文章编号：1003-0530(2022)06-1111-33

全媒体内容质量评价研究综述

颜成钢 孙垚棋 钟 昊 朱晨薇 朱尊杰 郑博仑 周晓飞
（杭州电子科技大学自动化学院，浙江杭州 310018）

摘 要： 在全媒体时代，媒体内容的表现形式逐渐丰富，开始成为影响信息传播的一个重要因素。内容质量评价仍停留在“流量思维”阶段，难以客观评价内容质量，亟需发展以用户为中心的全媒体内容质量评价方法。本文主要概述近十年来国内外公开发表的不同媒介的评价模型，回顾了图像、视频、音频、文本四类的客观质量评价在全媒体数据中的研究工作及相应的应用，主要介绍基于传统方法和基于深度学习方法两大方向中一些影响力较大的方法，每类方法有分成有参考和无参考的方法，对此总结了各方法特点，对一些具有代表性的方法进行了实验对比分析。最后对四种媒介内容质量评价领域仍面临的问题进行了总结并展望未来可能的发展方向。

关键词： 全媒体；图像质量评价；视频质量评价；音频质量评价；文本质量评价
中图分类号： N19 **文献标识码：** A **DOI：** 10. 16798/j. issn. 1003-0530. 2022. 06. 001

引用格式： 颜成钢，孙垚棋，钟昊，等. 全媒体内容质量评价研究综述[J]. 信号处理，2022，38(6)：1111-1143. DOI：10. 16798/j. issn. 1003-0530. 2022. 06. 001.
Reference format：YAN Chenggang，SUN Yaoqi，ZHONG Hao，et al. Review of omnimedia content quality evaluation [J]. Journal of Signal Processing，2022，38(6)：1111-1143. DOI: 10. 16798/j. issn. 1003-0530. 2022. 06. 001.

Review of Omnimedia Content Quality Evaluation

YAN Chenggang SUN Yaoqi ZHONG Hao ZHU Chenwei ZHU Zunjie ZHENG Bolun
ZHOU Xiaofei
（School of Automation，Hangzhou Dianzi University，Hangzhou，Zhejiang 310018，China）

Abstract: In the era of all media，the forms of media content are gradually enriched and become an important factor affecting the dissemination of information. Building models to assess the quality of omnimedia content has attracted increasing attention. Quality assessment methods are mainly divided into subjective models and objective models. Subjective methods aim to assess quality through human eyes and human senses. It requires a lot of manpower and material resources，and the evaluation process also takes a lot of time；therefore it is difficult to apply in practical application. Objective quality assessment simulates the human observation process，which can automatically predict quality input. This review mainly summarizes the evaluation models of different media published at home and abroad in the past ten years. Research work and corresponding applications in omnimedia data. We mainly list some influential methods in the two major directions based on traditional methods and methods based on deep learning. The quality assessment models of the video and audio parts are divided into traditional methods and deep learning-based models. Each type of model is divided into reference models and non-reference models. Compared with the methods with reference data，the performance of the non-reference method has some differences. However，the no-reference quality evaluation model has strong applicability because it does not need to rely on reference information，and has always been a research hotspot in the field

收稿日期：2021-10-21；修回日期：2022-03-21
基金项目：全媒体信息传播理论与基础服务技术研究，重点研发计划(SY2020YFB1406600)；国家自然科学基金(61931008，61671196，62071415，62001146，61701149，61801157，61971268，61901145，61901150，61972123)；浙江省自然科学基金(LR17F030006，Q19F010030)

of image quality evaluation. The image part is mainly developed by dividing the unreferenced quality assessment model into supervised learning and unsupervised learning. The unsupervised method does not require the support of manual scoring data, saves labor, and has good development prospects. The text quality assessment model is introduced from the two directions of automatic scoring system and text generation quality assessment. Finally, it is concluded that traditional or applied deep learning methods have their own characteristics. These methods are independent of each other and form their own systems. It also looks forward to the possible development direction of all-media content quality assessment in the future.

Key words: omnimedia; image quality evaluation; video quality evaluation; audio quality evaluation; text quality evaluation

1 引言

随着科技的发展和时代的进步,用来实现信息交互的手段越来越多,从一开始用文字进行信息传递,再到后来的图像、音频、视频等多种媒体技术的使用越来越频繁。早期时候,人们利用报纸、杂志等出版物作为文字的传播载体,使用广播来传递音频信息。之后,由于电视的普及,视频图像等可视化信息得以大规模的进行传播。进入到21世纪后,随着互联网和5G时代的到来,用户可以通过智能手机、网络电视以及电脑等各种设备将多种多样的信息进行融合并且传播。在这样的大环境下,有着许多与各种媒体相关的研究工作,例如,面向图像视频取证的相关方法[1-2]以及对于各种媒体下认知安全的研究[3],而“全媒体”也随之诞生,并且受到了越来越多的研究人员的广泛关注。

与“跨媒体”方法不同的是,作为当前信息传递手段的集成者,“全媒体”并不是对各种信息传播媒介进行简单的连接,而是指各种媒体间的全方位融合,从而实现其覆盖面广、技术方法多样、信息传播媒介全面等特点。

在“全媒体”当中,文本、图像、音频、视频等媒介发挥着重要的作用,成为了“全媒体”技术发展中的主力军。因此,对于信息在以这些方式进行传播的过程当中,能否做到很好的保留原始信息,让用户在接收信息的过程当中有着比较好的体验,是当前技术发展需要解决的一个重要问题。

针对这一问题,全媒体质量评价技术起到了关键性的作用,通过对全媒体中的各种媒介进行质量评价,得到客观真实的评价结果,然后根据该结果判断信息在传输过程中发生的损失,以此作为评判标准从而对传输过程进行改进,来提高用户获取信息的完整程度。

在信息传播的过程当中,不同的媒体的质量评价标准各有不同。例如对于文本而言,如果其存在语病或者语义不通顺等问题,那么用户在通过文本获取信息的难度就会提高;对于图像、视频和音频而言,如果图像视频音频在传输过程中收到了损失,从而会导致接收方获取的图像视频音频出现模糊、失真等问题。因此,也就有了分别针对各种媒体所出现的质量评价技术。

本篇文章将内容分为了四个部分,分别从文本、视频、音频和图像这四个角度来介绍关于全媒体质量评价的发展历程。同时也将四类质量评价领域的方法进行贯通,并对每个领域中的传统方法和基于机器学习、深度学习的方法做了比较。不论从领域类别的角度上,还是方法类别的角度上来说都更加全面。

2 图像质量评价

图像是人类视觉信息的一种来源,往往会含有较多的有用信息。在信息的传播中,图像可以承载更多的含义,现实生活中,我们常常需要通过获取、存储、传输图像等过程进行对图像信息的传递,但这些信息传递并不总是有效或者及时的,过程中又会存在一些模糊、噪声、数据丢失等干扰因素,从而引起图像的质量变化,比如降质或是失真。这又会直接影响到图片信息量的获取以及人们对图片信息最直观的主观感受。这里使用了图像处理的一种基本技术来衡量图像质量的好坏——图像质量评价(Image quality assessment, IQA),这是一种分析图像特征,然后评估图像质量,最终实现图像优化的技术,在图像处理领域占有极

其重要的地位。

在图像质量评价任务中，前常用的合成失真的图像数据库有 PNG 格式的 LIVE（Laboratory for image & video engineering）数据库[4]，CSIQ（Categorical Subjective Image Quality）数据库[5]，BMP 格式的 TID2008（Tampere Image Database）数据库[6]，以及对 TID2008 进行改进的以 Bitmap 格式保存的 TID2013 数据库[7] 等。最近两年出来的真实失真的数据库都有提供图像属性和 EXIF（ exchangeable image file format）信息，比如在 2020 年建立的 KonIQ-10k（Konstanz authentic image quality database）数据库[8]，SPAQ（smartphone photography attribute and quality）数据库[9]等。

在图像质量评价任务中，我们主要选择以下三个主流指标作为评估标准。斯皮尔曼秩序相关系数（SRCC）。它表示基本事实和预测分数之间的单调性；线性相关系数（PLCC），用于衡量 IQA 指标的预测线性；均方根误差（RMSE），用于计算映射分数与地面真相之间的误差。

图像质量评价方法可以分为主观评价方法和客观评价方法，而主观评价又可以根据有无标准参考的条件可以分为绝对主观评价和相对主观评价。主观质量评价的测试结果可以直接体现出人们对于图像质量的主观感受。失真图像的质量指标一般采用平均主观得分（Mean opinion score，MOS）或平均主观得分差异（Differential mean opinion score，DMOS）表示。虽然评价结果可以真实反映图像质量，但总会受到实验者的个人主观因素的影响，加上不能够通过数学模型加以描述，这种方法还需要进行多次人工的重复实验，可想而知耗费时间较长。客观评价方法则是由计算机根据人眼视觉系统算法建立模型来计算得到图像的质量指标，相比较主观评价，该方法不会受人为原因影响受到偏差，根据对于参考图像的依赖性又可以分为全参考质量评价、半参考质量评价和无参考质量评价等三类评价方法。

全参考图像质量评价方法（Full reference，FR）需要选择理想的图像进行参考，比对失真图像之间的差异，从而分析失真图像的失真情况，这类方法的算法可分为：使用像素误差统计[10]；使用人眼视觉特性[11]；使用结构相度等算法。例如 2012 年提出的梯度相似度 GSM[12]，该方法利用梯度的特性来提取出所需要的视觉信息，并将其与像素值相结合得到了良好的实验结果，并且加快了算法的计算速度；基于人类视觉系统与其他算法结合，例如 Liu 等人[13] 提出将结构相似性算法结合至人类视觉系统，利用 HVS 的特点将图像的局部像素和整体的几何拓扑结构相对应来评价图像质量。

半参考评价方法（Reduced reference，RR）只提供参考图像的部分信息或者从中提取部分特征，所以与图像整体相比较，该方法用到的数据量大大减少，灵活性更强。该方法的关键就是提取参考图像和失真图像的部分特征，并将其比较来进行图像质量评价。该方法又可以分为通过原始图像特征[14]、通过 DCT 域[15] 和通过 Wavelet 域统计模型[16] 等多种类型。

如今，无参考图像质量评价方法（No reference，NR）因其不需要使用参考图像、实用性强、应用范围广但实现难度大于有参考图像或特征的方法而成为 IQA 领域的热门话题。本文有关图像质量评估方法主要介绍 NR-IQA，其内容框架如图 1 所示。

2.1　基于传统机器学习算法的无参考图像质量评价

2.1.1　基于传统方法的有监督 NR-IQA 模型

有监督的，基于传统机器学习无参考图像质量评分模型大致可分为自然场景统计（NSS）方法和基于学习的方法，在传统方法中主要总结为特征提取，特征表示和特征映射三步骤，其中特征提取尤为关键。

基于自然场景统计方法的提出是一里程碑式的事件，自然场景统计特征指的是服从一定分布规律的图像特征，针对于不对类型程度的失真会对其产生相应影响的情况，2012 年 Mital 等人[17] 提出了一个基于自然场景统计特征的 NR-IQA 模型，该模型在空间域运行，称为盲/无参考图像空间质量评估器（Blind/Referenceless image spatial quality evaluator，BRISQUE）。该方法如何提取特征向量如下：提取亮度平均对比度归一化系数（MSCN），BRISQUE 在计算时选取四个不同方向分别计算 MSCN，这是由于假设失真会改变 MSCN 的分布情况，然后再将 MSCN 系数合成为非对称广义高斯分布（AGGD），设定其特征是经过广义高斯模型拟合之后获取得到

