基于多层 BP 神经网络的无参考视频质量客观评价

姚军财[1,2] 申 静[1] 黄陈蓉[1]

摘 要 机器学习在视频质量评价 (Video quality assessment, VQA) 模型回归方面具有较大的优势, 能够较大地提高构建模型的精度. 基于此, 设计了合理的多层 BP 神经网络, 并以提取的失真视频的内容特征、编解码失真特征、传输失真特征及其视觉感知效应特征参数为输入, 通过构建的数据库中的样本对其进行训练学习, 构建了一个无参考 VQA 模型. 在模型构建中, 首先采用图像的亮度和色度及其视觉感知、图像的灰度梯度期望值、图像的模糊程度、局部对比度、运动矢量及其视觉感知、场景切换特征、比特率、初始时延、单次中断时延、中断频率和中断平均时长共 11 个特征, 来描述影响视频质量的 4 个主要方面, 并对建立的两个视频数据库中的大量视频样本, 提取其特征参数; 再以该特征参数作为输入, 对设计的多层 BP 神经网络进行训练, 从而构建 VQA 模型; 最后, 对所提模型进行测试, 同时与 14 种现有的 VQA 模型进行对比分析, 研究其精度、复杂性和泛化性能. 实验结果表明: 所提模型的精度明显高于其 14 种现有模型的精度, 其最低高出幅度为 4.34 %; 且优于该 14 种模型的泛化性能, 同时复杂性处于该 15 种模型中的中间水平. 综合分析所提模型的精度、泛化性能和复杂性表明, 所提模型是一种较好的基于机器学习的 VQA 模型.

关键词 视频质量评价, 神经网络, 时延, 视频内容

引用格式 姚军财, 申静, 黄陈蓉. 基于多层 BP 神经网络的无参考视频质量客观评价. 自动化学报, 2022, **48**(2): 594−607

DOI 10.16383/j.aas.c190539

No Reference Video Quality Objective Assessment Based on Multilayer BP Neural Network

YAO Jun-Cai[1,2] SHEN Jing[1] HUANG Chen-Rong[1]

Abstract Machine learning has a great advantage in the regression of video quality assessment (VQA) model and can greatly improve the accuracy of built model. To this end, a reasonable BP neural network is designed, and taking the feature values of the distorted video contents, code and decode distortion, transmission distortion, and visual perception effect as inputs, a no reference VQA model is constructed by training them with the samples of the built video databases. In modeling, firstly, 11 features are used to describe the four main factors that affect video quality, which are the brightness and chroma of image and their visual perception, the gray gradient expectation of image, the blur degree of image, the local contrast, the motion vectors and their visual perception, the scene switching feature, the bitrate, the initial delay, the single interrupt delay, the interrupt frequency and the average time of interrupt. And the feature parameters of a large number of video samples in the two video databases established are extracted. Then by using these feature parameters as inputs, the BP neural network is trained to construct our VQA model. Finally, the proposed model is tested and compared with 14 existing VQA models to study its accuracy, complexity and generalization performance. The experimental results show that the accuracy of the proposed model is significantly higher than those of 14 existing models, and the lowest increase was 4.34%. And in the generalization performance, it is better than 14 models. Moreover, the complexity of the proposed model is at the intermediate in the 15 VQA methods. Comprehensively analyzing the accuracy, generalization performance and complexity of the proposed model, it is shown that it is a good VQA model based on machine learning.

Key words Video quality evaluation, neural networks, delay, video contents

Citation Yao Jun-Cai, Shen Jing, Huang Chen-Rong. No reference video quality objective assessment based on multilayer BP neural network. *Acta Automatica Sinica*, 2022, **48**(2): 594−607

收稿日期 2019-07-19 录用日期 2019-09-24

Manuscript received July 19, 2019; accepted September 24, 2019

国家自然科学基金 (61301237), 江苏省自然科学基金面上项目 (BK20201468), 南京工程学院高层次引进人才基金 (YKJ201981) 和西安交通大学博士后基金 (2018M633512) 资助

Supported by National Natural Science Foundation of China (61301237), Natural Science Foundation of Jiangsu Province, China (BK20201468), Scientific Research Foundation for Advanced Talents, Nanjing Institute of Technology (YKJ201981), and Postdoctoral Science Foundation of Xi′ an Jiaotong University (2018M633512)

本文责任编委 黄庆明

Recommended by Associate Editor HUANG Qing-Ming

1. 南京工程学院计算机工程学院 南京 211167 2. 西安交通大学电子与信息工程学院 西安 710049

1. School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167 2. School of Electronic and Information Engineering, Xi′an Jiaotong University, Xi′an 710049

视频技术的发展和应用改变了人们传统的生活、工作和学习等方式. 由此, 视频质量成为一个不可回避的重点话题. 实时、有效和便捷的视频质量评价 (Video quality assessment, VQA) 方法, 是保障视频有效通信的前提[1−2].

视频质量主要受到来自视频内容、编解码、传输环境和人类感知 4 个大的方面因素的影响[1−5]. 视频的压缩编码给视频带来模糊、块效应等损伤[5]; 视频传输中的缓冲延时、卡顿、误码等问题造成视频图像模糊、播放停顿等情况, 均会影响网络视频质量, 使得用户体验质量下降[2]; 对于视频内容, 相同的外在环境但不同的视频内容给人的感知效果也有较大的不同, 视频内容同样是影响视频质量的重要因素[3]; 人类是视频质量的最后接受者和评价者, 视频质量评价结果需要符合人类视觉特性[4−6]. 由此, 在 VQA 中需要考虑上述 4 个大的方面的影响.

VQA 一般分为 3 类: 全参考 (Full-reference, FR)、部分参考 (Reduced-reference, RR) 和无参考 (No-reference, NR) 视频质量评价[1]. 截止目前, 现有的大多数 VQA 模型均是 FR 和 RR, 其典型的有 PSNR (Peak signal-to-noise ratio)、VSNR (Visual signal-to-noise ratio)[7]、SSIM (Structural similarity index)[8]、VQM (Video quality model)[9]、ST-MAD (Spatiotemporal most apparent distortion algorithm)[10]、MOVIE (Motion-based video integrity evaluation)[11] 模型等. 对于 NR-VQA, 其不需要任何来源, 该方法进一步分为两类[12]: 1) NR-P (NR 视觉感知) 类型, 其用于完全解码的视频质量的评价; 2) NR-B (NR 编码) 类型, 其使用从比特流中提取的信息来评价视频质量. 另外, 神经网络方法在 VQA 模型回归方面具有较大的优势, 能够较大地提高构建模型的精度[13], 且由于 NR-VQA 不需要源视频, 其在视频传输中具有重要的实际应用价值, 因而, 结合神经网络的无参考视频质量评价方法成为视频通信的热门研究课题. 近些年报道相关领域的研究成果主要有 VQAUCA (NR VQA using codec analysis)[14]、V-CORNIA (Video codebook representation for NR image assessment)[15]、C-VQA (NR VQA method in the compressed domain)[16]、NR-DCT (Discrete cosine transform-based NR VQA model)[17]、V-BLIINDS (Blind VQA algorithm)[18]、NVSM (NR VQM using natural video statistical model)[19]、3D-DCT (NR-VQA metric based on 3D discrete cosine transform domain)[20] 和 COME (NR VQA method based on convolutional NN and multiregression)[21] 等 NR-VQA 模型, 但其目前仍存在较多问题, 主要有:

1) 失真特征提取数量问题: 在视频通信中, 可能会产生多种类型的视频失真, 在构建 NR-VQA 模型中, 虽然提取更多的视频失真特征可以提高其评估精度, 但同时也增加了其复杂度[12, 19, 22]. 因此, 构建 NR-VQA 模型时应尽量提取少量但有效的失真特征, 但这个度非常难把握;

2) 视频内容及其视觉感知问题: 现有的 NR-VQA 模型通常只关注于传输造成的视频失真, 很少考虑视频内容及其视觉感知效果对视频质量的影响[3, 14]. 因此, 其主客观评价结果一致性较差, 需要结合二者提高精度;

3) HVS 特性问题: 在 VQA 中引入合适有效的 HVS (Human visual system) 感知特性能够显著性提高 VQA 评价精度. 但是, 如果使用从比特流中提取的失真特征来构建 NR-VQA 模型时, 则很难有效地在模型中引入 HVS 特性[3−4]. 因此, 目前一般将 VQA-B 度量和 VQA-P 度量相结合, 构建综合的 NR-VQA 模型, 从而提高了模型的精度;

4) 模型的复杂性问题: 在视频通信中, VQA 需要实时进行, 其要求模型尽可能简单但有效. 然而, VQA 模型往往引入了部分 HVS 特性, 并且依赖于更多的视频失真特性, 同时, 采用了机器学习方法, 因此, 现有的 NR-VQA 模型往往非常复杂[17−22]. 因此, 在构建模型时, 需要对这些特征和方法进行适当的选择, 并对相应的参数进行优化;

5) 泛化性问题: 在 NR-VQA 中, 其方法往往使用机器学习工具获得视频质量评价分数, 然而, 机器学习需要训练样本; 目前, 其常见方法是使用视频数据库中的部分样本进行训练, 而其余部分进行测试, 其实验结果表明, 如此方式, VQA 模型精度较高; 然而, 当测试其他数据库中的视频时, 其模型精度则显著下降[15−22]. 实验表明, 基于机器学习方法的 VQA 模型的泛化性能往往较差. 因此, 有必要对 VQA 模型进行优化, 提高泛化性能.

