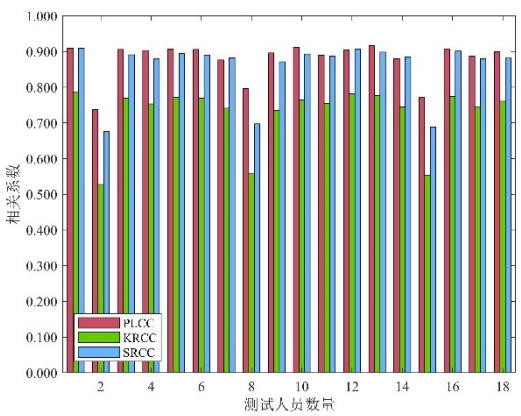


图5　每个测试人员与MOS值的相关性

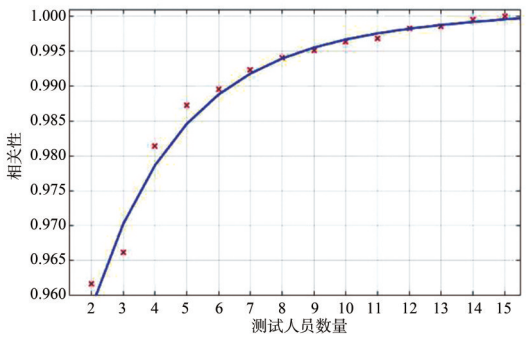


图6　MOS值饱和度变化曲线

3　基于注意力机制和长短期记忆网络的质量评价模型

由于人眼视觉感知特征是复杂且多样的,通过所提取到的手工特征虽能取得较好的预测结果,但在实际应用中可能缺乏普适性,即只能反映特定视频类型的失真. 同时,人类的记忆是连续的,且具有复杂的时间依赖关系. 这意味着视频观看过程中,记忆对用户QoE的影响是始终存在的,不仅会受到当前播放事件的影响,用户自身记忆特性也会对其产生影响. 残差网络在提取视频多层次的感知特征方面表现出优越的性能. 通过深层感知特征的捕获,以更好地描述感知对用户QoE的影响. 此外,文献[19]研究表明人类的记忆能力随着物体显著性的增加而增强,这表明视觉显著性较强的信息会优先进入人类记忆.

因此,本文提出参考感知与记忆的深度视频质量

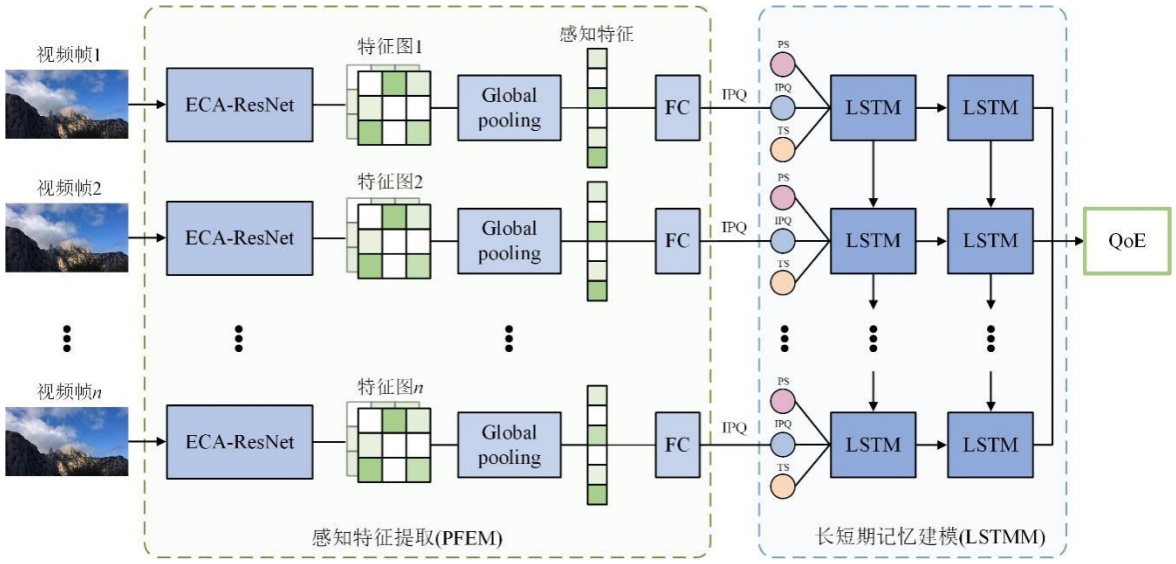


图7 PAM-DVQI框架图

表2 PAM-VQD的构成

参考视频	失真视频	播放类型	播放事件	视频时长	视频分辨率	评价尺度
10个	80个	8种	初始缓冲、卡顿、比特率波动	2 min	1 920×1 080	11分制

评价模型(Deep Video Quality Index with Perception And Memory,PAM-DVQI),该模型提取深层感知特征并反映记忆的时间依赖关系,以提升用户QoE的预测精度,总体框图如图7所示. 首先,将待测视频的所有视频帧依次输入本文所提出的感知特征提取网络,以生成显著区域增强后的特征图;接着,通过全局池化操作(Global Pooling,GP)获得感知聚合特征,并通过全连接层(Full Connection,FC)建立感知特征与用户QoE之间的映射关系,以获得用户的瞬时感知质量(Instantane-ous Perceptual Quality,IPQ);最后,将瞬时感知质量IPQ、播放状态(Playback Status,PS)以及自卡顿间隔(Time elapsed since last Stall,TS)共三个特征输入长短期记忆网络(Long Short-Term Memory,LSTM)[20]后端网络,对记忆效应的作用进行建模. 通过上述两个模型的统合,输出用户QoE的预测值.

3.1 基于ECA-ResNet的感知特征提取

为了获取视频的深层感知特征,并充分发挥深层网络模型的优势,本文以残差网络(Residual Network,ResNet)[21]为前端网络,并嵌入有效通道注意力(Efficient Channel Attention,ECA)模块[22],增强特征图中的视觉显著特征,以进一步提升特征的表达能力[23]. ECA-ResNet网络对输入的视频帧采取数据归一化预处理,4个残差单元共同作为ECA-ResNet网络的特征提取器,从视频帧中提取各类有助于用户QoE预测的感知特征. ECA-ResNet网络的具体框架如图8所示,主要分为输入层、4个残差单元和1个ECA模块.