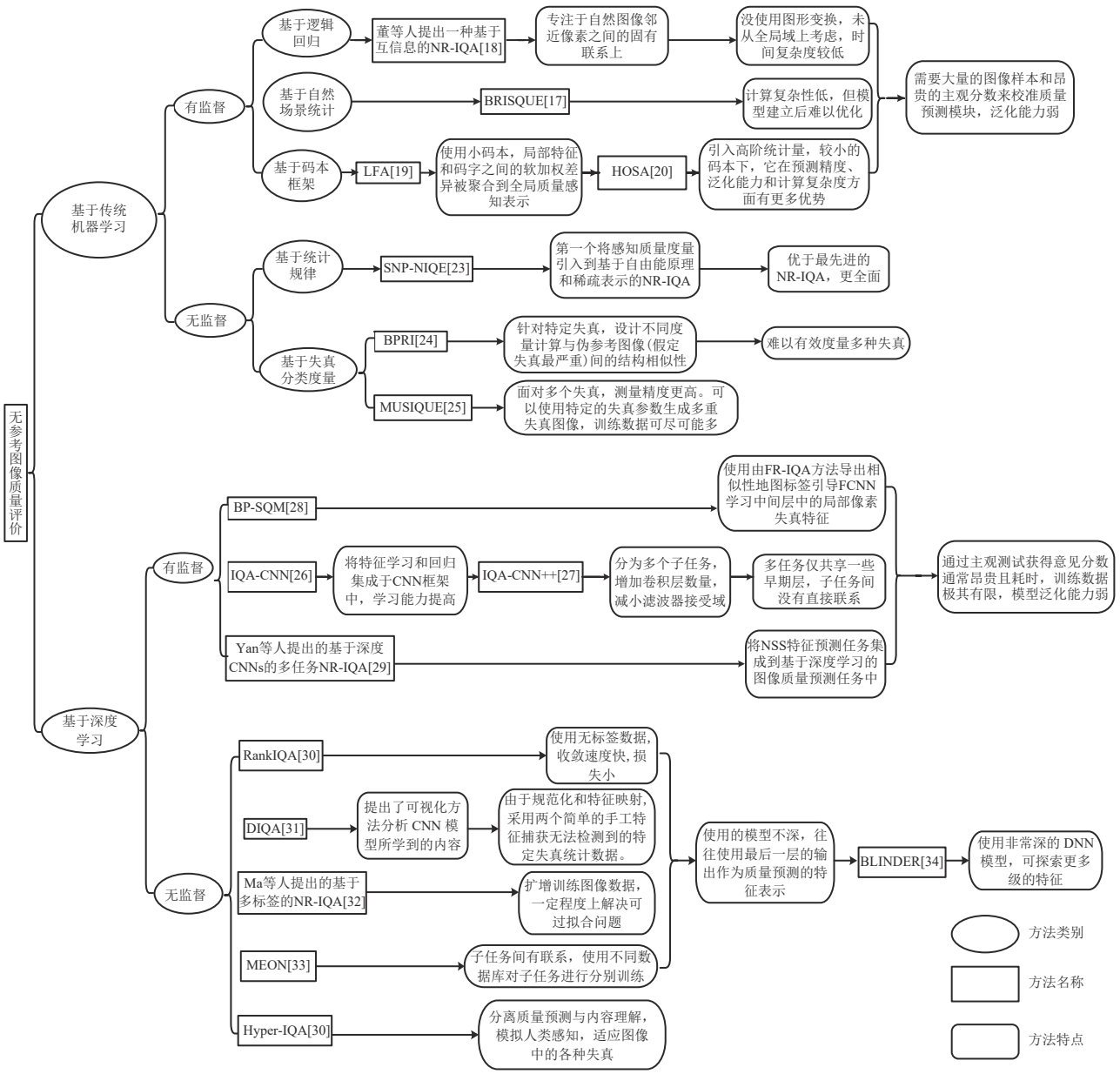

图 1 图像质量评价方法框架图

Fig. 1 Frame of image quality evaluation methods

的模型参数，接着再对特征使用多变量高斯模型进行进一步的描述，然后最后再使用支持向量机(Support vector machine，SVM)进行分类。该方法对图像数据只使用了归一化，使其呈现有规律的分布，且最后确定图像质量是使用通过判断比较失真图像与预先建立模型的特征参数距离的方法，模型简单且计算复杂度低。

MSCN 系数的计算如下：

$$\hat{\boldsymbol{I}}(i,j)=\frac{\boldsymbol{I}(i,j)-\mu(i,j)}{\sigma(i,j)+C} \tag{1}$$

$$\mu(i,j)=\sum_{k=-K}^{K}\sum_{l=-L}^{L}\omega_{k,l}\boldsymbol{I}_{k,l}(i,j) \tag{2}$$

$$\sigma(i,j)=\sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L}\omega_{k,l}\left(\boldsymbol{I}_{k,l}(i,j)-\mu(i,j)\right)^2} \tag{3}$$

$\boldsymbol{I}(i,j)$表示图像$\boldsymbol{I}$在位置(i,j)处的像素值，ω为高斯滤波器(可有效抑制噪声)，$\mu(i,j)$是高斯滤波之后的结果，$\sigma(i,j)$是标准差。

广义高斯分布(GGD)的计算如下：

$$f(x;\ \alpha,\sigma^2)=\frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)}\exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{4}$$

$$\beta=\sigma\sqrt{\frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}} \tag{5}$$

$$\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}\mathrm{e}^{-t}\mathrm{d}t \tag{6}$$

其中x表示像素值，Γ表示伽马函数，σ^2表示GDD模型的方差参数，α则为其形状参数，α值大于0。

但是基于传统有监督自然场景统计的方法只针对于一定的失真有抑制效果，面对现实情况所会出现的各种各样的噪声情况，模型的性能还有待提高。

在传统的基于学习的方法中常见算法包括逻辑回归等，其试图设计一个有效的视觉特征表达方法，而且一般可以使用支持向量机的方法来学习从特征空间到质量分数之间的映射模型。2014年董宏平等人[18]提出一种基于自相关互信息的NR-IQA模型可以对多类失真图像进行客观质量评价，该方法量化图像邻近像素间的相关性，其提取的多尺度特征来源于三种图输入：原始图像、原始图像对应的局部标准差图和亮度图，以及等到最后再使用SVM对该模型进行训练。但该方法没有使用图像变换，模型的时间复杂度较低。

更一种常见的传统的特征提取方法还有基于码本的框架。2015年Xu等人[19]提出了一种基于局部特征聚合的盲图像质量评价(Blind image quality assessment，BIQA)框架，所提出的方法简称为LFA，它所要用到的码本很小，而且不需要更新码本。大型码本包含许多可能干扰图像质量评估的类似码字。相比之下，LFA使用更自然和直接的方式来构建图像质量感知表示。将局部特征和码字之间的软加权差值进行聚合，形成特征向量。整个过程为：将归一化的原始图像块作为局部特征进行提取，Kmeans聚类应用于从CSIQ数据库[5]提取的局部特征，以获得100个码字码本，接着直接计算局部特征和码字之间的软加权差异，以保留最大化的图像信息。最后选择标准SVR学习聚合特征和主观得分之间的映射。

2016年，Xu等人[20]在前期研究内容的扩展上研究如何利用码本和图像之间的统计差异提出了一种基于高阶统计聚集(High Order Statistics Aggregation，HOSA)的BIQA框架。除了每个聚类的平均值，还计算聚类的维度方差和偏斜，以形成一个详细的质量感知码本，以近似低层特征的分布，然后计算局部特征与相应聚类之间的软加权高阶统计量差异，该方法可以应用于各种图像类型，包括自然图像、屏幕内容图像和文档图像。它还可以很好地反映模拟和真实失真对感知质量的影响，泛化能力强。且由于使用了更小的码本，质量感知表示计算具有更快的速度，并且具有应用于实际应用的潜力。

尽管上述有监督的NR-IQA方法可以实现高预测性能，但是它们需要大量的图像样本和昂贵的主观分数来校准质量预测模块。此外，监督方法也可能遭受较弱的泛化能力。

2.1.2　基于传统方法的无监督NR-IQA模型

基于传统机器学习的无监督学习的数据并不被特别标识，样本数据实现并不需要主观评分来进行训练，而是对原有数据上直接建模，那么需要开发图像质量自身感知内容进行参考，目前根据内容的不同可以大致分为根据统计规律进行模型拟合以及根据失真分类度量进行感知质量融合这样两类。以下主要介绍最近的两类方法。

利用自然图像中的统计规律作为参考对于模型的预测性能有所提高，2015年Li等人[21]提出了一种新的无监督特征选择方案，即利用非负谱聚类和冗余分析，进行约束冗余的非负谱分析。该方法可以直接识别最有用和最冗余约束特征的判别子集。开发非负谱分析是为了学习输入图像的更准确的聚类标签，在此期间同时执行特征选择。集群标签和特征选择矩阵的联合学习能够选择最具鉴别性的特征。之后该团队又提出了一种新的半监督局部特征选择方法(S2LFS)[22]，允许为不同的类选择不同的特征子集。根据此方法，通过学习分别考虑每个类的特征的重要性来选择特定于类的特征子集。特别是，所有可用数据的类标签都是在对标记数据的一致约束下共同学习的，这使得所提出的方法能够选择最具辨别性的特征。2020年Liu等人[23]通过对原始自然图像中的失真图像的结构、自然度

以及其感知质量的度量，建立了一种新的无监督图像质量评价方法（SNP-NIQE），通过局部平均相减和对比度归一化（Mean subtracted contrast normalized，MSCN）系数和相邻MSCN系数对的乘积的分布变化来表征自然度变化。这里首次将感知质量度量引入到无监督质量评价的方法中。设计并提取三种有效的自然统计（Natural scene statistics，NSS）特征，分别表征结构、自然度和感知质量。在特征提取之后，从一组原始图像中学习具有质量感知特征的原始MVG模型，作为质量预测的“参考”。问题图像的MVG模型和所学习的原始MVG模型之间的距离被定义来测量问题图像质量。

传统的IQA度量方法一般为显式或隐式地对失真图像与完美质量图像的偏差进行测量，以此来预测图像质量，但Min等人（2018）[24]提出“伪”参考图像的概念打破了这一方式，并提出了一种基于优先级的NR-IQA模型。与传统的参考图像被认为有完美的质量不同，PRI是通过失真图像来生成的，认定其失真最大，为了能够进一步地模糊当前模糊图像去获得PRI，该方法使用了特定的平滑滤波器，同时还将一定强度的噪声加在了当前有噪声的图像。该方法开发了基于优先级的特定的质量度量来估计块效应、锐度和噪声。然后，通过失真识别后的两阶段质量回归框架，将基于PRI的失真特定度量集成到通用BIQA方法中，其框架如图2。然而事实是图像中往往不止一种失真，这就体现了该模型的一个局限性：难以同时有效度量同一图像中的多种失真。

与之不同的是Zhang和Chandler（2018）[25]提出了基于使用自然场景统计（NSS）特征预测失真参数，盲目评估多重失真和单失真图像的质量的方法（MUSIQUE），可面对多个失真，且测量精度更高。首先识别图像中可能存在的失真类型，然后通过学习不同失真类型和组合的不同NSS特征，使用特定的回归模型来预测高斯模糊、JPEG压缩和白噪声三个参数，最后根据质量映射曲线和最明显失真策略，将三个估计的失真参数值映射并组合成整体质量估计。