6) 模型精度问题: 对于基于机器学习方法的 NR-VQA, 往往选取的样本素材、测试和训练样本的比例、不同测试数据库样本等对评价模型的精度有较大的影响[16−19]. 因此, 在模型构建时需要从样本的多个方面来考虑, 以提高精度.

基于此, 在本研究中, 针对上述影响视频质量的 4 个大的方面, 结合多层 BP 神经网络研究了无参考视频质量评价方法, 并与现有模型进行对比分析, 研究了其精度、复杂性和泛化性能.

1 视频特征描述

NR-VQA 在视频传输中具有重要的实际应用

价值. 在视频通信中, 通过提取视频本身特征和传输引起的质量因素, 结合 HVS 特性, 能够很好地预测终端视频的质量; 并且在视频质量预测过程中, 采用机器学习的方法能够获得更高的评价精度, 如采用神经网络的方法. 结合机器学习方法, 需要提取影响视频质量的有效特征. 通过 VQA 的主观实验研究和前人的研究结果表明[1−5], 目前, 影响视频质量主要是上述的 4 个大的方面, 经过反复试验和研究, 其可更进一步细化到 11 个具体的视频特征, 即为图像局部对比度、灰度梯度分布、模糊度、亮度和色度的视觉感知、运动矢量、场景切换、比特率 (BitRate, BR)、初始时延、单次中断时延时长、多次中断平均时延时长、中断频率, 其说明如下.

在视频的时域和空域描述方法中, 视频可以看成是一帧一帧的图像信息与帧间的时域信息相结合而成[1, 9], 即在本研究中, 采用帧图像描述空域信息, 采用运动矢量和场景切换两个方面来描述时域信息.

人眼对视频空域信息的感知, 即是感知其每帧图像. 依据 HVS 特性, 人眼对图像的感知主要是获取图像的灰度和色度信息, 并分辨其清晰程度, 同时其感知效果受局部对比度和人眼敏感程度的影响. 由此, 采用局部对比度、灰度梯度分布、模糊度、及亮度和色度的视觉感知共 4 个特征来表征每帧图像的内容, 即人眼感知到的图像是该 4 个特征的综合结果.

在传输过程中, 影响视频质量的主要因素有: 压缩编码、信道传输条件及其视觉感知. 对于压缩编码, 主要表现为视频压缩编码后的码率 BR; 信道条件对传输视频质量影响最大的是时延 (在 TCP 协议下不考虑丢包), 其包括初始时延、中间单次中断时延、中断频率、中断平均时长.

对于上述的 11 个特征, 采用表 1 中的参数进行定量描述.

2 视频特征参数计算

表 1 中的各参数值的计算或获得方法如下.

1) 图像局部对比度

采用图像中每一点与周围最近的 8 个点的对比度的平均值作为该点受局部环境影响的对比度, 然后乘以对应中心的归一化亮度值, 再对所有点求平均, 其值可以表明人眼在该对比度下的亮度的敏感结果. 记帧图像为 I, 其局部对比度计算表达式为:

$$\text{Value}_{contrast}=\frac{1}{m\times n}\sum_{i=1,j=1}^{m,n}\left\{\frac{1}{8}\left\{\sum_{\substack{L_1=-1,0,1\\L_2=-1,0,1}}\frac{I(i,j)-I(i-L_1,j-L_2)}{\frac{1}{2}[I(i,j)+I(i-L_1,j-L_2)]}\right\}\cdot\frac{I(i,j)}{256}\right\} \tag{1}$$

其中, m、n 表示图像像素的行和列的数目. 找出所有帧图像中该参数值的最大值, 同时对所有帧图像

表 1 所提视频特征及其参数描述

Table 1 Video features and description of their parameters

信息描述	特征	特征名称	参数值描述
空域信息及其感知	特征 1	图像局部对比度	对比度平均值
			对比度最大值
	特征 2	亮度色度视觉感知	亮度色度感知平均值
			亮度色度感知最大值
	特征 3	图像模糊度	模糊度平均值
			模糊度最大值
	特征 4	图像灰度梯度分布及其视觉感知 (内容复杂性视觉感知)	结合 HVS 的灰度梯度期望平均值
			结合 HVS 的灰度梯度期望值的最大值
			每次中断时前 3 帧的结合 HVS 的灰度梯度期望平均值
时域信息及其感知	特征 5	运动信息及其感知	结合 MCSFst 的运动矢量平均值
			结合 MCSFst 的运动矢量最大值
	特征 6	场景切换	复杂性变化对比感知平均值
			复杂性变化对比感知最大值
编解码	特征 7	码率	比特率
传输时延	特征 8	初始时延	初始中断 (缓冲) 时延时长
	特征 9	中间中断时延	中间单次中断 (缓冲) 时延时长
	特征 10	平均中断时长	多次中断平均中断时长
	特征 11	中断频率	单位时间中断次数

求平均, 以其平均值和最大值作为定量描述该视频内容空域信息的特征之一. 表 1 中的其他特征参数的最大值和平均值的计算方法均与该方法一致.

2) 亮度和色度视觉感知

图像亮度和色度的感知主要表现为对其强度大小的感知, 且受空间位置的影响, 不同空间位置, 人眼的敏感程度不同. 则依据 Nadenau[23] 和 Barten[24] 分别提出的人眼觉察静止目标时的亮度和色度对比敏感度函数 CSF (Contrast sensitivity function), 计算图像上每一点像素所在位置人眼的敏感值, 并归一化; 再将敏感值乘以相应像素点的灰度值, 且归一化; 然后按像素点个数求平均, 其平均值为人眼对图像亮度和色度的感知结果; 最后对所有帧图像求平均, 得出的结果和其最大值作为描述该视频内容空域信息的第 2 个特征, 其计算式为:

$$\text{Value}_{I\_\text{CSF}}=\frac{1}{m\times n}\cdot \frac{\sum\limits_{i=1,j=1}^{m,n}\text{IDCT}\left\{\text{CSF}(f_{x_i},f_{y_j})\cdot\text{DCT}[I(i,j)]\right\}}{\max}\cdot\frac{I(i,j)}{256} \tag{2}$$

3) 图像的模糊度

采用文献 [25] 中提出的改进的点锐度算法函数 EAV (Entity attribute value) 来计算, 其表达式为:

$$\text{EAV}=\frac{\sum\limits_{i=1}^{m\times n}\sum\limits_{a=1}^{8}\left|\frac{\text{d}f}{\text{d}x}\right|}{m\times n} \tag{3}$$

式 (3) 中, m、n 为图像像素的行和列, $\text{d}x$ 表示距离增量 (即像素间隔), $\text{d}f$ 表示灰度变化幅值. 同上理, 得到其平均值和最大值, 其结果作为描述该视频内容空域信息的第 3 个特征.

4) 图像的灰度梯度分布及其视觉感知

无论是亮度图还是色度图, 其对人眼的刺激仍然是其强度的感知效果. 其强度值在存储和显示中一般采用 256 级的灰度值来描述. 同时灰度的变化快慢对人眼的刺激也影响人眼对图像亮度和色度的感知. 基于此, 需要计算图像的灰度和梯度分布. 为了更好地说明其分布情况, 采用灰度梯度共生矩阵来描述图像的灰度梯度分布概率情况. 再利用期望值的普遍计算方法, 将各梯度值乘以灰度梯度共生矩阵中对应的值, 其结果作为图像的灰度梯度期望值. 该值亦作为描述图像内容的一个特征, 其能够较好地反映图像内容的复杂性, 记为 Value_c ($\text{Value}_\text{complexity}$), 其计算如表达式 (4).

$$\text{Value}_\text{c}=\sum_{i=0,j=0}^{i=256,j=32}\|\text{gradients}_j\|\cdot\frac{H(\text{gray}_i,\text{gradients}_j)}{m\times n} \tag{4}$$

式中, 图像的梯度被量化为 32 级.

人眼对不同内容复杂性的图像敏感程度明显不同. 结合人眼对图像内容复杂程度的敏感效果, 通过反复主观实验和定性分析, 人眼对图像内容复杂程度的感知结果可用灰度梯度期望值与人眼敏感程度值的乘积来定量描述, 记为结合 HVS 的灰度梯度期望值, 其计算如式 (5).

$$\text{Value}_{\text{c\_sensitivity}}=\begin{cases}K_1\cdot\text{Value}_\text{c}, & \text{Value}_\text{c}<0.15\\ 0.15K_2, & 0.15\le\text{Value}_\text{c}\le0.25\\ K_3\cdot(-0.75\cdot\lg\text{Value}_\text{c}-0.3), & \text{Value}_\text{c}>0.25\end{cases} \tag{5}$$

其中 K_1、K_2 和 K_3 为参数, 可通过主观实验测量数据拟合得出. 式 (5) 计算值越大, 表明人眼对该图像的复杂性越敏感, 感知图像效果越好. 同上方法, 得到所有帧图像中该值的最大值和平均值. 对于视频中的中断, 中断时间内, 人眼感知到的均是中断瞬间前的一帧图像, 为了表明其效果, 计算每次中断时前 3 帧的结合 HVS 的灰度梯度期望平均值, 则该值、所有帧图像的最大值和平均值共同作为描述该视频内容空域信息的第 4 个特征.