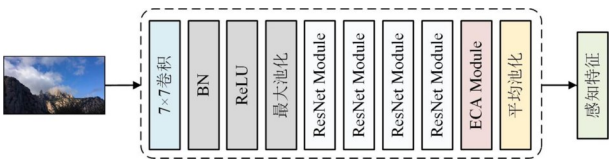


图8 ECA-ResNet网络框架图

输入层的主要作用是对原始数据进行简单的预处理,从而促进网络的训练. ECA-ResNet网络对输入的视频帧采取数据归一化预处理,该操作能有效防止由于输入数据范围大、单位不统一等原因,导致网络训练时间长、收敛速度慢.

4个残差单元共同作为ECA-ResNet网络的特征提取器,从视频帧中提取各类有助于用户QoE预测的感知特征. 每个残差单元包含两个残差模块,而残差模块主要由卷积层、正则化层和ReLU激活函数构成. 残差模块是ECA-ResNet的核心功能模块,可以显著降低网络优化的难度,从而使网络的深度能够更深,计算式如下:

$$y=F(x,\{W_i\})+x \tag{3}$$

其中,x和y表示残差模块的输入和输出特征图,W为卷积层的参数,函数$F(x,\{W_i\})$表示待学习的残差映射. 在每个残差模块中,输入特征图x经两次卷积操作、2次正则化操作以及一次ReLU操作后会学习到残差映射$F(x,\{W_i\})$,随后通过跳跃连接的方式,将输入x直接向后传递,并与残差映射$F(x,\{W_i\})$逐元素相加,从而得到特征图y. 通过4个残差单元共8个残差模块的特征提取后,可以获得每个视频帧的感知特征图X_t,其中,$t=1,2,\cdots,T$,T为视频帧的总数.

ECA模块首先对残差单元输出的尺寸为$H\times W\times C$的特征图X_t进行分通道全局平均池化(Global Average Pooling,GAP). 接着,将全局平均池化后获得的聚合特征Z_t,输入卷积核大小为k的1D快速卷积层,并利用以sigmoid函数$\sigma(\cdot)$为激活函数的全连接层,生成每个通道特征图的权重W_t. 最后,对原始特征图X_t进行加权操作$F_{\text{scale}}(\cdot)$,得到最终特征图$\widetilde{X}_t$,其模块结构图如图9所示,计算过程如式(4)~(6)所示.

$$Z_t=\frac{1}{W\times H}\sum\sum X_t \tag{4}$$

$$W_t=\sigma\left(\text{C1D}_k(Z_t)\right) \tag{5}$$

$$\widetilde{X}_t=F_{\text{scale}}(X_t,W_t) \tag{6}$$

本文将ECA模块放置于4个残差单元之后,可以选择性地强调有意义的特征通道,并抑制不重要的特征通道,进而优化网络的特征表达能力. 接着,对ECA-ResNet网络输出的特征图$\widetilde{X}_t$进行空域的GAP操作,计算式如下:

$$f_t=\text{GAP}(\widetilde{X}_t) \tag{7}$$

其中,f_t表示第t个视频帧的感知特征. 通过对所有视频帧的特征图执行上述操作,可获得所有视频帧对应的感知特征.

3.2　基于LSTM的长短期记忆建模

研究表明,QoE不仅取决于视频感知质量,还取决于视频播放过程中不同时刻的卡顿事件和比特率波动等因素的影响. 由于这些事件,HVS会产生滞后效应,即卡顿或比特率波动等事件产生的不良影响仍然保留在用户记忆中,即使视频后期质量得到提升,用户QoE也不会有明显改善. 滞后效应的存在意味着QoE具有非马尔可夫性[24-26],这是因为用户QoE不仅受到当前播放事件的影响,还会受到用户自身记忆特性的影响. 在以往的研究中,LSTM网络已被证明能够模拟序列数据中的长短期依赖关系[24]. 这是因为LSTM中具有特殊的门控结构和记忆单元,能使其有效捕获复杂的依赖关系,并解决了循环神经网络中梯度消失等问题. 因此,为了捕获用户的记忆特性,本文采用LSTM网络对QoE进行建模,以捕获用户QoE的长期依赖关系. 假设用户在t时刻的QoE实际值和预测值分别为$y(t)$和$\hat{y}(t)$,$x(t)$为输入特征集. 特征集$x(t)$的选择对于用户QoE的预测至关重要. 所输入的特征需要能够有效描述控制用户QoE变化的重要影响因素. 为此,本文根据上述所提数据库PAM-VQD中视频播放模式的设置方式,选择以下3种特征对用户QoE进行建模.

(1)IPQ:IPQ是指用户对当前视频内容的瞬时感知质量. 在以往的研究中,IPQ已成功应用于QoE预测[27]. 为了获取用户的瞬时感知质量,本文首先采用ECA-ResNet网络进行感知特征的提取. 接着,通过FC层建立感知特征与感知质量分数之间的映射关系,计算式如下:

$$\text{IPQ}=Wf_t+b \tag{8}$$

其中,f_t为感知特征,W和b为FC层中的参数. 通过上述计算,可输出视频帧的预测质量分数,即用户的瞬时感知质量.

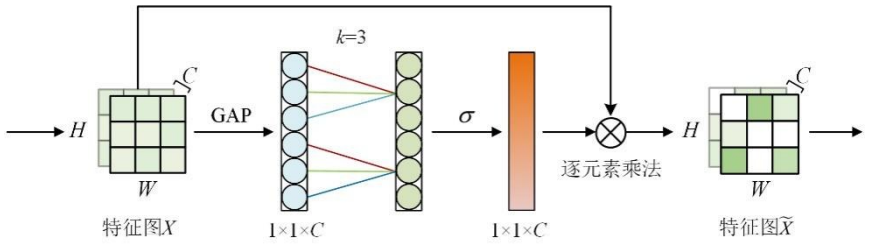


图9　ECA模块结构图

(2)PS:PS是一个二进制变量,用于显示视频当前是处于播放状态还是处于卡顿状态. 具体而言,当视频出现卡顿事件时,PS将显示为1. 当视频正常播放时,PS则显示为0. 此前的研究已经证明了PS对QoE预测的有效性[24].