2.1.3 基于传统方法的NR-IQA模型性能分析

统计了多篇基于传统机器学习的NR-IQA方法在LIVE图像数据集和TID2013数据集上SROCC和PLCC两个指标数据，结果如表1所示。

表1中的前三行是基于有监督的传统机器学习的NR-IQA方法的实验结果，后三行则是基于无监督的实验结果。从中不难看出，在LIVE数据集上这些方法的SROCC和PLCC两个指标都在0.90以上，使用MOS做监督训练的BRISOUE与无监督的SNP-NIQE实验结果相比，各有所长，指标性能各有突出，这显然可以得出基于NSS的模型可以不用大量数据进行监督训练，基于无监督的传统机器学习方法前景广泛。且HOSA在TID2013数据集上的表现优异可达到0.95以上，这表明其所提出的特征聚合方案能够更好地表示不同图像内容和失真类型的图像质量。

BPRI于众方法中在CISQ数据集上的RMSE指标上表现最优，且与其他方法差距较大。BPRI方法使用的质量特征并不局限于自然场景，大多数NA-IQA模型在接近常见失真的失真上表现出良好的性能，但在对比度变化等独特失真方面表现不佳。

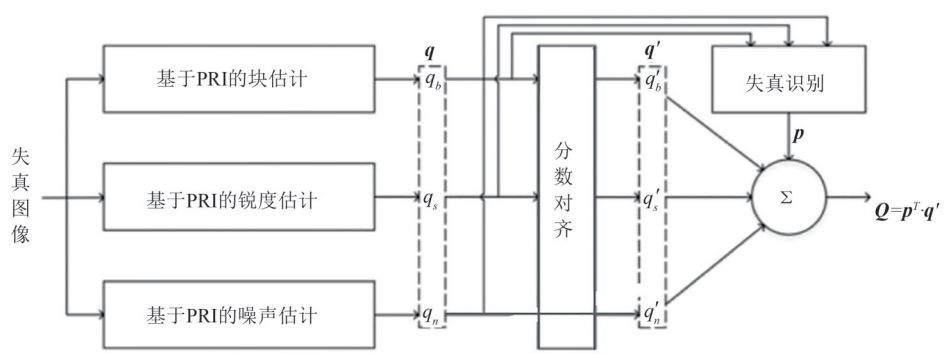

图2 基于PRI的通用BIQA度量的框架[19]

Fig. 2 PRI-based general BIQA measurement framework[19]

表 1　基于传统 IQA 方法在 LIVE、CSIQ 和 TID2013 数据集上的性能

Tab. 1　Performance of traditional IQA methods on LIVE, CSIQ and TID2013 datasets

算法 \ 数据集	LIVE-IQA		TID2013			CSIQ		
	SROCC	PLCC	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
BRISUQE(Mittal 等[17],2012)	0.939	0.942	0.573	0.651	/	0.775	0.817	/
LFA(Xu 等[19],2015)	0.944	0.946	0.681	0.758				
HOSA(Xu 等[20],2016)	0.950	**0.953**	**0.952**	**0.959**	**0.394**	**0.93**	**0.948**	0.089
SNP-NIQE(Liu 等[23],2020)	/	/	0.857	0.847	0.742	0.901	0.906	0.119
BPRI(Min 等[24],2018)	0.930	0.932	0.894	0.88	0.659	0.899	0.919	**0.012**
MUSIQUE(Zhang 等[25],2018)	**0.963**	/	0.903	/	0.515	/	/	0.092

但是，除了 HOSA 在 TID2013 上的优异表现，其他模型在相比较于其他的数据集上，它们在 TID2013 上的实验效果就差了很多。这或许是因为早期的传统方法是针对于一定的数据集来手工设计相应特征，这就导致模型的泛化能力受到了限制，在面对多种失真的情况下，模型学习能力有限。

2.2　基于深度学习算法的无参考图像质量评价

在基于深度学习的算法中，图像到图像质量的映射这类关系能够采用端到端的方式加以学习，而且此类算法往往会运用卷积神经网络（Convolutional neural network，CNN）的模型。

2.2.1　基于深度学习的有监督 NR-IQA 模型

与传统方法借助手工设计特征来学习的方式不同，基于深度学习的有监督的 BIQA 模型通常需要依靠失真的图像及借助 MOS 来学习，然后进行特征映射得到图像质量结果。

2014 年，Kang 等[26]使用卷积神经网络（CNN）来学习 NR-IQA 任务的判别特征。该方法可以说开启了 IQA 进入深度学习时代的大门，极大的提高了图像质量评价算法的鲁棒性。为了能够加深网络并且提高学习能力，在 CNN 框架中，该方法使用了特征学习和回归，还可以使用反向传播来训练整个网络，能够更好的改进 dropout 和纠正线性单元的技术，并更方便的进行技术结合。2015 年，Kang 等[27]基于 IQA-CNN 进一步提出一个紧凑的多任务卷积神经网络（CNN），将 NR-IQA 任务分成多个子任务，可同时估计图像质量和识别失真，通过增加卷积层数量，修改全连接层，减小滤波器的接受域来提高学习能力，获取更多的信息。

Pan 等人（2018）[28]提出了一个基于深度 CNN 的 NR-IQA 模型（BP-SQM），其架构如图 3 所示，一个基于由卷积神经网络组成的新框架（FCNN）和一个深度池化网络（DPN）来有效模拟人类视觉的属性

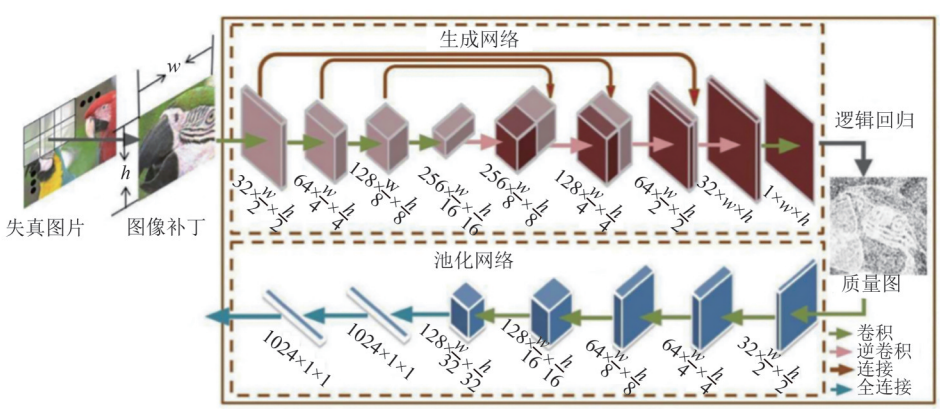

图 3　BP-SQM 架构[28]

Fig. 3　BP-SQM architecture[28]

数据驱动的系统，它可以指引网络生成不同的图像质量图，前提是给定一个相似性的索引图标签，该模型可以生成一个与人眼视觉相关的质量图，从而在像素畸变水平上逼近相似性索引图，用一个主观评分标签来进行指导训练，融合进生成的不同类型的质量图，再至DPN中进行回归。

2019年Yan等人[29]选择开发一种基于CNNs的方法，并利用基于自然图像统计方法的优势来提高基于CNNs方法的泛化能力，由此提出了一种自然场景统计辅助深度神经网络（NSSADNN）用于NR-IQA，该网络是通过多任务学习方式设计的，自然场景统计（NSS）特征预测任务和质量分数预测任务。NSS特征预测是一项辅助任务，它有助于质量预测任务提高表征能力，该模型主要是将NSS特征预测任务集成到基于深度学习的图像质量预测任务中，以提高表示能力和泛化能力。

上述均为基于深度学习的有监督的NR-IQA模型，训练可靠的有监督的图像质量评价模型要有较多的人工评分样本数据作为支持，主观测试来获得人工感知意见分数的数据过程往往耗时且操作繁杂。且这类模型的泛化能力也不强。

2.2.2 基于深度学习的无监督NR-IQA模型

基于深度学习的无监督的NR-IQA模型的训练并不需要人工评分数据的支持，例2017年Liu等人[30]提出一种基于从排序图像数据集中学习的NR-IQA方法（RankIQA），训练一个Siamese网络，该方法不依靠含有MOS的图像而是通过使用图像的失真程度来对图像质量的排序，同时还提出了一种有效的反向传播方法，提高收敛速度。

2018年，Kim等人[31]提出深度图像质量评估器DIQA，DIQA的训练过程包括两个阶段：客观失真部分和人类视觉系统相关部分。使用两个独立的CNN分支，每个分支分别用于学习客观失真和人类视觉敏感性，视觉敏感度分支通过查看扭曲图像的三元组、其客观误差图和其基本真实主观分数来预测客观误差图的局部视觉权重第二阶段模型学习预测主观得分。客观误差图和灵敏度图相乘，得到一个感知误差图，可以从HVS的角度解释失真程度。

在训练的第一阶段，目标误差图被用作代理回归目标，以获得增加数据的效果。损失函数由预测误差图和地面真实误差图之间的均方误差定义，表达式为：

$$L_1\left(\hat{\boldsymbol{I}}_d;\ \boldsymbol{\theta}_f;\ \boldsymbol{\theta}_g\right)=\left|g\left(f\left(\hat{\boldsymbol{I}}_d;\ \boldsymbol{\theta}_f\right)-\boldsymbol{e}_{gt};\ \boldsymbol{\theta}_g\right)\otimes\hat{\boldsymbol{r}}\right|_2^2 \tag{7}$$

$$\boldsymbol{e}_{gt}=\left|\hat{\boldsymbol{I}}_r-\hat{\boldsymbol{I}}_d\right|^P \tag{8}$$

$\boldsymbol{\theta}$为CNN参数，$\hat{\boldsymbol{I}}_r$为参考图像的高频信息图，$\hat{\boldsymbol{I}}_d$为畸变图像的高科技信息图，P为指标参数。

$$\boldsymbol{r}=\frac{2}{1+\exp\left(-\alpha\left(\left|\hat{\boldsymbol{I}}_d\right|\right)\right)}-1 \tag{9}$$

其中$\boldsymbol{r}$是纹理和平面部分的重量分布（高频部分的重量大大增加），α控制可靠性图的饱和特性。由于模型输入了畸变图像的高频信息，结合损失函数，该系数可以消除平坦部分对预测误差图的不利影响。为了标准化可靠性图，在式中使用了sigmoid函数的正半部分，以便为具有小值的像素分配足够大的可靠性值。

为防止可靠性图直接影响预测得分，将其除以其平均值，式中$\hat{\boldsymbol{r}}$为可靠度图：

$$\hat{\boldsymbol{r}}=\frac{1}{\frac{1}{H_rW_r}\sum_{(i,j)}\boldsymbol{r}(i,j)}\boldsymbol{r} \tag{10}$$

其中H_r和W_r是$\boldsymbol{r}$的高度和宽度。

一旦模型被训练来预测客观误差图，就进入下一个训练阶段，为了补偿丢失的信息，考虑了两个额外的手工特征：非标准化可靠性图$\boldsymbol{\mu}_r$的平均值和失真图像的低频标准差$\boldsymbol{\sigma}_{I_d^{\text{low}}}$。