5) 运动矢量及其视觉感知

采用全搜索算法匹配获得每一个子块的运动矢量. 结合人眼对运动目标的敏感特性, 将运动矢量乘以敏感阈值, 其结果定量描述为人眼感知视频时域信息的两个特征之一. 其计算方法为:

a) 采用 Kelly[26] 提出的运动目标感知敏感函数 MCSF_st:

$$\text{MCSF}_\text{st}(f_\theta,f_t)=4\pi^2f_\theta\cdot f_t\text{e}^{-4\pi(2f_\theta+f_t)/45.9}\times\left(6.1+7.3\left|\lg\frac{f_t}{3f_\theta}\right|^3\right) \tag{6}$$

其中: MCSF_st 定量表征 HVS 感知运动目标 (即光栅) 时的敏感阈值, f_θ、f_t 为角频率和时间频率[26]. 该式可以计算得出模拟观看视频时, 人眼感知视频任一像素点对应的运动矢量的敏感阈值.

b) 将式 (6) 计算的敏感阈值乘以该点的运动矢量, 并按像素点数求平均, 该值即定量描述为感知视频两帧间时域信息的一个特征值; 将所有两帧间的计算结果对总帧数 N 求平均, 其计算如式 (7). 其平均值和最大值为人眼感知视频时域信息的特征

之一.

$$\text{Value}_{\text{temporal}} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{m\times n}\sum_{i=1,j=1}^{m,n}\frac{\text{MCSF}_{\text{st}}(i,j)\times Mv(i,j)}{\max}\right\}_k \tag{7}$$

6) 视频内场景切换

视频中往往存在较多的场景切换, 而且是快速的, 如此给用户带来非常负面的评价, 但是客观评价分数往往变化不大, 致使主客观评价分数明显不一致. 为了克服该影响, 引入场景切换来描述时域信息, 即采用亮度、色度和纹理 (期望值) 的对比差异来反映场景的切换, 其计算方法为: 首先分别计算当前帧与之前 30 帧各自的灰度梯度期望值; 再基于物理中对比度的定义, 分别计算当前帧与之前 30 帧的灰度梯度期望值的对比度, 其定义表述如式 (8), 并计算其平均值, 则该值用来反映当前帧场景与之前 30 帧场景的切换快慢; 最后按视频帧数求其平均值.

$$\text{Switch}_N = \begin{cases}\dfrac{1}{30}\displaystyle\sum_{L=1}^{30}\frac{\text{Value}_{\text{c},N}-\text{Value}_{\text{c},N-L}}{\frac{1}{2}[\text{Value}_{\text{c},N}+\text{Value}_{\text{c},N-L}]}, & N\geq 30\\ \dfrac{1}{N-1}\displaystyle\sum_{L=1}^{N-1}\frac{\text{Value}_{\text{c},N}-\text{Value}_{\text{c},N-L}}{\frac{1}{2}[\text{Value}_{\text{c},N}+\text{Value}_{\text{c},N-L}]}, & N<30\end{cases} \tag{8}$$

7) 码率 (BR)

码率能够较好地描述视频传输中的连接速度、传输速度、信道容量、最大吞吐量和数字带宽容量. 不同的 BR, 最直接的表现为在终端播放出不同质量的视频, 对于目前广泛采用的 H.264 技术, 其质量主要表现为模糊和块效应. 在视频传输中, 可以直接提取传输视频的 BR. 在仿真实验中, 为了更好地阐述 BR 对 VQA 分数的影响, 将参考视频经 H.264 进行压缩, 其压缩后的视频的 BR 分别取 50 至 10 000 的范围, 从而来探讨 BR 对视频质量的影响.

8) 时延

在视频传输中, 当传输 1 个 segment 的往返时间 RTT (Round trip time) 大于 2 s 时, 播放该 segment 必定发生卡顿, 其卡顿带来的时延大小为 (RTT−2) s; 而在传输视频码流时, RTT 可以通过发送和接受 ACK 的时间戳来计算得到, 即两个时间点的时间差即可求出 RTT; 则在视频码流传输时, 可以预测每一个 segment 在即将播放时是否出现时延、出现时延的位置、时延长短等相关信息. 基于此, 结合视频质量评价的需要, 视频传输可能影响其质量的时延特征量为: 初始时延、中间单次中断时延、中断频率 (次数) 和平均中断时长, 其值可从视频传输时直接获得.

3 视频质量客观评价方法

3.1 建立的视频数据集

为了获得性能优异的 VQA 模型和结果, 需要采用大量的样本对 BP 神经网络进行训练. 为此, 采用开源视频库 LIVE[27] 中的 10 个源视频和 VIPSL[28] 中的 8 个源视频作为参考视频, 分别对其进行不同 BR 的压缩编码和不同方式的时延处理, 得到两个数据集, 分别记其为 LIVEour 和 VIPSLour, 共 1 658 个失真视频. 并采用 21 位观察者, 通过主观实验, 得到其失真视频的主观质量评价分数 (Mean opinion score, MOS). 同时, 计算出每个失真视频的空域信息及其感知特征和时域信息及其感知特征的参数值 (即表 1 中的特征 1~ 特征 6 的参数值), 且提取码率和时延参数值 (即表 1 中的特征 7 ~ 特征 11 的参数值), 共 18 个参数值; 以所有视频的每个参数值作为 1 列, 组成一个 18 维的视频特征参数矩阵; 由于各个特征参数值范围不同、物理意义不同, 再对其进行归一化处理, 归一化处理的方式即为每列的原始值除以该列的最大值. 则以此方法得到失真视频特征参数矩阵及其主观质量评价分数 MOS, 如此, 则每一 “视频特征参数矩阵加上 MOS 分数”即为一个样本, 从而构建训练样本. 实验中, 以此两个数据库中的视频特征参数矩阵作为神经网络的输入, 主观质量评价分数 MOS 值作为其输出, 对所提 VQA 模型进行训练和测试实验. 其中, 考虑到主观实验中个体的差异带给 MOS 分数的偏差, 采用所有观测者的质量评价分数的平均 MOS 值作为神经网络的输出; 另外, 通过神经网络回归模型得出预测视频质量的分数, 其对应为 0 ~ 1 之间的结果, 为了与不同模型的对比, 可将其映射至 0 ~ 100 之间, 以其作为视频的客观质量分数. 由于失真视频来源于不同的数据库, 其特征差异较大, 完全能够满足 11 个特征量较大范围的取值.

3.2 基于多层 BP 神经网络的 VQA 方法基本框架

多层 BP 神经网络对数据的回归在精度上具有较大的优势, 其是一种多层的前馈神经网络. 所提的基于多层 BP 神经网络的 VQA 方法 (BP-VQA) 主要分为 3 步: 第一步特征提取, 第二步对设计的多层 BP 神经网络进行训练, 第三步采用训练好的 VQA 模型对失真视频进行质量评价测试. 其基本

思路为: 依据提出的方案和 BP 神经网络的特征, 采用上述描述的 11 个特征的 18 个参数值作为输入, 主观 MOS 作为输出, 采用构建的 LIVEour 和 VIPSLour 两个数据集中的样本数据来进行训练; 通过大量训练, 得到所需的多层 BP 神经网络; 再用训练好的神经网络评价其余部分的失真视频, 预测出客观质量评价分数, 并与其主观质量评价分数做相关性分析, 计算其参数 PLCC (Pearson linear correlation coefficient)、SROCC (Spearman rank order correlation coefficient)、OR (Outlier ratio) 和 RMSE (Root mean squared error), 从而检验基于多层 BP 神经网络的 VQA 模型的评价性能. 其构建的基本框架如图 1.

3.3　多层 BP 神经网络设置

对于多层 BP 神经网络设置问题, 需要综合模型精度、复杂性和泛化性能 3 个方面来考虑. 隐含层数越多、节点数越多、精度越高, 但耗时越长, 泛化性能越差, 则需要合理选取隐含层数和节点数. 文中采用理论分析与实验效果验证相结合的方法, 不断优化选择, 通过研究, 设置的多层 BP 神经网络如图 2.

对 BP 神经网络的隐含层数目和每一层的节点数, 设计方法如下: 理论上, 对于每一隐含层的节点数, 依据式 (9)[29] 初步估计其值范围 $[a, b]$, 再对范围中每个值进行实验, 结合实验实际效果得出其最优值, 同时兼顾模型复杂性, 并且通过增加隐含层节点数来抵消因减小隐含层数目而带来的误差, 从而尽量减少隐含层的数目.

$$a = \frac{n+l}{2} \le m \le (n+l)+10 = b \tag{9}$$

式中, m 为当前隐含层节点数, n 为前一层或输入节点数目, l 为后一层或输出节点数目. 通过式 (9) 计算和结合其实际效果, 得出各层节点数目依次为 26、24、22、25、19 和 18.

仿真中, BP 神经网络的训练参数设置为: 学习率 (Learning rate) 为 0.01, 误差目标期望值为 0.001, 迭代次数 epochs 为 500, 显示间隔次数为 25.

4　实验及结果

按照以上设计的 VQA 方法, 对构建的视频数据集进行仿真实验.实验中, 训练和测试样本共 1658 组数据. 实验中, 为了使选取的数据具有更好的随机性, 在训练和测试之前, 将所有组的数据按组进行乱序, 即选取训练和测试样本时, 其是置乱后任意选取的一定比例的数据作为训练和测试样本.