(3)TS:TS是一个用于记录两次卡顿事件之间时间间隔的变量. 具体而言,该变量在每次卡顿事件结束后会开始进行时间的累加,直至下一次卡顿事件发生时才会归零重新开始记录. 由于在测试序列中设置了不同频率的卡顿事件,本文考虑采用TS特征来描述卡顿事件之间的时间间隔对用户QoE的影响. 该特征也被成功用于QoE预测[25].

综上,对于任意给定时刻t,本文使用特征向量$x(t)=\left(\text{IPQ}(t),\ \text{PS}(t),\ \text{TS}(t)\right)$来预测用户当前的QoE,表达式如下:

$$p\left(y(t)|y(t-1),y(t-2),\cdots,y(1)\right)\neq p\left(y(t)|y(t-1)\right) \tag{9}$$

其中,条件概率$p\left(y(t)|y(t-1),y(t-2),\cdots,y(1)\right)$表明用户QoE具有高阶时间依赖关系. 由于上述依赖关系较为复杂,单个LSTM可能无法对其进行有效建模. 通过实验表明,当LSTM网络配置为2个LSTM层和11个单元时,能够达到最佳的QoE预测性能. 因此,本文将LSTM

的层数设置为2，单元数设置为11. 整体网络$\mathrm{LSTM}_{l,d}$由多个LSTM单元串联而成，共同堆叠形成LSTM层，其中l表示LSTM层数，d表示LSTM单元数. 通过特征向量$x(t)$的输入，$\mathrm{LSTM}_{l,d}$可以在每个时刻t计算QoE的预测值$\hat{y}(t)$. 每个LSTM单元通过保持内部记忆状态$c(t)$来描述用户QoE中的长期依赖关系，并且记忆状态转换由输入特征向量$x(t)$进行驱动. 通过多个LSTM单元的建模，$\mathrm{LSTM}_{l,d}$网络可以学习控制用户记忆状态转换的潜在分布，并预测每个时刻的用户QoE，其表达式如下：

$$p\left(y(t)|y(t-1),y(t-2),\cdots,y(1)\right)=p\left(y(t);\ g(c(t))\right) \tag{10}$$

其中，$g(\cdot)$是一个可微函数，可以将LSTM单元中的记忆状态$c(t)$映射到潜在的QoE分布. 记忆状态的整体更新，可由下式进行计算：

$$c(t)=\mathrm{LSTM}_{l,d}^{c}(c(1:t-1),\hat{y}(1:t-1)),\quad \forall t>1 \tag{11}$$

其中，$c(1:t-1)$为$t-1$时刻的记忆状态，$\hat{y}(1:t-1)$为$t-1$时刻的QoE预测值，将上述两个值输入$\mathrm{LSTM}_{l,d}^{c}$网络可确定t时刻的记忆状态$c(t)$. 最后，用户QoE的预测值$\hat{y}(t)$可表示为

$$\hat{y}(t)=\mathrm{LSTM}_{l,d}^{o}(x(t),c(t-1)) \tag{12}$$

其中，$\mathrm{LSTM}_{l,d}^{o}$为t时刻的QoE预测值，$x(t)$为输入特征向量，$c(t-1)$为$t-1$时刻的记忆状态. 通过上述计算过程，可以构建感知和记忆与用户QoE之间的映射关系.

3.3 模型训练

（1）ECA-ResNet网络的训练

基于本文提出的PAM-VQD数据库训练ECA-ResNet网络，将所有视频样本以8∶2的比例按视频内容进行划分，80%的数据用于训练，其余20%用于测试. 为了获取用户的瞬时主观质量，本文将所有视频都以视频帧的形式输入ECA-ResNet网络进行训练. 每个视频帧的训练标签为所属视频的MOS值. 由于PAM-VQD中原始视频和测试序列的总数为90个，视频样本数量较少，本文利用了图像资源丰富的ImageNet图像数据库，在该数据库的基础上预先训练好的ResNet模型，通过迁移学习的方法，使用数量较少的视频样本将获取到的权重参数进行微调训练，避免ECA-ResNet网络在训练时产生过拟合的现象[26]. ECA-ResNet网络的优化目标：最小化MOS值与视频帧预测得分之间的误差，并采用带有L2正则项的均方误差（Mean Square Error，MSE）函数对该目标进行约束. 计算式如下：

$$\mathrm{Loss}=\frac{1}{N}\sum_{i=0}^{N}||y_i-y_i'||^2+\lambda||\theta||^2 \tag{13}$$

其中，y_i表示输入视频帧的MOS值，y_i'表示第i帧的预测分数，N为输入视频帧的总数，λ为正则化系数，大小设置为0.000 5.

本文采用SGD优化器对该目标函数进行优化，并引入动量优化算法加速梯度的下降，将动量最大值设定为0.9. 其中，SGD的初始学习率设为0.001，并采用回调函数进行学习率的优化，即当损失值超过10轮都没有得到改善时，该函数会自动降低学习率，从而使模型达到最优状态.

（2）LSTM网络的训练

LSTM网络在PAM-VQD数据库上进行训练和测试. 以8∶2的比例按视频内容进行数据划分，80%用于训练，20%用于测试. 将所有视频对应的IPQ、PS和TS特征输入LSTM网络进行训练. 每个视频的训练标签为该视频的MOS值. 在训练过程中，输入特征需要通过输入层，并以适当的时间步长输入到网络中. 本文参照文献[28]中使用的三阶时间依赖关系，将时间步长设置为4. 视频的采样率设置为1 s. 因此，在测试过程中，用户QoE预测值$\hat{y}(t)$将以每秒的方式动态输出，即时间步长为1，并在时间密集分布层的末端获得. 在LSTM网络的训练过程中，采用MSE函数和Adam优化器进行目标约束.