损失函数定义为：

$$L_2\left(\boldsymbol{I}_d;\ \boldsymbol{\theta}_f,\boldsymbol{\theta}_h\right)=\left|\left(h\left(\boldsymbol{V},\boldsymbol{\mu}_r,\boldsymbol{\sigma}_{I_d^{\text{low}}};\ \boldsymbol{\theta}_h\right)-\boldsymbol{S}\right)\right|_2^2 \tag{11}$$

其中$\boldsymbol{S}$是输入失真图像的基本真实主观得分，$\boldsymbol{V}$是汇集的特征向量。$\boldsymbol{V}$的定义如下：

$$\boldsymbol{V}=\text{GAP}\left(f\left(\hat{\boldsymbol{I}}_d;\ \boldsymbol{\theta}_f\right)\right) \tag{12}$$

其中GAP表示全球平均操作。

不同于Liu等使用的无标签数据，2019年Ma等人[32]提出一个基于多标签学习的NR-IQA模型，由于很多对基于卷积神经网络（CNN）的数据驱动BIQA模型都是根据平均意见分数MOS进行训练的，而这些数据往往无法训练以百万计的大量的模型参数。将若干种人工合成失真添加到高清图像数据中获取大量的图像数据对，并使用多个IQA注释器来计算二进制标签，指示两个图像中哪一

个质量更高，然后我们训练CNN使用成对学习排序算法计算质量分数和相关的不确定性。每个IQA注释器和CNN参数的可靠性通过最大化其可能性进行联合优化。由于图像数据不足，IQA模型很有可能会出现模型的过拟合问题，而该方法对使用的训练图像数据进行扩增，在很大程度上解决了该问题。

2018年，Ma等[33]提出一个基于多任务学习的端到端优化的深度神经网络MEON。这里将BIQA问题分解为两个子任务：子任务1从一组预定义的类别中将图像分类为特定的失真类型；子任务2利用从子任务1获得的失真信息预测同一图像的感知质量。该方法定义了一个与卷积激活和子任务1的输出都不同的层，以保证反向传播的可行性。在预训练之后，使用随机梯度下降法的一种变体对整个网络进行端到端优化，还使用广义分裂归一化GDN联合非线性作为激活函数，可保持相似的质量预测性能。

以上提到的RankIQA模型[30]、DIQA模型[31]、MEON模型[33]等往往使用最后一层的输出作为特征表示，使用的模型都很浅。2018年，Gao[34]等人开发了一个在图像分类任务中预先训练的非常深的DNN模型用于特征提取，然后使用浅层学习技术进行质量预测——BLINDER模型。在BLINDER中，该方法在每一层当中提取特征向量，并使其得到每一层的质量分数并进行平均，来得到最后的质量。该方法使用支持向量回归(SVR)来学习预测。

而且其浅层架构的网络不能很好地处理真实失真，提取的特征会随着图像的变化而发生变化，从而导致预测结果偏离真相，但深度模型只学习用于分类的全局特征，当图像其余部分显示极好质量时人体视觉系统对局部失真的敏感性也会有所提高，为此2020年，Su等人[35]提出了一种基于超网络的NR-IQA模型(Hyper-IQA)，该模型能够自适应地调整质量预测参数，该网络以内容感知的方式预测图像质量，即将质量预测与内容理解分离，以模拟人类对图像质量的感知，适应图像中的各种失真。

2.2.3　基于深度学习的NR-IQA模型性能分析

统计了多篇基于深度学习的NR-IQA方法在LIVE图像数据集和TID2013数据集上SROCC和PLCC两个指标数据，结果如表2所示。

表2　基于深度学习的NR-IQA方法在LIVE-IQA和TID2013数据集上的性能

Tab. 2　Performance of some deep learning-based NR-IQA methods on LIVE-IQA and TID2013 datasets

算法 \ 数据集	LIVE-IQA		TID2013	
	SROCC	PLCC	SROCC	PLCC
CNN(Kang等[26], 2014)	0.956	0.953	0.718	0.778
BP-SQM(Pan等[28], 2018)	0.973	0.963	0.862	0.885
NSSADNN(Yan等[29], 2019)	**0.986**	**0.984**	0.844	**0.910**
Rank-IQA(Liu等[30], 2017)	0.980	0.982	0.780	0.792
DIQA(Kim等[31], 2018)	0.975	0.977	/	/
BLINDER(Gao等[34], 2018)	0.966	0.959	0.819	0.838
Hyper-IQA(Su等[35], 2020)	0.983	0.966	**0.903**	**0.910**

表2的前三行是基于有监督的方法，可以看出在监督下的NSSADNN在这两个数据集和指标下性能更优，后四行是基于无监督的方法。可以从表中看到即使是最早的深度学习方法CNN在LIVE数据集上，SROCC和PLCC两个指标都在0.95以上，展现出深度学习的应用在该领域的潜力无限。

TID2013数据集包含失真的类型范围很广，是一个更具挑战性的IQA数据库。随着近些年的深度学习的发展，在面对失真多样性的问题上，从表中可以看到后来的深度学习方法有很好的得到改善，TID2013数据集下的SROCC和PLCC都有所上升。Rank-IQA对于特定应用场景简单有效，其在LIVE数据集上的两个指标性能都在0.98以上，但在TID2013数据集上可以反应其对于不同失真表现的效果欠佳，算法不够稳定。而BLINDER的结果显示在不训练特定DNN模型的情况下学习还是非常有效果的。

而作为较新提出的算法NSSADNN和Hyper-IQA在两指标上表现都很好，Hyper-IQA针对失真内容以及失真多样性的问题采用内容理解模块以及学习人类对图像质量的感知规律，与NSSADNN最根本的不同则是在于无监督下，虽然在该两指标上稍落于NSSADNN，但不需要人工评分数据的支持，节省劳力。

随着深度学习的发展以及在该领域的不断研究，模型的泛化能力已经有很大提升，但在TID2013数据集上的总体情况来看，解决失真的多样性问题

方向仍有欠缺。在现实生活中要面对的失真种类复杂多样,能应对多种失真情况的方法还待更好的去研究。而且伴随着深度学习应用的深入,模型结构趋于复杂化,参数量的增加也会带来计算量扩增,模型效率降低的问题。但是深度学习方法带来的好的结果是有目共睹的,其应用仍是图像质量评价领域的主流。

3 视频质量评价

在当前环境下,移动互联网信息技术得到了迅猛的发展,人们使用各种媒体设备的频率越来越高,观看视频的时间也越来越长,视频被传输和分享的次数也越来越多。因此对于视频的质量高低与否,也成了人们关注的问题。但是视频在压缩或者传输时,容易发生失真、丢包或者受到一些高斯噪声的损伤,从而导致视频质量降低。由此可见,对视频的质量做出评估是比较重要的。视频质量评价(video quality assessment, VQA)是视频服务系统中的重要技术,该技术可以对视频编码器的性能进行评估和测量,同时也能够对视频的质量起到一定程度的监测作用,能够有效的提高视频的质量。视频质量评价分为主观视频质量评价和客观视频质量评价。

其中,主观视频质量评价方法分为质量评价和损伤评价两种方式,用于对视频质量的优劣程度进行评价,从而确定视频系统性能的好坏的方法为质量评价。损伤评价是指视频经过压缩、编码、解码等多个环节后,通过视频在这些环节中受到的损伤,然后再对视频的质量进行评估。主观评价法需要通过评价人员的评估得出评价的结果,评价人员可以是专业的人员,也可以是非专业的人员。虽然对于评价人员的专业性要求不高,但是对于评价人员的选择要有比较广泛的代表性,该评价人员应该具备一定的分析判断能力。同时,为了保证最后得到的评价数据有一定的可靠性,应该最少要选择15个评价人员。在评价过程中,评价人员首先要严格按照已经规定好的评价标准再结合自己的经验对视频的质量进行评估;然后综合所有的评价人员的结果,将质量由高到低分为五个等级。由于该方法的评价标准容易建立,而且实施起来比较方便,所以这种方法是目前常用的视频图像质量主观评价方法。

客观质量评价是指让计算机通过已经设计好的模型和算法来对视频的质量进行自动评估。客观质量评价与主观方法相较而言,具有成本低、速度快、容易实现等优点,并且还能对视频质量进行实时的监控。客观方法在结果上要尽量与主观方法的结果接近,可以通过评价结果的一致性、正确率和稳定性来判断客观质量评价方法的优劣程度。

视频的客观质量评价可以分为三种类别:无参考、部分参考和全参考,这种分类是根据需要评价的视频在评价过程中对原始视频的依赖程度来划分的。因为现实生活当中,在进行质量评价时很难得到用来参考的视频,所以无参考视频评价方法有着巨大的研究应用价值,该方法也逐渐成为视频质量评价方面的研究热点。本文有关视频质量评估方法的内容框架如图4所示。

3.1 主观视频质量评价

比较常见的主观视频质量评价方法与图像方法类似,包括SSM、SSCQE、DSIS、DSCQS等。

主观评价可能会面临许多的干扰,比如评价人员所处的环境变化以及评价的时长和距离,还有评价人员观看原始视频和待评价视频的先后顺序,这些因素都可能影响到主观评价的最后结果。而且评价人员还可能受到时间掩蔽效应[36]的影响,如果评价的视频当中存在幅度比较大的运动时,或者存在比较鲜明突出的事物,观察者的注意力可能就集中在这些事物上,从而忽略了视频中其他部分不太明显的变化,导致最后的评价结果受到影响。同时主观视频质量评价需要消耗大量的人力、物力以及资源,所需成本较大,因此更加先进的客观视频质量评价方法的需求越来越大。

3.2 传统视频质量评价方法

3.2.1 有参考视频质量评价

由于传统的客观方法,如MSE和PSNR[37]等,没有能够与人眼的视觉特性相结合,可能会导致最后的评价结果和人眼实际的观测效果不一致,因此基于人眼视觉特性(human visual system, HVS)的算法受到研究者们的广泛关注。Seshadrinathan等人[38]提出了一种利用了人眼的延迟效应,并将该效应通过池化的方式去计算帧级之间的SSIM值的方法。同时又提出了MOVIE[39]方法,该方法使用了评估动