在 BP-VQA 模型设计中, 训练样本集及其来源、测试样本集及其来源、训练样本和测试样本比例等均对模型的精度有较大的影响. 则实验采用以下 3 种方式进行: 1) 采用同一数据库中的样本进行

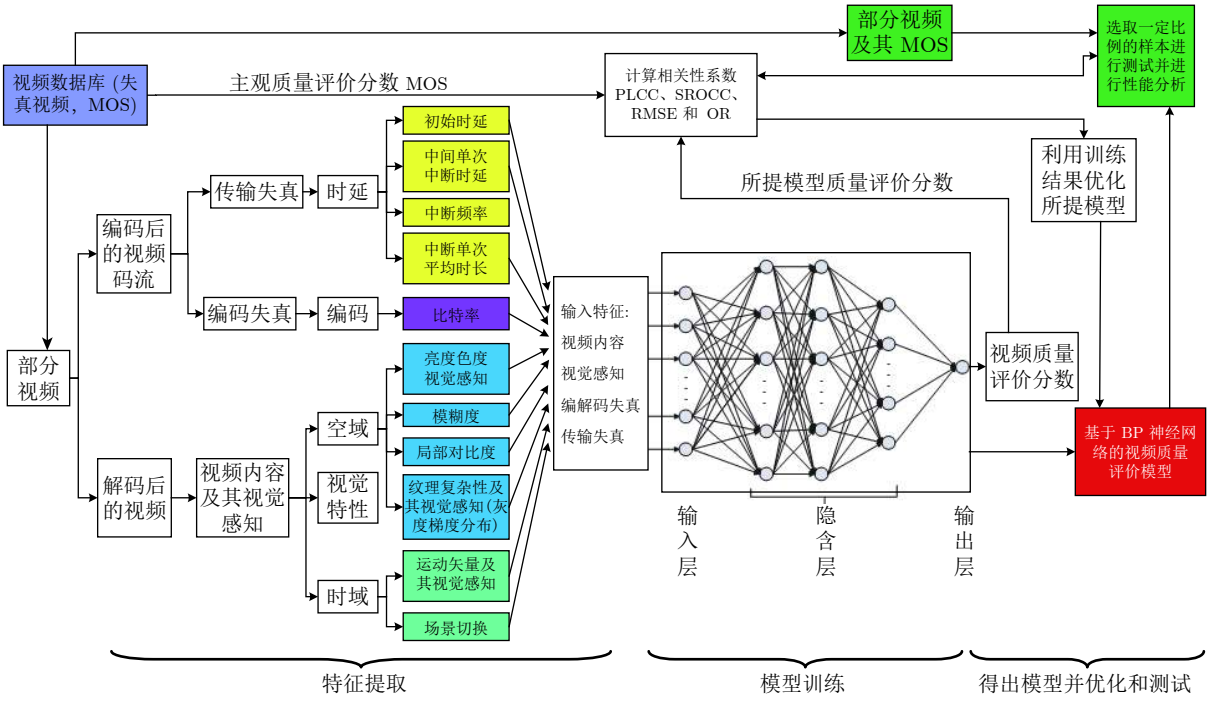


图 1　基于多层 BP 神经网络的无参考视频质量客观评价方法流程图

Fig. 1　Flow chart of no reference video quality objective evaluation method based on multilayer BP neural network

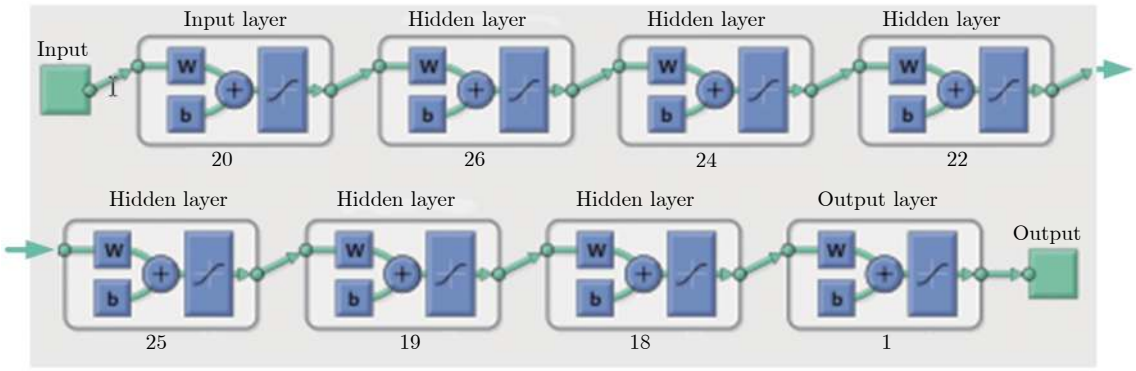


图 2 设置的多层 BP 神经网络结构图

Fig. 2 Designed multilayer BP neural network structure

训练和测试, 2) 采用两个数据库中的样本交叉训练和测试, 3) 采用不同比例的样本训练和测试. 依据此 3 种方式仿真, 其实验结果分别如下.

1) 同一数据库中的样本进行训练和测试

分别采用 LIVEour 和 VIPSLour 数据库中各自 80 % 的数据作为训练样本, 剩余 20 % 的数据作为测试样本进行测试, 得到其质量客观分数, 然后结合数据库中的主观分数, 计算其主客观评价分数之间的相关性参数值 PLCC、SROCC、OR 和 RMSE, 同时得到主客观评价分数之间的散点图, 以及预测值与原始值之间的比较图. 其结果如表 2、图 3 和图 4.

上述实验结果表明, 采用同一数据库中的样本进行训练并测试时, 所提 VQA 模型精度非常高.

2) 两个数据库中的样本交叉训练和测试

实验采用两种方式进行: a) 采用 LIVEour 中 100 % 的数据作为训练, 任意选取 VIPSLour 中 20 % 数据作为测试; b) 采用 VIPSLour 中 100 % 的数据作为训练, 任意选取 LIVEour 中 20 % 数据作为测试. 从而得到客观 VQA 分数. 结合主观 MOS值, 得出其主客观分数一致性分析结果, 即 4 个相关性参数值、散点图和预测值与原始值之间的比较图, 结果如图 5、图 6 和表 3.

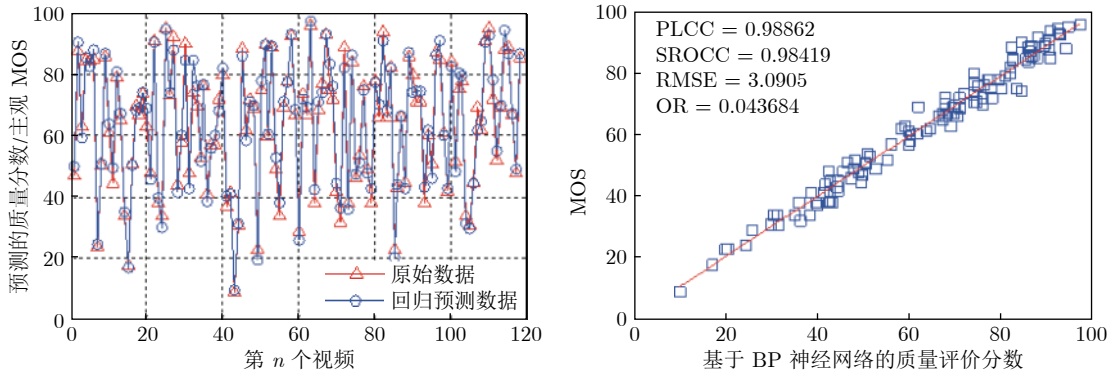
表 2 计算的 4 个相关性参数值

Table 2 Calculated results of four correlation parameters

样本数据库	PLCC	SROCC	RMSE	OR
LIVEour (80 % 训练、20 % 测试)	0.9886	0.9842	3.0905	0.0437
VIPSLour (80 % 训练、20 % 测试)	0.9842	0.97899	3.4389	0.04463

图 5、图 6 和表 3 的结果表明, 该方式得出的散点图和比较图的相关性效果明显差于采用相同数据库作为训练和测试时的效果, 相应的 4 个相关性参数值与之相比, 也明显偏小; 但相对于目前现有的常用 VQA 模型和基于传统方法构建的 VQA 模型来说, 所提模型的 PLCC 和 SROCC 值均明显高于 0.8, 其精度明显高于传统方法.

3) 采用不同比例的样本训练和测试



(a) 原始数据和回归预测数据对比

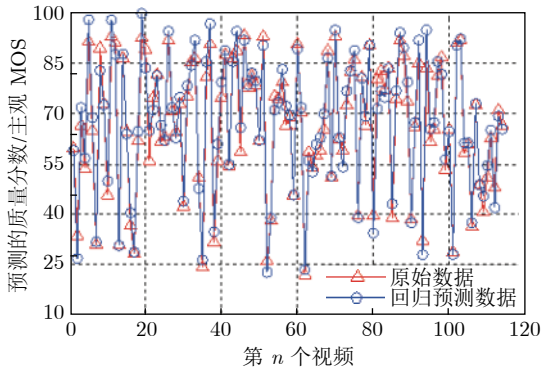
(a) Comparing original data with regression prediction data

(b) LIVEour 80 % 训练和剩余 20 % 测试

(b) LIVEour 80 % training and remaining 20 % testing

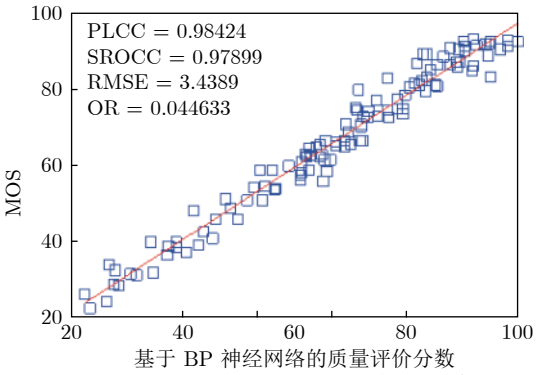
图 3 LIVEour 数据库中 (80% 训练, 20% 测试) 实验结果

Fig. 3 Experimental results in LIVEour database (80 % training, 20 % testing)