4 实验测试与分析

4.1 与VQA算法的性能比较与分析

为了验证PAM-DVQI模型在预测用户QoE方面的准确性，本文在数据库PAM-VQD以及LIVE[10]、Waterloo SQoE-Ⅰ[7]和LFOVIA QoE[29]视频数据库上分别进行对比实验. 对比算法包括通用的全参考视频质量评价（Full-Reference VQA，FR-VQA）算法：PSNR、SSIM和MS-SSIM，半参考视频质量评价（Reduce-Reference VQA，RR-VQA）算法：SpEED-QA[30]和STRRED[31]，无参考视频质量评价（No-Reference VQA，NR-VQA）算法：TLVQM[32]、BRISQUE[33]、NIQE[34]和BIQI[35]，以及基于深度学习的算法：HSTVQA[36]和DeepVQA[37]. 在本次实验过程中，网络的基本参数、结构的设定、训练集和测试集的分配比例、相关性的计算方法等方面都保持一致. 在表3~表6分别列出了各VQA算法的皮尔逊线性相关系数（Pearson Linear Correlation Coefficient，PLCC）、肯德尔等级相关系数（Kendall Rank-Order Correlation Coefficient，KRCC）和斯皮尔曼等级相关系数（Spearman Rank-order Correlation Coefficient，SRCC）性能.

由表3至表6可知：第一，基于PAM-VQD数据库，本文提出的PAM-DVQI模型的PLCC、KRCC和SRCC性能均高于其他VQA对比算法，表明PAM-DVQI模型在不同播放事件下的QoE预测性能与用户保持了较高的一致性. 此外，基于深度学习的HSTVQA和DeepVQA算法也取得了较好的性能，这是由于这两种算法都提取了视频多层次的感知特征，与其他VQA算法相比，能够更好地反映视频的内容信息. 然而，这两种算法无法

表3　各VQA算法在PAM-VQD上的评价性能

算法类型	算法	PLCC	KRCC	SRCC
FR	PSNR	0.611 7	0.357 6	0.493 3
	SSIM	0.700 7	0.455 5	0.585 3
	MS-SSIM	0.698 4	0.384 5	0.502 9
RR	SpEED-QA	0.596 1	0.369 0	0.468 3
	STRRED	0.626 7	0.335 4	0.463 7
NR	TLVQM	0.567 7	0.577 4	0.525 7
	BRISQUE	0.522 2	0.283 5	0.398 1
	NIQE	0.594 1	0.198 3	0.289 9
	BIQI	0.600 5	0.433 1	0.579 8
	HSTVQA	0.625 5	0.530 6	0.574 7
	DeepVQA	0.760 2	0.547 8	0.711 4
本文算法	PAM-DVQI	**0.883 4**	**0.678 2**	**0.812 6**

表4　各VQA算法在LIVE数据库上的评价性能

算法类型	算法	PLCC	KRCC	SRCC
FR	PSNR	0.575 0	0.448 8	0.531 2
	SSIM	0.621 2	0.458 8	0.593 6
	MS-SSIM	0.744 0	0.532 5	0.695 6
RR	SpEED-QA	0.811 6	0.609 8	0.772 9
	STRRED	0.826 3	0.628 5	0.792 1
NR	TLVQM	0.576 5	0.565 5	0.542 7
	BRISQUE	0.196 5	0.106 7	0.160 5
	NIQE	0.230 4	0.069 2	0.103 0
	BIQI	0.304 5	0.189 7	0.187 5
	HSTVQA	0.749 2	0.753 9	0.732 9
	DeepVQA	0.881 2	0.710 9	0.894 6
本文算法	PAM-DVQI	**0.896 4**	**0.733 6**	**0.918 4**

表5　各VQA算法在Waterloo SQoE-Ⅰ数据库上的评价性

算法类型	算法	PLCC	KRCC	SRCC
FR	PSNR	0.660 1	0.521 8	0.654 3
	SSIM	0.841 3	0.668 4	0.811 6
	MS-SSIM	0.812 7	0.643 4	0.798 6
RR	SpEED-QA	0.831 6	0.690 1	0.802 9
	STRRED	0.846 3	0.718 5	0.812 1
NR	TLVQM	0.548 9	0.547 1	0.467 5
	BRISQUE	0.526 5	0.390 5	0.471 8
	NIQE	0.310 2	0.223 5	0.263 4
	BIQI	0.438 9	0.308 6	0.382 0
	HSTVQA	0.716 4	0.713 1	0.677 7
	DeepVQA	0.854 1	0.728 7	0.823 1
本文算法	PAM-DVQI	**0.910 8**	**0.775 9**	**0.892 3**

描述不同播放事件对用户QoE的影响,因此与本文提出的PAM-DVQI模型性能仍有一定差距. 第二,对于LIVE视频数据库,由于LIVE数据库中只包含压缩失真和传输失真,不存在卡顿或比特率波动等播放事件,此时的PAM-DVQI模型仅通过ECA-ResNet网络计算当前视频的感知质量. 从表4中可以看出,本文提出的PAM-DVQI模型以及基于深度学习的HSTVQA和Deep-VQA算法的PLCC、KRCC和SRCC性能均优于其他对比算法,能够较好地预测视频感知质量. 此外,NR-VQA算法的预测性能较弱,这是由于NR-VQA算法没有原始视频的信息加以参考,而直接利用失真视频的特征预测用户的感知质量,因此评估结果的稳定性较弱. 第三,对于Waterloo SQoE-Ⅰ数据库,由于该数据库中测试序列的时长较短,视频复杂度不高,且播放事件出现的位置和数量都较为固定,因此各个VQA算法的预测性能都有显著提升. 第四,对于LFOVIA QoE数据库,由于该数据库包含多种视频类型,且播放事件出现的位置和时长较为随机,使得用户QoE的准确评估具有一定难度.

表6　各VQA算法在LFOVIA QoE数据库上的评价性能

算法类型	算法	PLCC	KRCC	SRCC
FR	PSNR	0.434 7	0.326 8	0.384 0
	SSIM	0.604 8	0.513 9	0.567 4
	MS-SSIM	0.667 5	0.571 5	0.612 6
RR	SpEED-QA	0.746 0	0.614 1	0.693 0
	STRRED	0.776 7	0.638 5	0.708 1
NR	TLVQM	0.592 4	0.582 5	0.542 9
	BRISQUE	0.409 1	0.290 8	0.348 4
	NIQE	0.386 7	0.265 1	0.321 4
	BIQI	0.368 5	0.254 9	0.298 7
	HSTVQA	0.733 8	0.729 2	0.676 2
	DeepVQA	0.801 2	0.640 9	0.739 6
本文算法	PAM-DVQI	**0.868 5**	**0.727 2**	**0.804 5**

由表6可知,各类VQA算法的性能都出现了不同程度的下降. 与其他算法相比,本文提出的PAM-DVQI模型保持了较好的预测性能,这表明通过视觉注意力模块提取的感知特征以及时域中时间依赖关系的捕获,能够适应变化类型多样的视频序列,从而实现对用户QoE的有效预测.