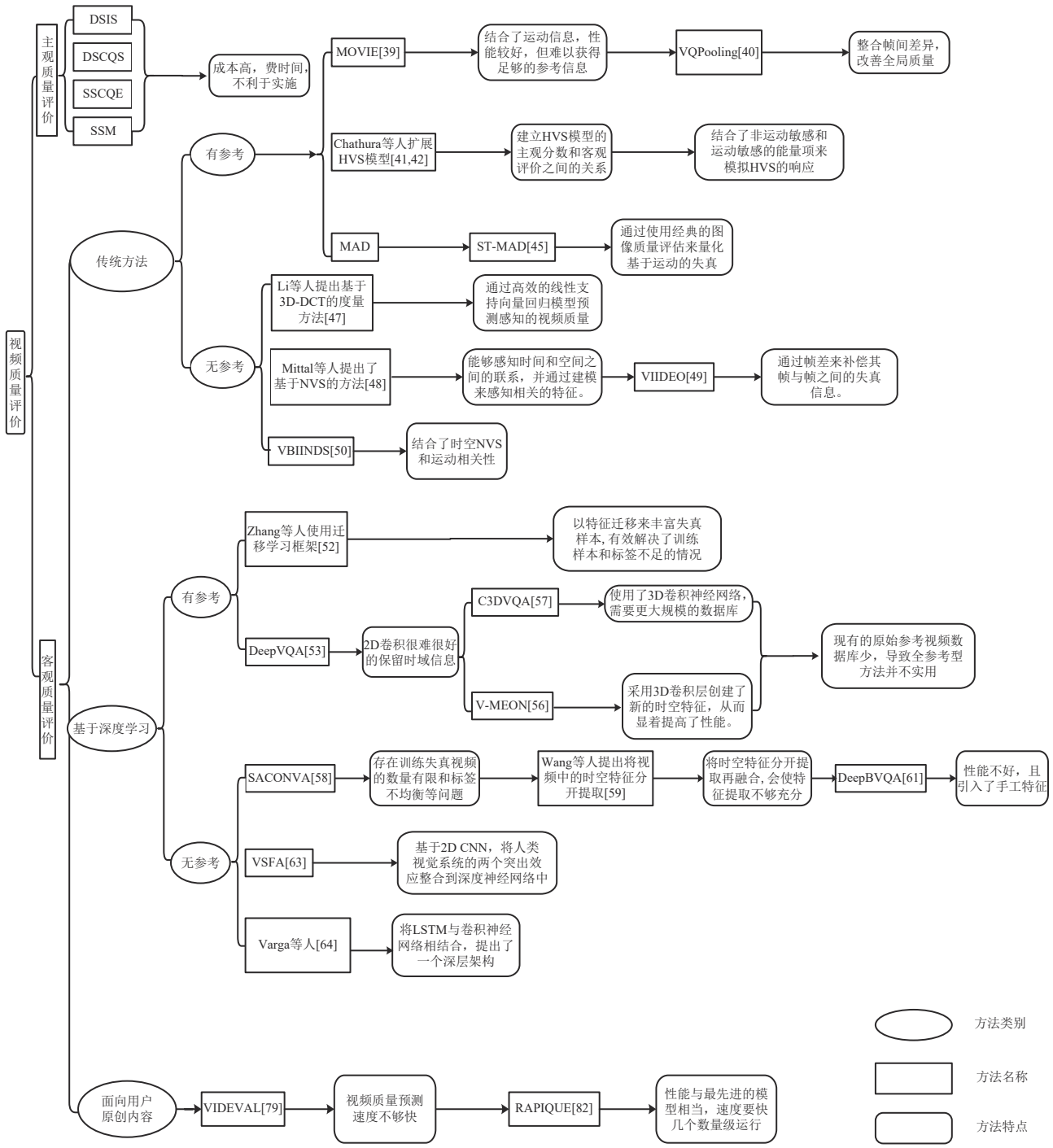

图 4　视频质量评价方法框架图

Fig. 4　Frame of video quality evaluation methods

态视频保真度的空间光谱局部多尺度框架，此框架对空间和时间方面进行失真评估，同时能够结合运动信息，展现了比较优秀的评价性能，但是需要依赖比较多的参考信息，并且这些信息往往难以收集。之后 Park 等人在此基础上提出了一种自适应池化方法 VQPooling[40]，该方法通过对视频中每一帧的质量进行整合来解决评价过程中不同帧之间质量变化大的问题，从而改善了整个视频的全局质量。

Chathura 等人[41]提出了基于人类视觉系统（HVS）模型的全参考立体图像和视频质量指标，该

模型结合了双目视觉的重要生理学发现。它引入了一种新的HVS模型，扩展了以前的模型，包括双眼抑制和反复激发的现象。最后，引入了优化的时间池策略以将评估扩展到视频域。图像和视频质量指标都是通过训练过程获得的，以建立HVS模型的主观分数和客观评价之间的关系。之后该团队在2020年又提出了一种新的HVS模型[42]，其灵感来自生理学发现，该模型通过估计光流以测量不同尺度和方向的场景速度来表征简单和复杂细胞的行为，它独特地结合了非运动敏感和运动敏感的能量项来模拟HVS的响应。其中，运动响应加权客观分数的计算公式为：

$$\overline{\boldsymbol{X}}_i=\sqrt[\beta]{\frac{1}{f}\sum_{t=1}^{f}\left|\boldsymbol{X}_i(t)\boldsymbol{w}_h(t)\right|^{\beta}} \tag{13}$$

关键思想是测量特定类型运动在每一帧中的表现程度，并使用此信息来通知池权重的选择。对于给定的感知通道c，t帧的运动支持被定义为具有非零运动响应的像素数。用$\boldsymbol{w}_h(t)$表示运动支持，即第i个目标分数随时间$\boldsymbol{X}_i(t)$取值的序列，t=1…f被合并成一个单一的运动响应加权客观分数。

许多评价方法利用结构信息来完成视频质量评价，Wang等人[43]对参考视频和测试视频的对应帧进行计算，分别得到其SSIM值，并通过运动信息来对其进行加权平均从而得到最后的质量评价。Chen等人[44]对图像进行处理，并获取其显著信息，然后通过该信息使用SSIM算法进行信息处理，从而得到对视频评价的预测结果。该方法利用统计方法，对待评价的视频或者图像的局部亮度进行归一化处理，根据处理过后的信息来推测待评价的视频可能产生了多大程度的失真，之后根据失真信息来得到最后的评测。

由于基于结构信息的方法大部分只着重处理了在空间维度的信息，而在时间维度的信息并未充分利用，许多研究者在原有的基础上对时间信息进行处理，来完善评价方法。Phong等人[45]在之前基于图像的算法（MAD）基础上进行了扩展，提出了ST-MAD方法，该方法对原始视频和失真视频进行基于时间的切片，这使得人们可以通过使用经典的图像质量评估来量化基于运动的失真。Wang等人[46]对视频中的结构信息进行扩展，在考虑空间信息的同时也提取了时间维度的信息，并将三维结构张量作用于空间边缘特征和时间运动信息的提取。

3.2.2 无参考视频质量评价

在多个视频处理和计算机视觉应用中设计通用无参考视频质量评估（NR-VQA）模型是一项重要任务。但是，大多数现有的NR-VQA指标都是为特定的失真类型设计的，这些类型在实际应用中通常无法察觉。Li等人[47]提出了一种基于3D离散余弦变换（3D-DCT）域时空自然视频统计的新型NR-VQA度量。在所提出的方法中，首先基于3D-DCT系数的统计分析提取了一组特征，以表征不同视角下视频的时空统计。这些功能用于通过高效的线性支持向量回归模型预测感知的视频质量。

无参考视频质量评价在不依赖于原始视频的情况下，也常常使用统计的方法将得到的视频质量向期望的真实值进行拟合。Mittal等人[48]提出了基于空间域自然视频统计（natural video statistic，NVS）的模型，该模型能够感知时间和空间之间的联系，并通过建模来感知相关的特征。之后其又提出了基于帧差的带通滤波系数来提取特征从而预测视频质量的方法VIIDEO[49]，该方法通过帧差来补偿其帧与帧之间的失真信息。同样是对帧进行处理，Saad等人[50]提取运动的信息并将其与时空NVS相结合，提出了一种无参考的评价方法VBLIINDS，该方法是对图像质量评价方法BLIINDS-II的扩展。其原理是对视频帧进行处理，通过将帧与帧之间的变换系数拟合NIQE方法和模型提取的特征，之后输入SVR来映射为最后的视频质量。

3.3 基于深度学习的视频质量评价方法

随着卷积神经网络的出现，研究人员发现使用深度学习可以提取出视频中的更多特征和信息，这些信息使得客观视频质量评价的结果更加趋近于主观评价。Callet等人[51]首次提出使用卷积神经网络应用到客观视频质量评价上，该方法虽然仅仅解决了对SSCQE方法的预测问题，但是开辟了一条从传统方法通向深度学习方法的道路。Kang等人[26]提出了一种使用CNN进行无参考图像质量评价（image quality assessment，IQA）的方法，这是卷积神经网络与视频图像质量评价方向结合的一大进步。

3.3.1 有参考视频质量评价

图5是使用深度学习的全参考评价方法的流程

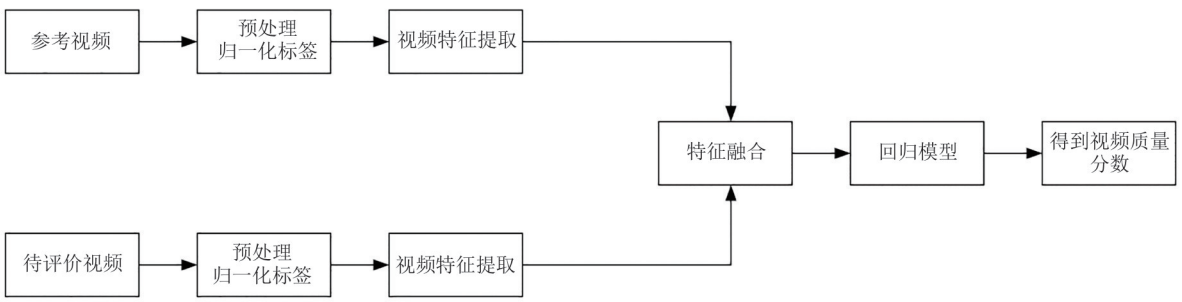

图5　FR-VQA方法流程

Fig. 5　FR-VQA method flow

图。FR-VQA方法流程大致分为四个步骤：(1)对视频数据进行预处理，(2)对视频的特征进行提取，(3)将提取到的视频特征进行融合，(4)建立回归模型输出视频的质量分数。

预处理指的是对输入视频数据的时间长度以及每一帧视频图像的长和宽进行归一化处理。输入前要将视频分解为一帧一帧的图像或者是连续几秒的视频块，然后再将其进行输入。然后使用卷积神经网络对预处理过后的参考视频和待评价视频数据分别进行特征提取，并获得其相应的时空特征。然后将两部分的视频特征进行融合，得到融合后的时空特征。接着将融合后的特征和原始参考视频的主观评价分值作为输入，得到回归模型。最后通过得到的回归模型输出失真视频的质量分数。

由于FR-VQA过于依赖样本数据，样本数据不够充足会对该方法产生的结果产生较大影响，而就目前而言同时包含失真视频和原始的参考视频的数据库中，样本数据的资源非常匮乏，并且已有的样本还存在分布不平衡、失真程度多样、标签不够完善等诸多问题。Zhang等人[52]针对样本数据不足的问题，使用迁移学习框架来提取特征，将待评价的视频进行预处理，以特征的转换来对失真的样本进行评估。该方法能够减少在质量评价过程当中预测不准确标签的影响。但是，该方法的模型更具复杂性。

另一方面，视频质量评价的目的是使得最后的评价结果能够更加准确地与人眼的感知质量相贴合。因此，Kim等人[53]提出了一种名为深度视频质量评估器(DeepVQA)的新型全参考(FR)VQA框架，该框架如图6所示，以通过卷积神经网络(CNN)和卷积神经聚合网络(CNAN)量化时空视觉感知。该方法加入了“注意力机制”的思想[54-55]。使用了CNAN的模型相比较未使用该模型的方法而言，展现了更优秀的整体预测性能。

但是，由于2D卷积容易丢失时域上的信息，为了能够更好地保留视频的时域信息，3D卷积被提出用来处理视频的信息。Liu等人对图像质量评价的MEON[28]方法进行了改进提出了视频多任务端到端优化的深度神经网络(V-MEON)[56]方法，该方法采用3D卷积层创建了新的时空特征，从而显着提高了性能。其将特征提取阶段和回归阶段合并为一个阶段，其中特征提取器和回归器联合优化，可预测最终的质量分数。该方法首先对早期卷积层进行预训练，提取与时空质量相关的特征。然后初始化预训练的特征提取器，将整个网络与两个子任务联合优化。