(a) 原始数据和回归预测数据对比

(a) Comparing original data with regression prediction data

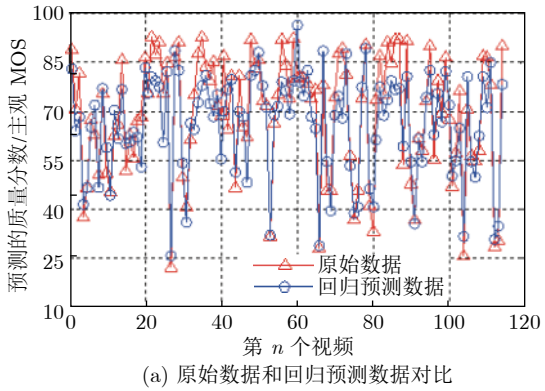


(b) VIPSLour 80% 训练和剩余 20% 测试

(b) VIPSLour 80% training and remaining 20% testing

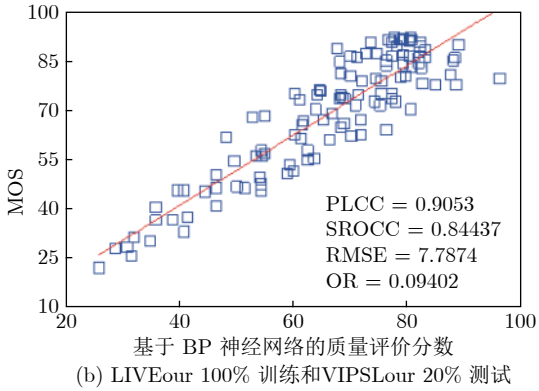
图 4　VIPSLour 数据库中 (80% 训练, 20% 测试) 实验结果

Fig. 4　Experimental results in VIPSLour database (80% training, 20% testing)



(a) 原始数据和回归预测数据对比

(a) Comparing original data with regression prediction data

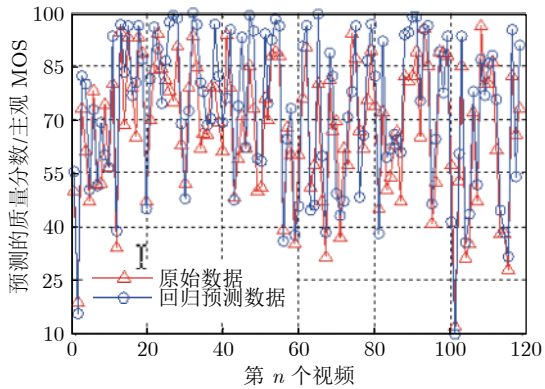


(b) LIVEour 100% 训练和VIPSLour 20% 测试

(b) LIVEour 100% training and VIPSLour 20% testing

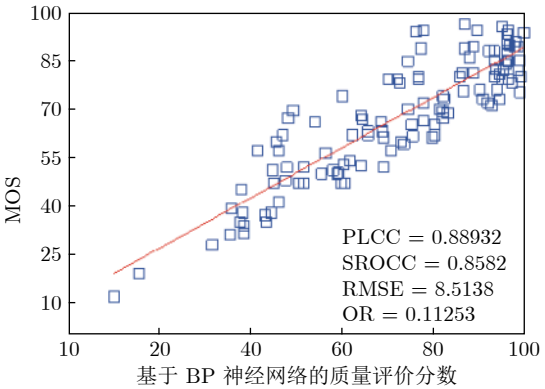
图 5　LIVEour 100 % 训练和 VIPSLour 20 % 测试的实验结果

Fig. 5　Experimental results from training by 100 % samples in LIVEour and testing 20 % samples in VIPSLour



(a) 原始数据和回归预测数据对比

(a) Comparing original data with regression prediction data



(b) VIPSLour 100 % 训练和 LIVEour 20 % 测试

(b) VIPSLour 100 % training and LIVEour 20 % testing

图 6　VIPSLour 中训练 (100 % 样本) LIVEour 中测试 (20 % 样本) 的实验结果

Fig. 6　Experimental results from training by 100 % samples in VIPSLour database and testing 20 % samples in LIVEour database

a) 同一数据库中不同比例的样本训练和测试　分别在 LIVEour 和 VIPSLour 各自数据库中, 依次采用 90 %、70 %、50 %、30 % 的数据作为训练, 其对应剩余 10 %、30 %、50 %、70 % 的数据作

表 3 计算的 4 个相关性参数值
Table 3 Calculated results of four correlation parameters

样本说明 (100 %训练, 20 % 测试)	PLCC	SROCC	RMSE	OR
LIVEour 训练和 VIPSLour 测试	0.9053	0.8443	7.7874	0.0940
VIPSLour 训练和 LIVEour 测试	0.8893	0.8582	8.5138	0.1125

为测试来进行实验. 并做主客观分数一致性分析和计算其相关性参数, 其结果如表 4.

表 4 计算的 4 个相关性参数值
Table 4 Calculated results of four correlation parameters

样本说明	PLCC	SROCC	RMSE	OR
LIVEour 中 90 % 训练和 10 % 测试	0.9897	0.9819	2.8792	0.03375
LIVEour 中 70 % 训练和 30 % 测试	0.9775	0.9753	4.6518	0.07064
LIVEour 中 50 % 训练和 50 % 测试	0.9663	0.9587	5.5681	0.07566
LIVEour 中 30 % 训练和 70 % 测试	0.9504	0.9456	6.3464	0.08892
VIPSLour 中 90 % 训练和 10 % 测试	0.9847	0.9715	3.4695	0.04362
VIPSLour 中 70 % 训练和 30 % 测试	0.9751	0.9694	4.3601	0.05401
VIPSLour 中 50 % 训练和 50 % 测试	0.9668	0.9648	5.1316	0.06715
VIPSLour 中 30 % 训练和 70 % 测试	0.9471	0.9434	6.3954	0.07859

b) 不同数据库中不同比例的样本训练和测试

在 LIVEour 中分别选取 80 % 和 50 % 数据作为训练和对应在 VIPSLour 中选取 20 % 和 50 % 作为测试, 以及在 VIPSLour 中分别选取 80 % 和 50 % 数据作为训练和对应在 LIVEour 中选取 20 % 和 50 % 作为测试来进行实验. 并做主客观分数一致性分析和计算其相关性参数, 其结果如表 5.

表 5 计算的 4 个相关性参数值
Table 5 Calculated results of four correlation parameters

样本 (训练、测试) 比例说明	PLCC	SROCC	RMSE	OR
LIVEour 80 % 和 VIPSL 20 %	0.8876	0.8588	8.4442	0.1116
LIVEour 50 % 和 VIPSL 50 %	0.8735	0.8066	8.4322	0.0970
VIPSLour 80 % 和 LIVE 20 %	0.8780	0.8486	9.3577	0.1267
VIPSLour 50 % 和 LIVE 50 %	0.8507	0.8403	10.7100	0.1449

表 4 和表 5 中相关性数值表明: 1) 对于所提的 BP-VQA 模型, 当在相同数据库中选择样本训练和测试时, 即使在 30 % 的数据训练下, 其精度仍能达到 PLCC 为 0.9471, 而且当训练样本为 90 % 时, 其评价精度能够实现达到 PLCC 为 0.9897, 其结果表明所提模型精度非常高, 而且是在两个不同类型的视频数据库中得出的结果; 2) 对于两个数据库中交叉选择训练和测试样本时, 其模型仍然表现为精度较高, 在训练样本为 50 % 时其能够达到 0.8507, 表明该模型具有较高的泛化性能.

分析上述 3 种情况的实验结果, 可以得出: 1) 在同一数据库中, 不论训练和测试样本比例如何, 所提模型的精度均较高, 即使在 30 % 的训练样本下, 其 PLCC 值也能达到 0.9471; 2) 对于同一数据库中不同比例的训练和测试样本的实验结果, 训练样本占比越大, 测试时, PLCC 值越大, 所提模型的精度越高, 说明训练和测试样本比例对模型精度影响较大, 分析其原因, 主要在于, 当训练样本不足时, 训练样本不足以囊括测试样本中所有有效视频特征, 以致训练得出的模型对部分测试样本不能有效地预测, 部分预测值与主观评价分数存在差异; 3) 对于交叉数据库中不同比例的训练和测试样本的实验结果, 所提模型的精度 PLCC 值也能均在 0.8507 以上, 表明所提 VQA 模型具有较高的精度和泛化性能; 4) 对比现有的基于传统数学建模构建的 VQA 模型, 如 PSNR、VSNR、SSIM、VQM、ST-MAD 和 MOVIE, 无论是其精度、还是其泛化性能, 均得到了较大的提升.

5 讨论

对于视频质量评价模型, 评估其性能的要点主要为: 1) 精度, 2) 泛化性能, 3) 模型复杂性. 基于此, 从此 3 个方面来分析所提 VQA 模型的性能.

5.1 VQA 模型精度对比分析

在视频质量评价中, 评估 VQA 模型精度优劣的方法是分析主客观质量分数之间的一致性, 其一般采用 PLCC 和 SROCC 值来描述 VQA 模型的精度. 为了说明所提模型的精度, 将其与现有的 VQA 模型作以下对比: 1) 与基于机器学习的 VQA 模型的精度对比, 2) 与基于传统数学建模方法的 NR-VQA 模型的精度对比, 3) 与全参考 VQA 模型的精度对比, 其对比结果如下.

1) 与基于机器学习的 VQA 模型的精度对比

将所提模型与近些年提出的 6 种基于机器学习的 VQA 模型进行精度对比, 其 6 种模型为: VQAU-CA[14]、V-CORNIA[15]、NR-DCT[17]、V-BLIINDS[18]、3D-DCT[20] 和 COME[21], 其实验仿真结果如图 7.