为进一步评估本文提出的PAM-DVQI模型在4个数据库上所表现出来的整体性能,在表7给出各个算法在4个数据库的加权平均PLCC、KRCC和SRCC. 从表7可以看出,本文提出的PAM-DVQI模型的性能优于其他的VQA对比算法,这表明本文提出的PAM-DVQI模型对不同的视频数据库都体现出高相关性,因此对于用户QoE的评估具有显著优势.

4.2　与QoE算法的性能比较与分析

为了进一步验证本文提出的PAM-DVQI模型在预测用户QoE方面的准确性,本文在LFOVIA QoE、LIVE NETFLIX[6]和LIVE QoE[38]视频数据库上与现有的主流视

表7 各VQA算法在四个数据库上的总体性能对比

算法类型	算法	PLCC	KRCC	SRCC
FR	PSNR	0.604 6	0.452 1	0.562 2
	SSIM	0.723 0	0.547 2	0.678 0
	MS-SSIM	0.757 4	0.553 9	0.695 9
RR	SpEED-QA	0.775 7	0.599 4	0.723 9
	STRRED	0.794 6	0.613 1	0.734 5
NR	TLVQM	0.546 4	0.568 1	0.519 7
	BRISQUE	0.405 3	0.267 8	0.343 9
	NIQE	0.340 5	0.170 4	0.218 9
	BIQI	0.417 0	0.286 6	0.345 3
	HSTVQA	0.715 5	0.706 7	0.640 4
	DeepVQA	0.842 1	0.683 2	0.829 7
本文算法	PAM-DVQI	**0.897 6**	**0.740 2**	**0.879 7**

频QoE预测模型进行对比实验,如表8所示. 对比算法包括SVR-QoE[37]、NLSS-QoE[39]、LSTM-QoE[28]、NARX[40]和HW[38]. 在本次实验过程中,网络的基本参数、结构的设定、训练集和测试集的分配比例、相关性的计算方法等方面都保持一致. 在表8中,分别列出了各QoE算法和VQA算法的PLCC和SRCC性能. 实验结果表明,与其他算法相比,本文提出的PAM-DVQI模型保持了较好的预测性能,这表明通过视觉注意力模块提取的感知特征以及时域中时间依赖关系的捕获,能够适应变化类型多样的视频序列,从而实现对用户QoE的有效预测.

4.3 消融实验

为了分析所提出的PAM-DVQI模型中相关模块的性能增益,消融实验是必要的. 消融实验结果如表9至表13所示,其中加粗形式表示最佳的性能.

表8 各QoE算法在三个数据库上的评价性能

VQA 算法	QoE Model	LFOVIA QoE		QoE Model	LIVE NETFLIX		QoE Model	LIVE QoE	
		PLCC	SRCC		PLCC	SRCC		PLCC	SRCC
STRRED	SVR-QoE	0.686	0.648	NARX	0.621	0.557	HW	0.742	0.732
MS-SSIM		0.737	0.683		0.598	0.549		0.727	0.705
NIQE		0.797	0.750		0.605	0.537		0.511	0.509
STRRED	NLSS-QoE	0.767	0.685	NLSS-QoE	0.655	0.483	NLSS-QoE	0.723	0.707
MS-SSIM		0.781	0.680		0.583	0.420		0.883	0.871
NIQE		0.825	0.794		0.527	0.300		0.211	0.189
STRRED	LSTM-QoE	0.800	0.730	LSTM-QoE	0.802	0.714	LSTM-QoE	0.892	**0.893**
MS-SSIM		0.786	0.712		0.745	0.689		0.344	0.417
NIQE		0.858	0.808		0.683	0.609		0.473	0.475
STRRED	PAM-DVQI	**0.873**	**0.815**	PAM-DVQI	**0.831**	**0.793**	PAM-DVQI	**0.915**	0.887
MS-SSIM		0.791	0.724		0.802	0.736		0.832	0.811
NIQE		0.685	0.628		0.615	0.578		0.697	0.639

(1)ECA模块的性能增益

为了进一步验证ResNet网络的优越性能,本文引入流行网络VGG[41]、AlexNet[42]替换ResNet网络. 表9给出了其性能对比. 实验结果表明,以ResNet作为前端网络的总体性能均更高,证明了本文所选取的ResNet更加有效地提取视频的感知特征,从而提高QoE的预测性能.

为了验证注意力模块对PAM-DVQI模型的增益效果,表10给出了本文所提出的PAM-DVQI模型在加入ECA模块前后在4个视频数据库上的PLCC和SRCC性能. 同时,表中最后一列给出了PAM-DVQI模型在这4个视频数据库上的总体性能,用加权平均PLCC和SRCC表示,每个数据库的权重取决于该数据库中包含视频的数量,最优性能以加粗形式给出. 结果显示,与未加入ECA模块的算法性能比较,嵌入ECA模块的PAM-DVQI模型的总体性能均更高,证明了ECA模块能够模拟显著信息对用户视觉记忆的引导作用,从而提升PAM-DVQI模型对用户QoE的预测性能.