Xu等人[57]提出了一种新颖的架构，即C3DVQA(Convolutional neural network with 3D kernels (C3D) for video quality assessment)，该网络结构图如图7所示，它使用带有3D内核的卷积神经网络(C3D)来完成全参考VQA任务。C3DVQA将特征学习和分数池结合到一个时空特征学习过程中，并使用2D卷积层来提取空间特征，使用3D卷积层来获得时间上的特征，捕获视频的时间掩蔽效应。

该方法在失真阈值掩蔽之后使用全局平均池化层来表示感知失真的程度。两个全连接层用于学习感知失真和主观质量之间的非线性关系。然后，所提出架构的目标函数定义为：

$$L\left(\boldsymbol{x}_n, \boldsymbol{y}_n;\ \theta\right) = \lambda_1 \left| f_\theta\left(\boldsymbol{x}_n\right) - \boldsymbol{y}_n \right|_2^2 + \lambda_2 L_2 \quad (14)$$

其中λ_1和λ_2是超参数，$\boldsymbol{x}_n$表示失真视频，$\boldsymbol{y}_n$是主观

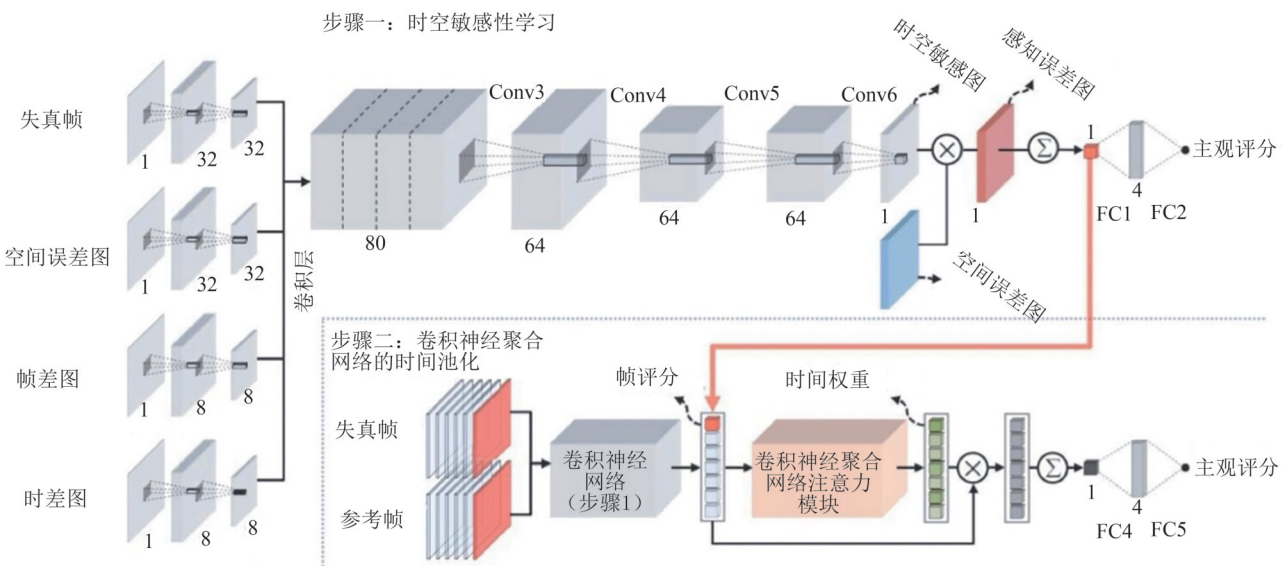

图6 DeepVQA网络架构[53]

Fig. 6 DeepVQA network architecture[53]

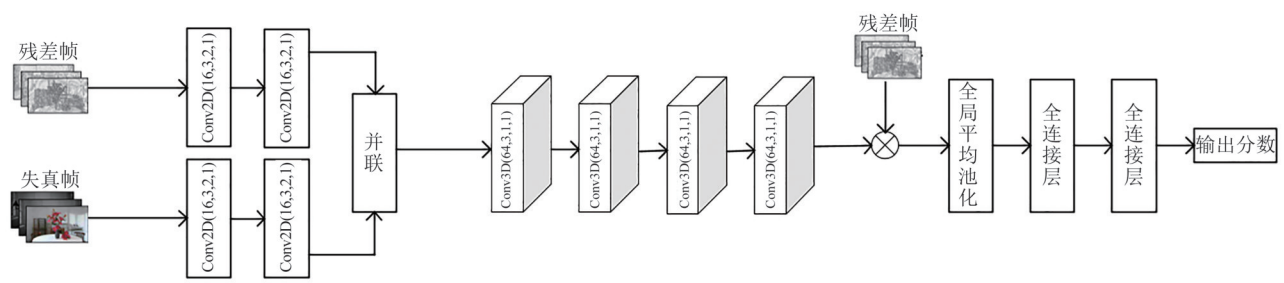

图7 C3DVQA网络结构[57]

Fig. 7 C3DVQA network architecture[57]

质量得分，$f_\theta(\cdot)$表示参数为θ的预测系统，L_2表示正则化项。

虽然随着2D卷积神经网络和3D卷积神经网络的引入，可以使FR-VQA方法得到比较满意的评价结果，但是由于目前现有的数据库当中原始参考视频的数量较少，样本数据也面临一定程度上的短缺，即使2D卷积神经网络模型能够采用迁移学习的方法缓解这一问题，但是在迁移学习的过程当中会引入大量的参数，而且2D卷积神经网络也无法充分利用时域信息；而3D卷积的泛化能力还有待提高，且实际问题中也没有原始视频进行参考，导致全参考模型并不适用。综上所述，全参考的评价方法并不适合解决实际问题，因此无参考视频质量评价方法得到了研究者的广泛关注，也具有更大的发展前景。

3.3.2 无参考视频质量评价

NR-VQA是无参考视频质量评价又称为盲视频质量评价（blind video quality assessment，BVQA），该方法不需要依赖原始参考视频，只需要提取待评价视频的特征就可以进行质量评估。与FR-VQA相比，NR-VQA适用范围更加广泛，只需要充分利用好失真视频的信息便可以进行质量评估。如图8是无参考视频质量评价方法的一般框架结构图。该方法相比较全参考视频质量评价方法，去掉了对参考视频进行操作的部分。

对于无参考视频质量评价的模型，要求其能够对任何失真类型都适用，并且评价结果也要尽可能的与人类主观视觉的感知一致。如Li等人[58]从NR-IQA方法中得到启发，提出了SACONVA（Shearlet and CNN-based NR-VQA）方法，该算法先将待测

图8　NR-VQA方法流程

Fig. 8　NR-VQA method flow

视频进行分割，使其变为一个一个的视频块，然后将这些分割好的视频块进行特征的提取，这里提取时空特征的方法用到了三维剪切波变换，因为该方法能够有效处理时域信息，在提取完时空特征之后，接着使用平均池化方法对得到的特征进行处理，最后建立回归模型对该视频的质量进行预测。不过，虽然该算法能够有效地贴合人类主观视觉的感知，但由于样本数据的数量少和标签乱影响了算法的性能和泛化能力。

由此，Wang等人[59]提出CNN-MR框架（如图9所示），该方法将视频中的时间和空间特征分开进行提取。对于视频中帧与帧之间的空间质量特征，该方法使用了CNN进行提取；而对于视频中基于时间的运动特征，该方法使用了自然场景统计特性（natural scene statistics，NSS）[60]对其进行提取。最后根据人的主观视觉感知，将提取出来的时空特征一起输入并训练一个回归模型来得到最终的视频质量分数。该方法的性能优于当时的其他无参考视频质量评价方法。但是该方法在提取时间特征时需要进行手工提取，并且该方法要先分别对时间和空间特征进行提取，然后再将其进行融合，这使得在提取特征时可能会丢失许多的关键信息。之后通过使用迁移学习提出了DeepBVQA方法[61]，该方法使用卷积神经网络提取空间特征，但对于时间特征依然需要进行手工提取。Lomotin等人[62]通过研究无参考图像质量评价方法，提出了一个复杂的框架来评估图像和视频的质量。该框架通过图像质量评估来应用到短视频上，以实现短视频快速稳定的逐帧评估。评分过程由几个并行的收集步骤和最后的分数聚合步骤组成。大多数评分模型基于深度卷积神经网络（CNN）。通过添加或删除这些步骤，可以灵活地扩展或减少框架。

由于大多数的VQA模型在人为制造的失真视频上可以实现有效的评估，但在自然的视频上的评估往往并不是那么理想，训练过程中存在一定的过拟合问题。

基于此问题，Li等人[63]提出了基于2D CNN的VSFA方法，该方法将人类视觉系统的两个突出效应整合到深度神经网络中，即内容依赖效应和时间记忆效应。对于内容依赖效应，其从预先训练的图像分类神经网络中提取特征，以获得其固有的内容感知属性。对于时间记忆效应，比如长期依赖关系，尤其是时间滞后，其通过门控循环单元和受主观启发的时间池层集成到网络中。

Varga等人[64]引入了长短时记忆网络（long short term memory，LSTM），并将其与卷积神经网络相结合，提出了一个深层架构，架构图如图10所示。该方法通过将待测视频分为一帧一帧的图像输入预训练好的CNN模型，以此来提取视频图像中的深度特征，预训练的CNN运行所有连续的视频帧以创建$d\times N$序列数据，其中d代表视频序列的长度，N是帧级深度特征向量的长度。然后将生成的序列作为LSTM网络的输入，以此训练LSTM网络来预测质量分数。LSTM网络具有一定的记忆功能，能够对之

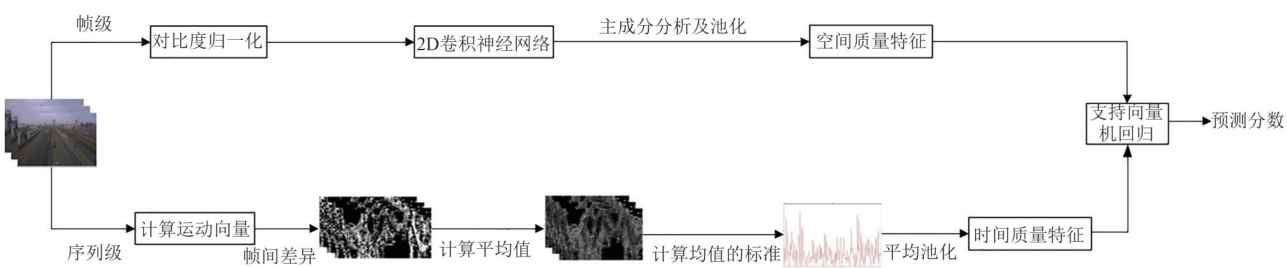

图9　CNN-MR框架[59]

Fig. 9　CNN-MR architecture[59]