该对比的 6 种模型的结果均为采用 LIVE 数据库中 150 个视频的实验结果; 该 150 个视频数据包括 H.264、MPEG2、Wireless 和 IP4 个方面的失真, 其对应的失真视频数据为 40、40、40 和 30 组[21]; 其评价结果均为采用数据的 80 % 训练和 20 % 测试的实验结果. 为了对比的公平性, 选取本研究中同一数据库中的不同训练和测试样本比例下的实验结

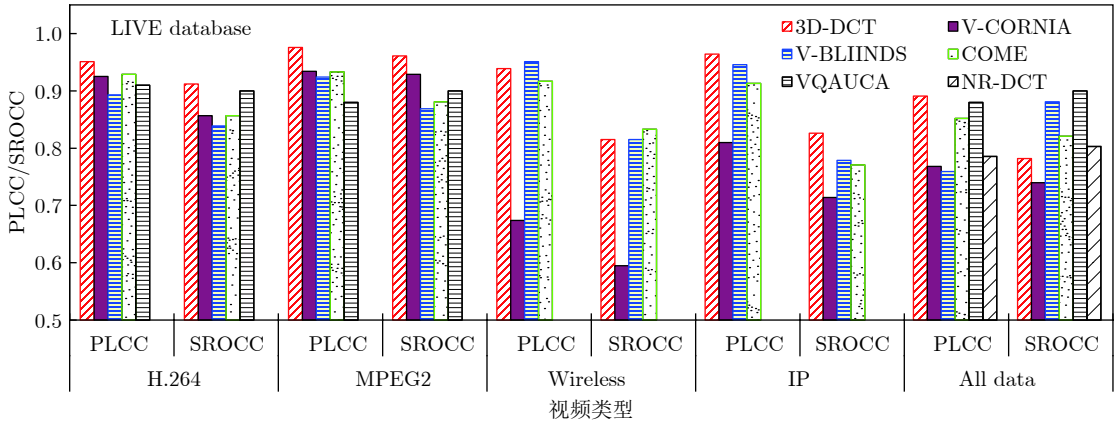


图 7　采用现有 6 种基于机器学习的 VQA 模型评价结果的 PLCC 和 SROCC

Fig. 7　PLCC and SROCC of VQA results with 6 existing models based on machine learning

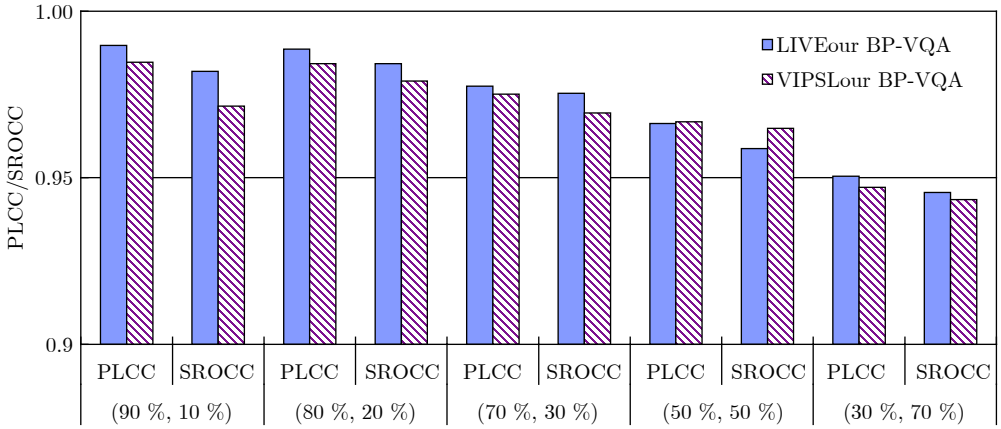


图 8　不同训练和测试样本比例下采用所提 BP-VQA 模型评价结果的 PLCC 和 SROCC

Fig. 8　PLCC and SROCC from applying the proposed BP-VQA model to evaluate video quality under different training and test sample ratios

果与其对比, 其结果如图 8.

从图 7 和图 8 的对比结果可以看出, 在相近的训练和测试样本比例下, 本文所提的 BP-VQA 模型的 PLCC、SROCC 值均高于现有的 6 种 VQA 模型的值, 其幅度为 PLCC 平均高 13.96 %; 则从精度对比上, 所提模型优于此 6 种 VQA 模型. 分析其原因, 主要因为: a) VQA 模型在考虑影响视频质量因素的全面性上存在较大差异, 6 种前人所提模型只考虑了 4 大影响视频质量因素 (即视频内容、编解码失真、传输失真和视觉特性) 中的部分特征, 即, VQAUCA[14] 主要是针对 H.264/AVC 失真视频、通过帧内预测的量化分析而提出的一种基于编解码分析的 NR-VQA 模型, V-CORNIA[15] 是一种基于码本表达的 NR-VQA 模型, NR-DCT[17] 是一种基于 DCT 和 6 个传输失真特征而构建的 VQA模型, V-BLIINDS[18] 是一种基于视频场景特征统计和感知特征的 VQA 模型, 3D-DCT[20] 是一种基于三维 DCT 域的时空自然视频失真统计特征的 NR-VQA 模型, COME[21] 是一种基于视频空域和时域失真特征的 VQA 模型; 而所提模型不仅考虑了视频内容特征、编解码和视觉特性, 还考虑了传输中的时延失真特征; 则相对于此 6 种 VQA 模型, 所提模型考虑影响视频质量的因素更为全面; b) 采用的机器学习方法不同, 6 种现有模型中, VQAUCA、V-CORNIA、V-BLIINDS 和 3D-DCT 均是采用 SVM 进行视频质量预测、NR-DCT 是采用多层神经网络, COME 是采用卷积神经网络提取视频失真特征和采用多元回归预测视频质量分数; 对于 SVM、神经网络和多元回归方法而言, 在相同条件下, 神经网络方法效果优于 SVM 和多元回归方法; 而本文所提模型采用了多层 BP 神经网络, 所以其效果优于此 6 种现有 VQA 模型.

2) 与基于传统数学建模方法的 NR-VQA 模型的精度对比

将所提模型与现有的 3 种基于传统数学建模构建的无参考 VQA 模型的精度进行对比, 3 个模型为 C-VQA[16], NVSM[19] 和 BRVPVC[30]. 为了对比的公平性, 所提 BP-VQA 模型的结果均采用同一数据库中 50 % 训练和 50 % 测试的实验结果, 其对比结果如表 6.

从表 6 中的对比结果可以看出, 所提 BP-VQA 模型的精度明显高于基于传统数学建模方法构建的 VQA 模型的精度.

3) 与全参考 VQA 模型的精度对比

目前, 在 VQA 的研究和实际应用中, 主要仍是采用传统的全参考质量评价方法, 如 PSNR、SSIM[8] 和 MOVIE[11] 等. 为此, 将所提 BP-VQA 模型与 6 种经典的全参考视频质量评价模型 (即 PSNR、VSNR[7]、SSIM[8]、VQM[9]、ST-MAD[10] 和 MOVIE[11]) 的精度进行对比, 其结果如图 9. 其中, 所提模型均是采用 50 % 训练和 50 % 测试的结果.

通过图 9 的精度对比, 可以发现, 所提模型的精度均明显高于基于传统数学建模方法构建的全参考 VQA 模型的精度, 而且其高出幅度最小为 10.67 %.

综合分析表 6 和图 9 的结果, 可以得出, 相对于传统方法构建的无论是 NR-VQA 模型, 还是 FR-VQA 模型, 采用多层 BP 神经网络构建的 VQA 模型, 其精度明显得到了较大的提升. 表明采用机器学习方法构建 VQA 模型在提高精度上具有较大的优势.

5.2 模型复杂性分析

对于 VQA 模型复杂性, 一般采用评价视频质量时算法的运算时间来描述, 其运算时间越短, 模型的复杂性越低. 基于此, 将所提 BP-VQA 模型与现有 10 种 VQA 模型的复杂性进行对比, 其 10 种模型为: PSNR、VSNR[7]、MS-SSIM[8]、VQM[9]、ST-MAD[10]、MOVIE[11]、V-CORNIA[15]、V-BLIINDS[18]、COME[21] 和 BRVPVC[30]. 实验采用 MATLAB-2014a 编程语言, 在 64 位操作系统的 DELL Core i7 笔记本上进行仿真, 其处理器为 Intel(R) Core(TM) i7-8550U CPU @1.80 GHz 1.99 GHz. 为了对比的需要, 采用每种模型平均评价 10 帧图像时算法运行的耗时来比较, 其结果如图 10.

从图 10 中模型复杂性对比的结果上看, 所提 BP-VQA 模型的复杂性处于 11 种方法中的中等水平. 比 PSNR 和 SSIM 模型的复杂性高, 但低于 VQM、MOVIE 和 ST-MAD.