表9 不同骨干网络嵌入ECA模块的PLCC、KRCC和SRCC性能对比

相关性	前端网络	PAM-VQD	LIVE	Waterloo SQoE-Ⅰ	LFOVIA QoE
PLCC	VGG	0.783 5	0.863 1	0.876 2	0.808 5
	AlexNet	0.753 1	0.890 4	0.867 8	0.794 1
	ResNet	**0.883 4**	**0.896 4**	**0.910 8**	**0.868 5**
KRCC	VGG	0.563 2	0.681 7	0.740 1	0.643 6
	AlexNet	0.538 4	0.726 3	0.734 6	0.639 3
	ResNet	**0.678 2**	**0.733 6**	**0.775 9**	**0.727 2**
SRCC	VGG	0.728 5	0.867 0	0.857 4	0.740 4
	AlexNet	0.698 7	0.902 3	0.834 6	0.728 0
	ResNet	**0.812 6**	**0.918 4**	**0.892 3**	**0.804 5**

为了进一步证明ECA模块使用的合理性,本文在表11中给出嵌入ECA模块与其它注意力模块后的PLCC、KRCC和SRCC性能对比以及在不同注意力模块下的总体性能对比. 实验结果表明,在嵌入不同的注意力模块后,模型

表10 ECA模块嵌入前后的PLCC、KRCC和SRCC性能对比

相关性	含ECA模块	PAM-VQD	LIVE	Waterloo SQoE-Ⅰ	LFOVIA QoE
PLCC	否	0.8643	0.876 2	0.899 8	0.824 5
	是	**0.883 4**	**0.896 4**	**0.910 8**	**0.868 5**
KRCC	否	0.651 6	0.719 1	0.758 5	0.678 0
	是	**0.678 2**	**0.733 6**	**0.775 9**	**0.727 2**
SRCC	否	0.788 0	0.887 6	0.872 3	0.775 2
	是	**0.812 6**	**0.918 4**	**0.892 3**	**0.804 5**

的性能均有所提升,这说明注意力机制在人类视觉认知中具有重大影响. 此外,与其他注意力模块相比,ECA模块的PLCC、KRCC和SRCC性能均更高,总体性能提升幅度更大. 通过上述实验,证明了ECA模块的有效性以及对HVS的作用,也表明嵌入ECA模块能有效改善PAM-DVQI模型评估用户QoE的性能.

表11 不同注意力模块的性能增益对比

相关性	模块名称	PAM-VQD	LIVE	Waterloo SQoE-Ⅰ	LFOVIA QoE
PLCC	无	0.864 3	0.876 2	0.899 8	0.824 5
	CBAM	0.881 2	0.872 4	**0.918 8**	0.855 8
	SE	0.870 8	0.889 2	0.907 9	0.844 7
	ECA	**0.883 4**	**0.896 4**	0.910 8	**0.868 5**
KRCC	无	0.651 6	0.719 1	0.758 5	0.678 0
	CBAM	0.682 5	0.724 1	**0.777 8**	0.722 3
	SE	0.671 8	**0.735 1**	0.767 1	0.713 5
	ECA	**0.678 2**	0.733 6	0.775 9	**0.727 2**
SRCC	无	0.788 0	0.887 6	0.872 3	0.775 2
	CBAM	0.792 5	0.901 6	**0.906 4**	0.797 8
	SE	0.801 5	0.913 8	0.882 6	0.782 4
	ECA	**0.812 6**	**0.918 4**	0.892 3	**0.804 5**

(2)IPQ、PS和TS特征对PAM-DVQI模型的性能增益

为了评估IPQ、PS和TS特征对PAM-DVQI模型的性能增益,本文研究了各个特征对QoE预测的贡献. 具体来说,本文将上述3个特征输入到LSTM-QoE网络,并在所提数据库上验证其对用户QoE的预测性能. 本文采用实验所证明的最佳网络配置进行实验,即2层11个单元的LSTM网络. IPQ特征采用的是本文上述提出的ECA-ResNet网络输出的视频感知质量预测值. 表12给出了不同特征组合下的QoE预测值与真实值之间的PLCC相关性. 各种特征组合如下:(a)IPQ,(b)PS,(c)TS,(d)IPQ+PS,(e)IPQ+TS,(f)PS+TS,(g)IPQ+PS+TS. 实验结果表明,当所有特征都用于QoE预测时,PLCC性能最佳. 因此,采用IPQ、PS和TS特征进行用户QoE预测是合理有效的.

表12 不同特征组合方式下的性能对比

特征	PLCC	特征	PLCC
IPQ	0.801 8	IPQ+TS	0.746 4
PS	0.624 5	PS+TS	0.691 0
TS	0.543 2	IPQ+PS+TS	0.883 4
IPQ+PS	0.753 6	—	—

(3)LSTM参数对QoE预测性能的影响

为了评估LSTM的层数和单元数对QoE预测性能的影响,图10显示了LSTM层数和单元数在不同配置下的预测性能. 实验结果表明:第一,在增加LSTM网络的单元数和层数后,用户QoE的预测性能得到了显著提升,这说明单个LSTM网络无法有效捕获QoE过程中复杂的时间依赖关系;第二,当LSTM层数在两层或以上,且LSTM单元数达到11个时,网络的预测性能趋于饱和;第三,当LSTM层数达到两层时,性能仅产生了微小提升,当超过3层时,网络性能则开始下降. 这是由于随着LSTM层数和单元数的增长,网络会变得越来越庞大,容易产生过拟合的问题,导致网络训练变得更加困难,从而降低预测性能. 图上结果显示,当LSTM网络配置为2个LSTM层和11个单元时,能够达到最佳的QoE预测性能. 因此,本文将LSTM的层数设置为2,单元数设置为11.

(4)跨数据集验证

为了进一步验证本文所提算法的泛化性能,本文进行跨数据集验证,分别在LIVE、Waterloo SQoE-Ⅰ、LFOVIA QoE和本文所构建的PAM-VQD数据库上训练,对于在每个数据集上训练的模型,分别以其他3个数据集为测试集,测试其性能,结果如表13所示,其中"—"处为同数据库实验,相关实验结果请看表3~表6,

表13 PAM-DVQI跨数据集验证结果

测试库训练库		LIVE	Waterloo SQoE-Ⅰ	LFOVIA QoE	PAM-VQD
LIVE	PLCC	—	0.453 6	0.476 3	0.357 4
	SRCC	—	0.468 2	0.465 8	0.449 5
Waterloo SQoE-Ⅰ	PLCC	0.619 8	—	0.597 8	0.574 3
	SRCC	0.621 3	—	0.602 3	0.587 9
LFOVIA QoE	PLCC	0.651 2	0.633 9	—	0.602 3
	SRCC	0.669 7	0.646 7	—	0.623 6
PAM-VQD	PLCC	0.723 2	0.703 2	0.687 4	—
	SRCC	0.731 5	0.713 5	0.693 8	—

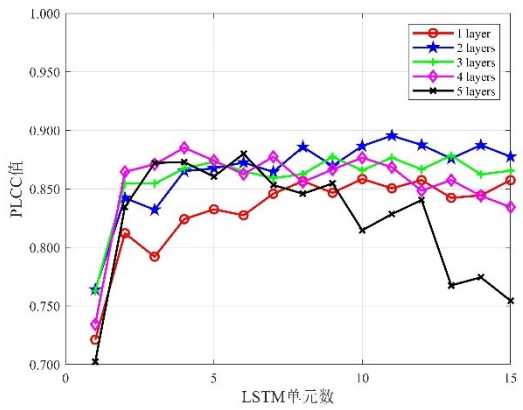


图 10 不同 LSTM 层数和单元数的性能对比

在此不再赘述.