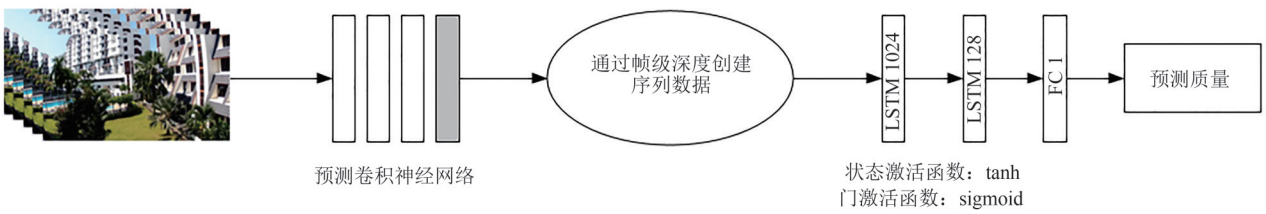

图 10 CNN 和 LSTM 算法框架[64]

Fig. 10 CNN and LSTM algorithm framework[64]

前视频质量的预测有一定的保留能力。算法在 KoNViD1k[65]视频数据库上进行测试，展现出了不错的性能。

3.4 面向 UGC 的视频质量评价方法

最近，社交媒体出现了巨大的增长，大量用户生成的视频内容（UGC）在各大媒体平台上分享。由于功能强大并且价格合理的移动设备和云计算技术的进步，再加上视频流的发展，使大多数消费者能够轻松地在全球范围内即时创建、分享和查看 UGC 图片/视频。事实上，UGC 的盛行已经开始将视频质量研究的重点从传统的合成失真数据库转移到更新的、更大规模的真实 UGC 数据集，这些数据集通常被用来解决 UGC-VQA 的问题。UGC-VQA 研究通常有着以下特点：

1）所有源内容都是用户生成的，因此存在未知且高度多样化的损伤；2）它们仅适用于测试和比较无参考模型，因为参考视频不可用；3）失真的类型是多种多样的，也有可能是各种情况混合的，包括但不限于捕获损伤、编辑和处理伪像、压缩、转码和传输失真。此外，与传统的 VQA 数据集和算法不同，压缩伪影不一定是影响视频质量的主要因素。这些不可预测的感知退化使得 UGC 视频的感知质量预测非常具有挑战性。

3.4.1 UGC 相关数据集介绍

UGC 视频的感知质量是一个宽泛的概念。除了压缩伪像，视频制作过程中引入的失真（如镜头模糊和相机抖动）也会影响观众的观看体验。最近发布了一些大规模的 UGC 图像数据集[8,66-67]，但是 UGC 的视频数据集仍然十分有限。

第一个包含真实失真的 UGC 视频数据集为 CVD2014[68]，该数据集中的视频并且使用了 78 个不同的视频捕获设备录制，之后还提出了相似的 LIVE-Qualcomm Mobile In-Capture Database[69]. 然而，这两个数据库具有比较大的局限性，它们仅对少量不太多样化的独特内容进行建模（相机）捕获失真。LIVE 数据集[70-72]为传统的公共视频质量数据集，主要用于分析原始数据的压缩失真，同时包含有限的 UGC 特征。YouTube-8M[73]和 AVA[74]这两个数据集是用来进行识别的，它们不提供原始视频数据和相应的 MOS 值，因此它们对质量评估的研究作用不大。在过去的几年里发布了一系列的大规模 UGC 质量数据集[75-77]，这些数据集会提供原始视频和 MOS。在这些数据集中，YouTube 的 UGC 数据集（YT-UGC）[77]是最具代表性的数据集之一。该数据集中的内容标签和 MOS 得分如图 11 和图 12 所示。该数据集从 150 万个 YouTube 视频中采样了 1500 个视频，具有共享权限。然而，虽然 YT-UGC 的一个主要目标是促进对视频压缩和质量评估实际应用的研究，但当前数据集不包含任何视频压缩版本和相应的差分 MOS（DMOS）。此外，提供的粗内容类别中的视频显示出高质量多样性，很难在内容和质量之间建立联系。

3.4.2 UGC 视频质量评价

Tu 等人提出了一种新的基于融合的 BVQA 算法[78]，该算法称之为 VIDeo quality EVALuator（VIDEVAL），它在现有的高效 BVQA 模型之上使用了特征集成和选择程序。通过在统一且可重复的评估框架内对当前比较好的视频质量模型进行系统评估，证明了用失真感知统计的视频特征和明确定义的视觉障碍特征能够以非常合理的计算成本提供最可靠的性能。为了量化使用的数据库在每个定义的特征空间上的覆盖率和均匀性，从而计算了覆盖率的相对范围和均匀性[79]，其中相对范围由下式给出：

$$R_i^k = \frac{\max\left(C_i^k\right) - \min\left(C_i^k\right)}{\max\limits_k\left(C_i^k\right)} \tag{15}$$

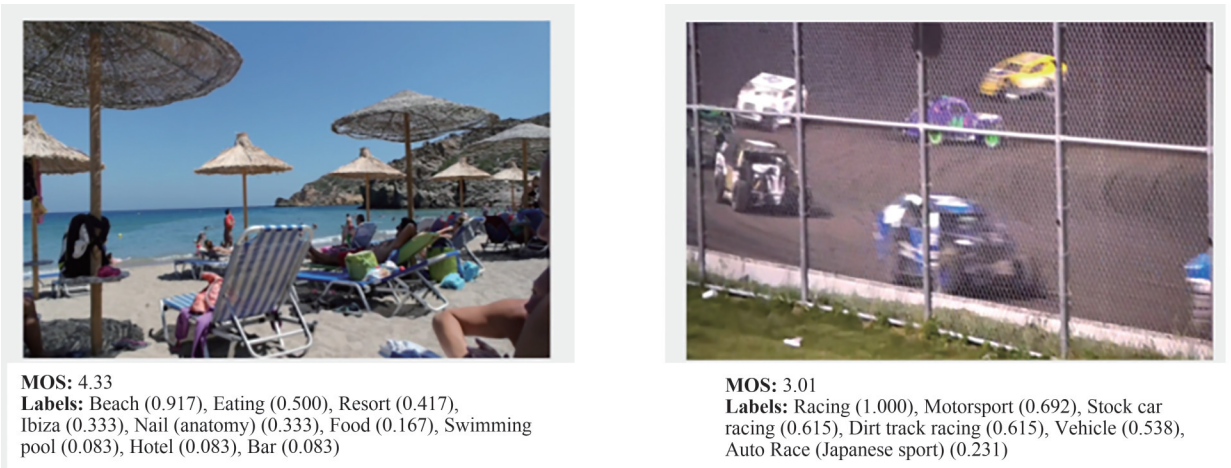

图 11　YT-UGC 内容标签[77]

Fig. 11　YT-UGC content label[77]

Video_id	Chunk_id	Category	Resolution	MOS
Vlog_1080P-18da	full	Vlog	1080P	4.38
Vlog_1080P-18da	chunk00	Vlog	1080P	3.93
Vlog_1080P-18da	chunk05	Vlog	1080P	4.23
Vlog_1080P-18da	chunk10	Vlog	1080P	4.62

图 12　YT-UGC MOS 得分[77]

Fig. 12　YT-UGC MOS score[77]

其中 C_i^k 表示给定特征维度 i 的数据库 k 的特征分布，并且 $\max\left(C_i^k\right)$ 指定所有数据库中该给定维度的最大值。覆盖均匀度衡量视频在每个特征维度中的均匀分布。我们将其计算为每个索引为 k 的数据库的所有源的 C_i^k 的 B-bin 直方图的熵：

$$U_i^k = -\sum_{b=1}^{B} p_b \log_B p_b \tag{16}$$

其中 p_b 是数据库 k 的特征 i 处 bin b 中源的标准化数量。统一性越高，数据库就越统一。

之后，该团队在针对处理图像视频质量的问题上使用回归还是分类进行了讨论，并提出了两种新方法——二元分类和序数分类[80]，这两种方法可以在较粗略的级别上评估和比较无参考质量模型的替代方法，而且在感知优化的 UGC 转码或媒体处理平台上的预处理方面传达了更实际的意义。紧接着为了加速视频质量的预测速度，该团队为 UGC 内容引入了一种有效且高效的视频质量模型，并将其称为快速准确的视频质量评估器（RAPIQUE）[81]，同时展示了该评估器的性能与最先进的模型相当，但速度要快几个数量级运行。

Wang 等人[82]创建了一个大规模数据集来全面研究通用 UGC 视频质量的特征。同时还提出了一

个基于DNN的框架，用来详细分析内容、技术质量和压缩级别在感知质量中的重要性。并且该模型能够提供质量分数以及人性化的质量指标，以弥合低级视频信号与人类感知质量之间的差距。

3.5 视频质量评价方法评估

3.5.1 基于LIVE-VQA数据集的分析

本节选取了上文中提到过的6种视频质量评价方法，并将其作用于LIVE-VQA数据集上，通过实验结果对这些方法进行比较和分析。如表3所示，失真类型分为Wireless、IP、MPEG2、H.264和ALL。这些失真类型代表了在各类压缩失真视频上的实验效果。

由实验结果可知，加入了运动信息的全参考方法MOVIE和ST-MAD在整个数据集上的SROCC和PLCC都达到了0.78以上，相较之前的传统方法来说性能得到了提升，其中ST-MAD尤其擅长处理MPEG-2压缩失真视频，其实验效果要优于其他方法，这就证实了加入运动信息可以使质量评价性能得到提升，但是仍然还存在较大的提升空间。而Park等人提出的在MOVIE方法上使用自适应池化去整合每帧的质量，取得了很不错的效果，该方法的实验结果相较于MOVIE方法得到了全面的提升，并且有三项指标在比较的实验方法中达到了最好的效果。

基于IQA方法改良的VBLIINDS算法在PLCC指标上得到了十分可观的结果，但是在SROCC上却没有展现出良好的性能。而且其对应的图像方法在LIVE-IQA数据集上都能达到0.91的效果，而作为视频方法出现了大幅度的下降，这从某种程度上也说明了在视频质量评价当中时间维度的信息对于整个视频质量评价有着很重要的影响。

而作为无参考方法的VIIDEO在各项指标上都无法与其他方法进行比较，这是由于无参考方法中不依赖原始参考视频，这也就导致了时间维的变化对实验结果造成了无法预估的误差，这也是无参考视频质量评价方法的难点所在。Varga等人的方法通过使用LSTM来预测时间维度信息的变化，使误差得到了减小，相较于VIIDEO方法提高了实验效果。

3.5.2 基于KoNViD-1k和LIVE-Qualcomm的分析

如表4所示，KoNViD-1k和LIVE-Qualcomm数据集是自然失真数据集，在该数据集上的进行质量评价要比以往的数据集更加具有挑战性。我们选取了四种算法进行比较，分别是传统算法VBLIINDS和VIIDEO，以及使用了深度学习的算法VSFA[63]和Varga等人[64]提出的方法。

由实验结果可以看出，传统的方法在面对自然失真的数据集所呈现出的实验效果就非常差了，其性能相比较使用深度学习的方法有着很大的差距。也就意味着传统方法中，人为设计的手工特征很难应对视频中的自然失真，而基于深度学习的方法能够很好的对自然失真进行预测，使其呈现出可观的效果。

综上，VQA的发展历程当中，有从IQA中获得启发进一步推进成为VQA的方法，有从FR-VQA演变成NR-VQA的方法，有传统方法的研究，也有基于深度学习方法的问世。数据集的稀缺使NR-VQA的研究成为一个必要的方向，深度学习的出现也给VQA提供了更多的思路。