对比分析图 10 的实验结果, 结合每种 VQA 模型的特点和计算公式, 其复杂性主要存在于 3 个过程中: 1) 视频特征提取, 2) 模型训练, 3) 视频质量评价. 分析每一过程, 模型复杂性的原因主要为: 1) 除了 PSNR、SSIM、VSNR 和 COME 外, 其余模型均需要提取大量的视频特征, 并对其进行计算, 该过程耗时基本上占据了整个模型运算耗时的 90 %, 其是模型复杂的主要原因; 2) 对于基于机器学习的 VQA 模型, 一般需要对模型进行训练, 其耗时同样

表 6 所提 BP-VQA 模型与 3 种 NR-VQA 模型的精度对比

Table 6 Accuracy comparison between the proposed BP-VQA model and three existing NR-VQA models

数据库 Metric	LIVE database			LIVEour BP-VQA	VIPSLour BP-VQA
	NVSM	C-VQA	BRVPVC		
PLCC	0.732	0.7927	0.8547	0.9663	0.9668
SROCC	0.703	0.772	0.826	0.9587	0.9648

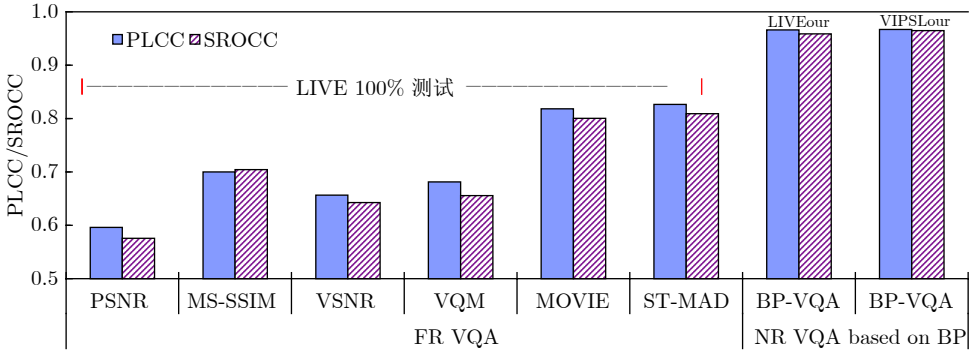


图 9 所提 BP-VQA 模型与 6 种现有 FR-VQA 模型的精度对比

Fig. 9 Accuracy comparison between the proposed BP-VQA model and six existing FR-VQA models

比较长, 而且, 为了使得所提模型的精度更高、模型更具有泛化性能, 一般需要采用大量的样本对其训练, 致使其耗时很长; 3) 在设计基于机器学习的 VQA 模型时, 为了提高精度, 对于每一隐含层, 尽量增加隐含层的数目和每一层的节点数目, 致使模型评价视频时耗时更长, 而且不仅耗时长, 其模型的泛化性能会明显下降; 4) 对于采用训练好的模型对视频质量的评价的耗时, 一般时间非常短, 基本上不影响模型的复杂性; 5) 对于所提模型, 其复杂性主要在视频特征提取上占时较多, 在其特征提取中, 其提取了 11 个特征, 共 18 个参数, 所以耗时稍长; 但综合精度和复杂性, 其实际应用效果更好.

5.3　泛化性能分析

为了说明所提 BP-VQA 模型的泛化性能, 将其与 8 种现有 VQA 模型的泛化性能进行比较, 其 8 种方法为: VQM、ST-MAD、MOVIE、3D-DCT、VQAUCA、V-BLIINDS、NR-SVR (NR-VQA used support vector regression)[31] 和 NR-MLP (NR-VQA used multilayer perceptron for nonlinear mapping)[31], 8 种模型的结果分别来自于 LIVE、VQEG、IRCCyN、CSIQ[32]、EPFL-PoliMI[1]、IVP[33] 和 Lisbon[34] 数据库中失真视频仿真的实验结果. 为了说明模型泛化性能的优劣程度, 采用实验结果的 PLCC 和 SROCC 大小来比较, 其对比结果如图 11.

从图 11 中可以直观地得到: 1) 在不同的数据库中, 除了 VQM 在 EPFL-PoliMI 数据库中的结果稍好外, 所提模型的评价结果的 PLCC 和 SROCC 值明显高于 8 种现有模型的评价结果的 PLCC 和 SROCC 值, 表明其泛化性能好于该 8 种模型; 2) 8 种模型中, 3D-DCT、NR-SVR、NR-MLP、VQAUCA、V-BLIINDS 同样是基于机器学习的 VQA 模型, 从图 11 中的结果可以看出, 无论是在同一数据库中训练和测试, 还是在不同数据库中训练和测试, 所

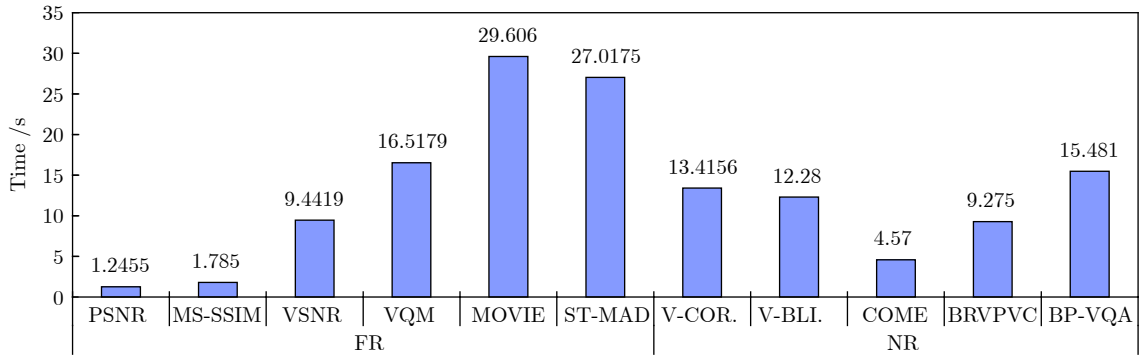


图 10　所提模型与 10 种现有 VQA 模型的运算耗时对比

Fig. 10　Comparisons of the computational time between the proposed model and 10 existing VQA models

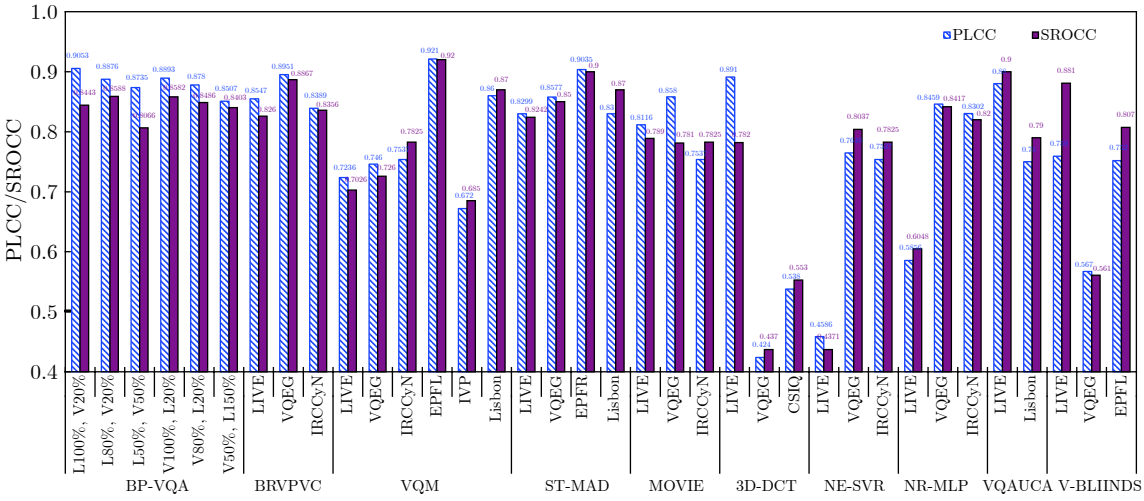


图 11　所提模型与 8 种现有模型的泛化性能对比

Fig. 11　Comparison of generalization performance between the proposed model and 8 existing models

提模型的测试结果即 PLCC、SROCC 值基本上都高于该 5 种模型的测试结果, 表明, 在结合机器学习的 VQA 模型中, 所提模型具有更高的泛化性能; 3) 本文所提模型在交叉数据库中进行了实验, 对于训练和测试样本采用不同数据库的做法, 在其他几种方法中基本没有相关报道, 而且, 在此情况下, 所提模型的评价结果的 PLCC 和 SROCC 值均超过了 0.8. 这也表明本文所提 VQA 模型具有更好的泛化性能; 4) 所提模型在不同数据库和交叉数据库中实验时, 其评价结果的 PLCC 和 SROCC 值均比较平稳, 而其他 8 种模型, 在不同数据库中, 均表现出评价结果的不稳定性; 且本文的模型的精度均超过了 0.8, 这也表明, 本文所提模型具有更好的泛化性能.

分析所提模型泛化性能较优的原因, 主要为: 1) 对于结合多层 BP 神经网络的 VQA 来说, 一般为了其模型的精度要求, 尽量增加隐含层数和每一层的节点数; 的确, 二者增加了, 模型精度能得到提高, 以致往往采用深度学习的方法进行, 甚至让隐含层增加到 20 层以上, 每一层的节点达到 30 个以上; 但是在精度提高的同时, 模型的复杂性也极大地增加了, 而且泛化性能也明显下降; 对于所提模型, 通过实验验证和理论分析相结合, 刚好精度达到极大时, 选取其合适的层数和节点数目, 并且如果再增加层数和节点数, 对模型精度的提高非常有限, 所以所选的隐含层数和节点数是最少的, 则此时, 相对来说, 层数和节点数对泛化性能的影响最小; 同时也说明, 在结合机器学习的 VQA 研究中, 应尽量减少隐含层数和节点数. 2) 对于结合机器学习的 VQA 评价中, 往往为了提高精度, 提取较多的影响因子, 并且每个影响因子采用多个参数来描述, 但是, 不同的视频, 其特征均有较大的差异, 则需要选择更多的特征, 由此反而限定了待测视频, 没有考虑其他视频的特征, 以致在评价其他数据库中的视频时, 其 PLCC 和 SROCC 值均明显下降, 表现出较差的泛化性能; 而对于文中所提模型, 基本上考虑了 VQA 中影响视频质量的 4 个大的方面因素, 并且只采用了 11 个特征进行描述, 相对来说特征相对较少, 所以对泛化性能的影响也较小. 而相对于 NR-SVR、NR-MLP、VQAUCA、V-BLIINDS4 种模型来说, 其提取的特征均较多, 甚至特征参量达到 36 个之多, 其不仅严重影响了 VQA 模型的泛化性能, 而且其复杂性也比较高. 所以, 通过对比分析表明, 所提模型的泛化性能优于该 8 种 VQA 模型.