跨库验证实验结果表明,PAM-VQD 算法对于不同的数据库具有较为良好的泛化能力. LIVE 数据库中测试序列不存在卡顿或比特率波动等播放事件,Waterloo SQoE-Ⅰ数据库中测试序列时长较短且具有较为固定的播放事件位置和数量,因此在这两个库上训练的模型表达能力不足. 而对于具有多种视频类型的 LFOVIA QoE 和 PAM-VQD 数据库,播放事件出现的位置和时长较为随机,算法的性能相对更好.

5 结论

本文通过主观实验探究了视觉感知和记忆特性对用户 QoE 的影响,构建了基于用户 QoE 的视频主观数据库,该数据库为后续研究提供数据支撑. 基于所提出的 PAM-VQD 数据库,本文将视觉注意力机制和记忆的时间依赖关系应用于视频质量评价中,实现对用户 QoE 的高精度预测;最后,本文利用与传统 VQA 算法和基于深度学习的 VQA 算法的对比实验和消融实验验证 PAM-DVQI 模型性能的优越性. 实验结果表明,本文提出的 PAM-DVQI 的总体性能优于所有的 VQA 对比算法,且在 4 个不同的视频数据库均表现出高预测性能,具有良好鲁棒性. 该模型可实现用户 QoE 的高精度预测,对相关 VQA 算法的性能评估和优化具有指导作用,有较高的实际应用价值. 此外,本文工作为基于感知和记忆的视频质量特别是长视频质量评测提供了新的方法和思路.

参考文献

[1] 曹燕, 董一鸿, 邬少清, 等. 动态网络表示学习研究进展[J]. 电子学报, 2020, 48(10): 2047-2059.
CAO Y, DONG Y H, WU S Q, et al. Dynamic network representation learning: A review[J]. Acta Electronica Sinica, 2020, 48(10): 2047-2059. (in Chinese)

[2] 易令, 李泽平. 基于深度强化学习的码率自适应算法研究[J]. 电子学报, 2022, 50(5): 1192-1200.
YI L, LI Z P. Research of adaptive bitrate algorithm based on deep reinforcement learning[J]. Acta Electronica Sinica, 2022, 50(5): 1192-1200. (in Chinese)

[3] 高敏娟, 党宏社, 魏立力, 等. 全参考图像质量评价回顾与展望[J]. 电子学报, 2021, 49(11): 2261-2272.
GAO M J, DANG H S, WEI L L, et al. Review and prospect of full reference image quality assessment[J]. Acta Electronica Sinica, 2021, 49(11): 2261-2272. (in Chinese)

[4] GHAFIL A S, ALI I H. Video streaming forecast quality of experience- A survey[C]//2021 1st Babylon International Conference on Information Technology and Science (BICITS). Piscataway: IEEE, 2021: 299-304.

[5] LI L, CHEN P, LIN W, et al. From whole video to frames: Weakly-supervised domain adaptive continuous-time QoE evaluation[J]. IEEE Transactions on Image Processing, 2022, 31: 4937-4951.

[6] BAMPIS C G, ZHI LI, MOORTHY A K, et al. Study of temporal effects on subjective video quality of experience[J]. IEEE Transactions on Image Processing, 2017, 26(11): 5217-5231.

[7] DUANMU Z, REHMAN A, WANG Z. A quality-of-experience database for adaptive video streaming[J]. IEEE Transactions on Broadcasting, 2018, 64(2): 474-487.

[8] SAWABE A, IWAI T. A QoS model to identify required QoS for guaranteeing quality of Internet video streaming services[C]//ICC 2021 - IEEE International Conference on Communications. Piscataway: IEEE, 2021: 1-6.

[9] WAHAB A, AHMAD N, SCHORMANS J. Direct propagation of network QoS distribution to subjective QoE for video on demand applications using VP9 codec[C]//2020 International Wireless Communications and Mobile Computing (IWCMC). Piscataway: IEEE, 2020: 929-933.

[10] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C, et al. Study of subjective and objective quality assessment of video[J]. Journal of Advanced Pharmaceutical Technology & Research, 2010, 19(6): 1427-1441.

[11] MOORTHY A K, CHOI L K, BOVIK A C, et al. Video quality assessment on mobile devices: Subjective, behavioral and objective studies[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 652-671.

[12] WU W, LIU Z Z, CHEN Z Z, et al. No-reference video quality assessment based on similarity map estimation[C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2020: 181-185.

[13] ITU-R. Methodology for the Subjective Assessment of

Video Quality in Multimedia Applications: BT. 1788: 2007[S/OL]. [2023-8-18]. https://www.itu.int/rec/R-REC-BT.1788-0-200701-W/en.

[14] KIANI MEHR S, JOGALEKAR P, MEDHI D. Moving QoE for monitoring DASH video streaming: Models and a study of multiple mobile clients[J]. Journal of Internet Services and Applications, 2021, 12(1): 1-26.

[15] REHMAN A, WANG Z. Perceptual experience of time-varying video quality[C]//2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). Piscataway: IEEE, 2013: 218-223.

[16] ITU-R. Methodology for the Subjective Assessment of the Quality of Television Pictures: BT. 500-13: 2012[S/OL]. [2023-08-18]. https://www.itu.int/rec/R-REC-BT.500-13-201201-S/en.

[17] ITU-T. Subjective Video Quality Assessment Methods for Multimedia Applications: Rec. P. 910: 2008[S/OL]. [2023-08-18]. https://www.itu.int/rec/t-rec-p.910/en

[18] ZHANG W, LIU H T. Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2424-2437.