6 结论

结合多层 BP 神经网络, 针对视频的 4 个主要影响因素, 即编解码影响、传输条件、视频内容和 HVS 特性, 研究了无参考视频质量评价方法, 构建了 VQA 模型. 在研究中, 首先采用图像的亮度和色度及其视觉感知、图像的灰度梯度期望值、图像的模糊程度、局部对比度、运动矢量及其视觉感知、场景切换特征、比特率、初始时延、单次中断时延、中断频率和多次中断平均时长共 11 个特征来描述 4 个主要影响视频质量的因素, 并提取相关的特征参数; 再以其作为输入节点, 采用多层 BP 神经网络, 通过建立的两个视频数据库中的大量视频样本对其进行训练学习, 构建 VQA 模型; 最后, 将所提模型应用于构建的视频数据库中, 对失真视频进行评价仿真实验, 且与 14 种现有的视频质量评价模型进行对比分析, 研究其精度、复杂性和泛化性能. 实验结果表明: 本文所提模型的精度明显高于其他 14 种模型的精度, 其精度最低高出幅度为 4.34 %; 且其泛化性能优于它们, 同时其复杂性处于该 15 种模型中的中间水平. 综合分析所提模型的精度、泛化性能和复杂性表明, 所提模型是一种较好的基于机器学习的视频质量评价模型.

References

1 Vega M T, Perra C, Turck F D, Liotta A. A review of predictive quality of experience management in video streaming services. *IEEE Transactions on Broadcasting*, 2018, **64**(2): 432−445

2 James N, Pablo S G, Jose M A C, Wang Q. 5G-QoE: QoE modelling for ultra-HD video streaming in 5G networks. *IEEE Transactions on Broadcasting*, 2018, **64**(2): 621−634

3 Demóstenes Z R, Renata L R, Eduardo A C, Julia A, Graca B. Video quality assessment in video streaming services considering user preference for video content. *IEEE Transactions on Consumer Electronics*, 2014, **60**(3): 436−444

4 Nan Dong, Bi Du-Yan, Ma Shi-Ping, Fan Zun-Lin, He Lin-Yuan. A quality assessment method with classified-learning for dehazed images. *Acta Automatica Sinica*, 2016, **42**(2): 270−278
(南栋, 毕笃彦, 马时平, 凡遵林, 何林远. 基于分类学习的去雾后图像质量评价算法. 自动化学报, 2016, **42**(2): 270−278)

5 Gao Xin-Bo. *Quality Assessment Methods for Visual Information*. Xi′an: Xi′an Electronic Science & Technology University Press, 2011.72−85
(高新波. 视觉信息质量评价方法. 西安: 西安电子科技大学出版社, 2011.72−85)

6 Feng Xin, Yang Dan, Zhang Ling. Saliency variation based quality assessment for packet-loss-impaired videos. *Acta Automatica Sinica*, 2011, **37**(11): 1322−1331
(冯欣, 杨丹, 张凌. 基于视觉注意力变化的网络丢包视频质量评估. 自动化学报, 2011, **37**(11): 1322−1331)

7 Chandler D M, Hemami S S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 2007, **16**(9): 2284−2298

8 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600−612

9 Pinson M H, Wolf S. New standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 2004, **50**(3): 312−322

10 Vu P V, Vu C T, Chandler D M. A spatiotemporal most-apparent-distortion model for video quality assessment. In: Proceed-

ings of the 2011 IEEE International Conference on Image Processing (ICIP). Brussels, Belgium: IEEE, 2011. 2505–2508

11 Seshadrinathan K, Bovik A C. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Transactions on Image Processing*, 2010, **19**(2): 335–350

12 Uzair M, Dony R D. No-reference transmission distortion modelling for H 264/AVC-coded video. *IEEE Transactions on Signal and Information Processing over Networks*, 2015, **1**(3): 209–221

13 Menor D P A, Mello C A B, Zanchettin C. Objective video quality assessment based on neural networks. *Procedia Computer Science*, 2016, **96**(1): 1551–1559

14 Jacob S, Søren F, Korhonen J. No-reference video quality assessment using codec analysis. *IEEE Transactions on Circuits & Systems for Video Technology*, 2015, **25**(10): 1637–1650

15 Xu J, Ye P, Liu Y, Doermann D. No-reference video quality assessment via feature learning. In: Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE, 2014. 491–495

16 Lin X, Ma H, Luo L, Chen Y. No-reference video quality assessment in the compressed domain. *IEEE Transactions on Consumer Electronics*, 2012, **58**(2): 505–512

17 Zhu K, Li C, Asari V, Saupe D. No-reference video quality assessment based on artifact measurement and statistical analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, **25**(4): 533–546

18 Saad M A, Bovik A C, Charrier C. Blind prediction of natural video quality. *IEEE Transactions on Image Processing*, 2014, **23**(3): 1352–1365

19 Galea C, Farrugia R A. A no-reference video quality metric using a natural video statistical model. In: Proceedings of the 2015 International Conference on Computer as a Tool (EUROCON). Salamanca, Spain: IEEE, 2015. 1–6

20 Li X, Guo Q, Lu X. Spatiotemporal statistics for video quality assessment. *IEEE Transactions on Image Processing*, 2016, **25**(7): 3329–3342

21 Wang C, Su L, Zhang W. COME for no-reference video quality assessment. In: Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). Miami, FL, USA: IEEE, 2018. 232–237

22 Song J, Yang F, Zhou Y, Gao S. Parametric planning model for video quality evaluation of IPTV services combining channel and video characteristics. *IEEE Transactions Multimedia*, 2017, **19**(5): 1015–1029

23 Nadenau M. Integration of human color vision models into high quality image compression [Ph. D. dissertation], École Polytechnique Fédérale de Lausanne, Switzerland, 2000

24 Barten P. Evaluation of subjective image quality with the square-root integral method. *Journal of the Optical Society of America A*, 1990, **7**(10): 2024–2031

25 Wang Hong-Nan, Zhong Wen, Wang Jing, Xia De-Shen. Research of measurement for digital image definition. *Journal of Image and Graphics*, 2018, **9**(7): 828–831
(王鸿南, 钟文, 汪静, 夏德深. 图像清晰度评价方法研究. 中国图象图形学报, 2018, **9**(7): 828–831)

26 Kelly D H. Motion and vision II Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 1979, **69**(10): 1340–1349

27 Sheikh H R., Wang Z, Bovik A C. LIVE image and video quality assessment database [Online], available: http://live.ece.utexas.edu/research/quality, May 20, 2018

28 Gao X B, Li J, Deng C. VIPSL image & video database [Online], available: http://see.xidian.edu.cn/vipsl/index.html, June 5, 2018

29 Tang Xian-Lun, Du Yu-Ming, Liu Yu-Wei, Li Jia-Xin, Ma Yi-Wei. Image recognition with conditional deep convolutional generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 855–864
(唐贤伦, 杜一铭, 刘雨微, 李佳歆, 马艺玮. 基于条件深度卷积生成对抗网络的图像识别方法. 自动化学报, 2018, **44**(5): 855–864)

30 Yao J C, Liu G Z. Bitrate-based no-reference video quality assessment combining the visual perception of video contents. *IEEE Transactions on Broadcasting*, 2019, **65**(3): 546–557

31 Song J R, Yang F Z, Zhou Y C, Gao S. Parametric planning model for video quality evaluation of IPTV services combining channel and video characteristics. *IEEE Transactions on Multimedia*, 2017, **19**(5): 1015–1029

32 Yao J C, Liu G Z. Improved SSIM IQA of contrast distortion based on the contrast sensitivity characteristics of HVS. *IET Image Processing*, 2018, **12**(6): 872–879

33 Blu T, Cham WK, Ngan KN. IVP video quality database [Online] , available: http://ivp.ee.cuhk.edu.hk/, July 12, 2018

34 Brandão T, Roque L, Queluz M P. IST–Tech. University of Lisbon subjective video database [Online], available: http://amalia.img.lx.it.pt/~tgsb/H264_test/, July 15, 2018

姚军财 博士, 南京工程学院计算机工程学院教授.主要研究方向为图像和视频处理, 计算机视觉与模式识别. 本文通信作者.
E-mail: yjc4782@163.com
(**YAO Jun-Cai** Ph. D., professor at the School of Computer Engineering, Nanjing Institute of Technology. His research interest covers image and video processing, computer vision and pattern recognition. Corresponding author of this paper.)

申 静 南京工程学院计算机工程学院副教授. 主要研究方向为图像和视频处理, 多媒体技术和人工智能.
E-mail: shenjingtg@163.com
(**SHEN Jing** Associate professor at the School of Computer Engineering, Nanjing Institute of Technology. Her research interest covers image and video processing, multimedia technology, and artificial intelligence.)

黄陈蓉 博士, 南京工程学院计算机工程学院教授. 主要研究方向为图像分割和编码, 计算机视觉与模式识别.
E-mail: huangcr@njit.edu.cn
(**HUANG Chen-Rong** Ph. D., professor at the School of Computer Engineering, Nanjing Institute of Technology. Her research interest covers image segmentation and code, computer vision and pattern recognition.)