[19] 梁永生, 柳伟, 周莺, 等. 基于视觉显著计算的视频流媒体渐进式表达方法[J]. 电子学报, 2017, 45(7): 1567-1575.
LIANG Y S, LIU W, ZHOU Y, et al. An approach to progressive description of video streaming based on visual saliency computation[J]. Acta Electronica Sinica, 2017, 45(7): 1567-1575. (in Chinese)

[20] TRAN H T T, NGUYEN D V, NGOC N P, et al. Overall quality prediction for HTTP adaptive streaming using LSTM network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(8): 3212-3226.

[21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2016: 770-778.

[22] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11534-11542.

[23] GU J P, HU J J, JIANG L, et al. Object detection of overhead transmission lines based on improved YOLOv5s[C]//2022 12th International Conference on Power and Energy Systems (ICPES). Piscataway: IEEE, 2022: 388-392.

[24] SHI W J, SUN Y J, PAN J Q. Continuous prediction for quality of experience in wireless video streaming[J]. IEEE Access, 2019, 7: 70343-70354.

[25] DONAHUE J, HENDRICKS L A, ROHRBACH M, et al. Long-term recurrent convolutional networks for visual recognition and description[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 677-691.

[26] CHEN P, LI L, WU J, et al. Contrastive self-supervised pre-training for video quality assessment[J]. IEEE Transactions on Image Processing, 2022, 31: 458-471.

[27] BAMPIS C G, LI Z, KATSAVOUNIDIS I, et al. Recurrent and dynamic models for predicting streaming video quality of experience[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3316-3331.

[28] ESWARA N, ASHIQUE S, PANCHBHAI A, et al. Streaming video QoE modeling and prediction: A long short-term memory approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(3): 661-673.

[29] JOSEPH V, DE VECIANA G. NOVA: QoE-driven optimization of DASH-based video delivery in networks[C]//IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. Piscataway: IEEE, 2014: 82-90.

[30] BAMPIS C G, GUPTA P, SOUNDARARAJAN R, et al. SpEED-QA: Spatial efficient entropic differencing for image and video quality[J]. IEEE Signal Processing Letters, 2017, 24(9): 1333-1337.

[31] SOUNDARARAJAN R, BOVIK A C. Video quality assessment by reduced reference spatio-temporal entropic differencing[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(4): 684-694.

[32] KORHONEN J. Two-level approach for no-reference consumer video quality assessment[J]. IEEE Transactions on Image Processing, 2019, 28(12): 5923-5938.

[33] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on Image Processing, 2012, 21(12): 4695-4708.

[34] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “Completely blind” image quality analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.

[35] MOORTHY A K, BOVIK A C. A two-step framework for constructing blind image quality indices[J]. IEEE Signal Processing Letters, 2010, 17(5): 513-516.

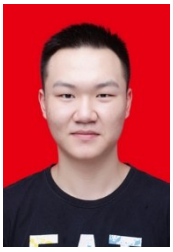
[36] SHEN W H, ZHOU M L, LIAO X R, et al. An end-to-end no-reference video quality assessment method with hierarchical spatiotemporal feature representation[J]. IEEE Transactions on Broadcasting, 2022, 68(3): 651-660.

[37] KIM W, KIM J, AHN S, et al. Deep video quality assessor: From spatio-temporal visual sensitivity to a convolutional neural aggregation network[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 224-241.
[38] CHEN C, CHOI L K, DE VECIANA G, et al. Modeling the time—Varying subjective quality of HTTP video streams with rate adaptations[J]. IEEE Transactions on Image Processing, 2014, 23(5): 2206-2221.
[39] ESWARA N, MANASA K, KOMMINENI A, et al. A continuous QoE evaluation framework for video streaming over HTTP[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(11): 3236-3250.
[40] BAMPIS CHRISTOS G, ZHI L, BOVIK ALAN C. Continuous prediction of streaming video QoE using dynamic networks[J]. IEEE Signal Processing Letters, 2017, 24(7): 1083-1087.
[41] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2015-04-10]. https://arxiv.org/pdf/1409.1556v6.
[42] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012, 60: 84-90.

作者简介

林丽群 女，1980 年出生，福建莆田人．2007 年和 2019 年分别获得福州大学硕士和博士学位．目前任福州大学副教授，硕士生导师，主要研究方向为视频质量评价、视频编码和计算机视觉等．
E-mail: lin_liqun@fzu.edu.cn

暨书逸 女，1997 年出生，福建武夷山人．2016 年获得福建师范大学学士学位，2022 年获得福州大学硕士学位，主要研究方向为视频质量评价和计算机视觉等．
E-mail: 973060009@qq.com



何嘉晨 男，2001 年出生，湖北武汉人．2023 年获得福州大学学士学位．主要研究方向为视频质量评价和计算机视觉等．
E-mail: hjc_18995643869@126.com

赵铁松 男，1984 年出生，河北衡水人．2006 年获得中国科学技术大学学士学位，2011 年获得香港城市大学博士学位．目前任福州大学教授、博士生导师，“媒体信息智能处理与无线传输”福建省重点实验室主任，在相关领域有十余年的研发经验，曾获得国家级青年人才项目及若干省级人才项目支持．同时担任 IEEE 高级会员、中国计算机学会高级会员、IET Electronics Letters 编委等，并入选团中央指导下的中国青年科技工作者协会会员，当选第六届理事．主要研究方向为多媒体通信系统、人工智能和视频编码等．中国电子学会会员编号：E190014840S.
E-mail: t.zhao@fzu.edu.cn

陈炜玲 女，1991 年出生，福建福州人．2009 年和 2018 年分别获得厦门大学学士和博士学位．目前任福州大学副教授，硕士生导师，主要研究方向为智慧海洋、计算机视觉、水下信号处理等．中国电子学会会员编号：E190157944M.
E-mail: weiling.chen@fzu.edu.cn

郭宗明 男，1966 年出生，江苏盐城人．1987 年和 1994 年分别获得北京大学学士和博士学位．目前任北京大学研究员，博士生导师，北京大学王选计算机研究所副所长，电子出版新技术国家工程研究中心主任，教育部中国文字字体设计与研究中心主任，主要研究方向为数字视频处理、数字水印、数字版权保护、计算机辅助卡通动画、视频编码、视频质量评价等．
E-mail: guozongming@pku.edu.cn