表3 LIVE-VQA数据集上的各种VQA方法的性能

Tab. 3 Performance of various methods on LIVE-VQA Dataset

算法	SROCC					PLCC				
失真类型	Wireless	IP	MPEG2	H.264	ALL	Wireless	IP	MPEG2	H.264	ALL
ST-MAD(Vu等[45],2011)	0.81	0.77	**0.90**	0.85	0.82	0.81	0.79	0.91	0.84	0.83
MOVIE(Seshadrinathan等[39],2010)	0.80	0.72	0.77	0.77	0.79	0.84	0.76	0.79	0.76	0.81
MOVIE(VQPooling)(Park等[40],2013)	**0.90**	**0.81**	0.83	0.85	**0.84**	0.85	0.80	0.84	0.85	0.86
VBLIINDS(Saad等[50],2014)	0.82	0.78	0.84	**0.87**	0.76	**0.95**	**0.95**	**0.92**	**0.89**	**0.88**
VIIDEO(Mittal等[49],2016)	0.53	0.61	0.67	0.56	0.62	0.63	0.74	0.69	0.63	0.65
(Varga等[64],2019)	/	/	/	/	0.70	/	/	/	/	0.69

表4　KoNViD-1k和LIVE-Qualcomm数据集上的一些经典VQA方法性能

Tab. 4　Performance of some classical VQA methods on KoNViD-1k and LIVE-Qualcomm Datasets

算法	SROCC			PLCC		
失真类型	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
VBLIINDS（Saad等[50]，2014）	0.60	0.59	0.49	0.56	0.69	9.01
VIIDEO（Mittal等[49]，2016）	0.30	0.30	0.61	0.13	0.10	12.31
VSFA（Li等[63]，2019a）	0.76	0.74	**0.47**	**0.74**	**0.73**	**8.86**
（Varga等[64]，2019）	**0.85**	**0.87**	/	/	/	/

4　音频质量评价

随着各种媒体技术的不断发展，语音通话、观看视频以及欣赏音乐等活动都离不开音频，并且音频作为传递信息的一个重要手段，在未来的应用场景会越来越多，好的音频体验也成为了人们的追求之一。因此，研究有效的音频质量评价方法能够对音频的质量起到一定的促进作用。

与其他多媒体信息的质量评价方法相似，音频质量评价可以根据评价的方式分为两大类：主观音频质量评价和客观音频质量评价。主观评价方法就是将待测音频播放给听声人员，然后让听声人员根据自己所听到的音频，然后再根据某种预先规定的标准或者尺度对音频的质量进行等级划分。主观方法反映的更多是听声人员对该音频的一种主观印象，这种评价一般都更贴近于人们对音频质量的真实感受。客观评价方法大多是收集音频的信息，然后制定一系列的参数标准，再根据收集到的音频信息的各个指定的参数去判断该音频的失真程度，从而来对音频的质量进行客观的评估。本文有关音频质量评估方法的内容框架如图13所示。

4.1　主观音频质量评价

1997年，ITU提出了BS.1116-1标准，该方法也叫《多声道音频系统中小损伤主观评价方法》，是音频质量评价领域的开山之作，之后很多方法都是根据该方法进行改进。其核心思想是将音频的级别分为优、良、中、差、劣五个等级，每个等级对应了一个MOS分数区间，然后让参与测试的人员对音频进行打分，分越高表明音频质量越好。最后综合所有的评分结果来得到最终的音频质量。之后在2003年时，ITU对BS.1116-1标准进行了改进，优化了测试的条件，对用来评价的参数进行了更加科学的调整，提出了更加完善的BS.1534-1标准作为主观评价方法，也称MUSHRA（Multiple Stimuli with Hidden Reference and Anchor）主观评价方法，该方法相比ITU-R BS.1116工作量较小，同时结果也更可靠。随着对BS.1116方法的不断改进，现在比较通用的受认可度比较高的主观方法为ITU-R BS.1284[83]评价标准。

4.2　客观音频质量评价

由于主观评价的结果容易受到外界干扰因素的影响，并且需要耗费大量的人力、物力和财力，成本过高，于是客观音频质量评价的研究变得十分有必要。

和图像视频的客观评价方法一样，音频的客观质量评价也可以根据对参考音频的依赖程度分为三大类型：全参考、部分参考和无参考。

传统的客观评价方法是将待测音频的一些提前制定好的参数类型进行提取出来，然后再将提取出来的参数和参考的音频参数进行对比，并设定相应的指标来判断该音频质量的优劣程度。常用的方法有峰值信噪比和总谐波失真等。但是，这些传统的方法没有考虑到人类的听觉特性，仅仅是将音频中一些参数提出进行测评，这导致了有些音频通过传统方法获得了较高的评价，但是人们听起来效果依然十分不理想，因为某些选定的参数对于人类的听觉来说并不那么容易感知到，这使得客观评价的结果与主观评价的结果有着较大的出入。

4.2.1　有参考音频质量评价

ITU-R组织在2001年提出了著名的BS.1387标准（即PEAQ标准，Perceptual Evalation of Audio Quality），该方法将心理声学模型与感知模型相结合，这是目前的音频质量客观评价国际标准[84]。PEAQ核心算法结构如图14所示。

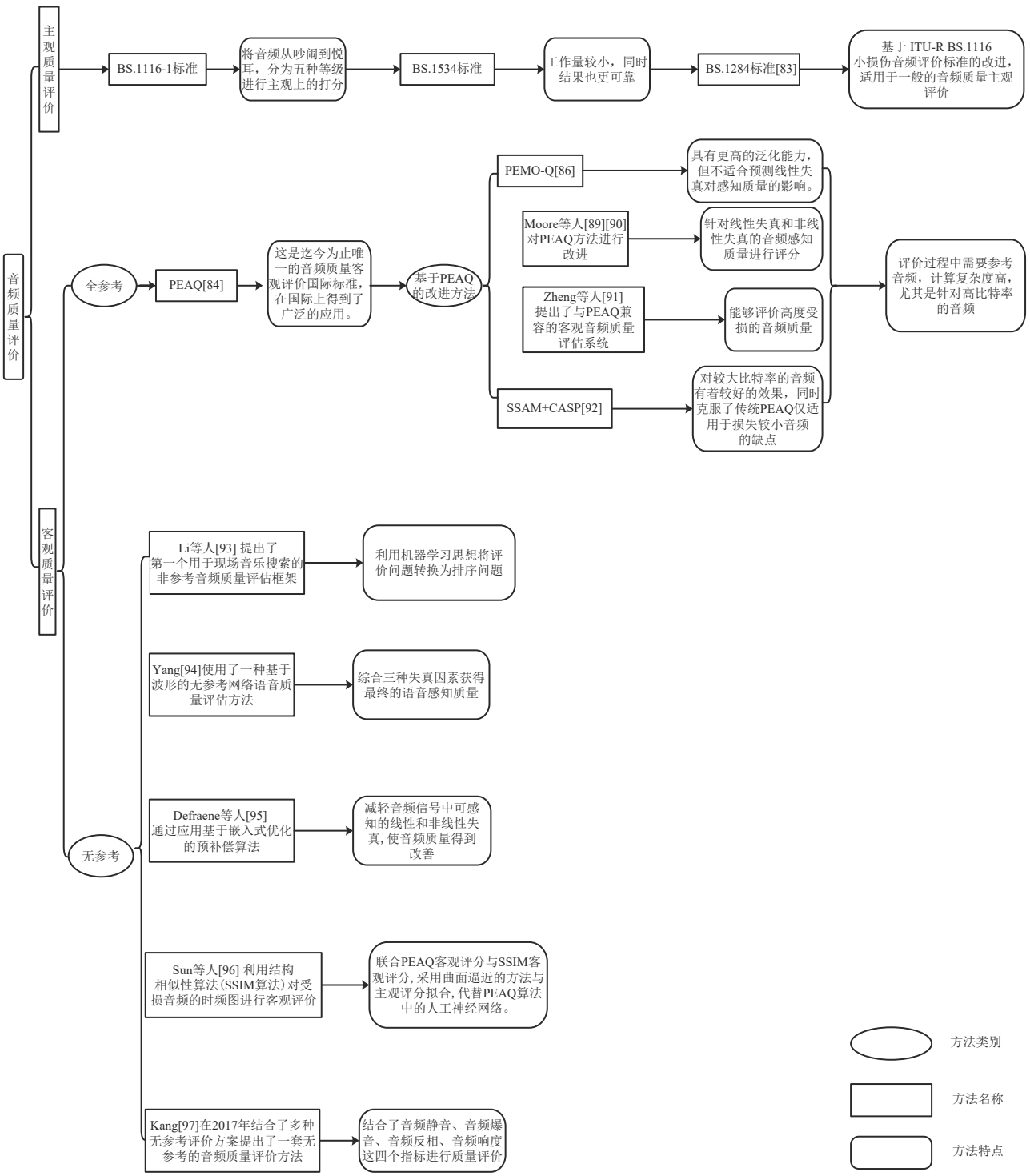

图 13 音频质量评价方法框架图

Fig. 13 Frame of audio quality evaluation methods

PEAQ 方法大概分为两个步骤，第一步是将参考音频和待测音频的信号输入心理声学模型，然后将该模型的输出分为两部分，将其中一部分输入感知模型进行特征的提取和综合，然后再与另一部分的输出一起输入神经网络，得到多维 MOV 值之后，利用神经网络训练测试求失真系数 DI。最后把失真系数 DI 值转化为客观评价分数 ODG。公式如下：

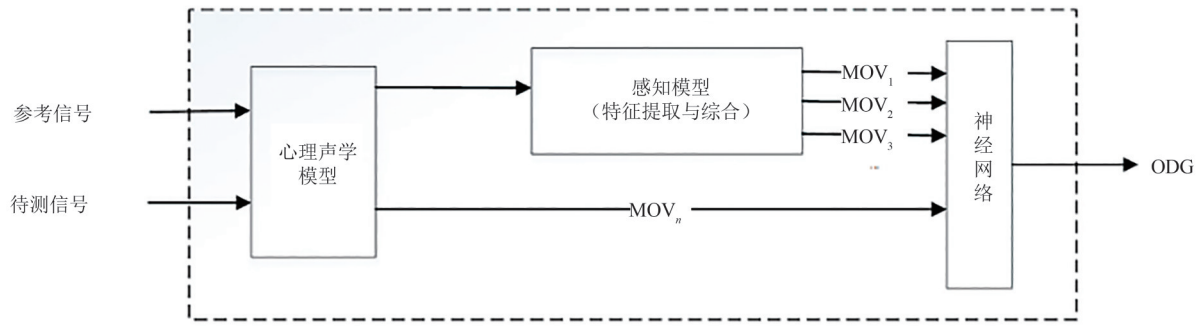

图 14　PEAQ 核心算法框架[84]

Fig. 14　PEAQ core algorithm framework[84]

$$\mathrm{ODG} = b_{\min} + \left(b_{\max} - b_{\min}\right) \times \mathrm{sig}(\mathrm{DI}) \quad (17)$$

其中，$b_{\min}$ 和 $b_{\max}$ 为设定好的权重参数，DI 为失真系数，sig(·)为阈值函数。ODG∈[-4,0]表示音频质量评价分数，其中 0 表示最好，-4 表示最差。

在 PEAQ 方法提出来之后，许多研究人员对其进行了改良，Cave 等人[85]将时域掩蔽模型和 SPL 计算方法引入 PEAQ 方法，以此来弥补该方法的不足。Huber[86]等提出了 PEMO-Q 方法，该方法基于经过心理声学验证的听觉模型，对于不同类型的音频信号和信号衰减，可以预测非常细微的以及更严重的质量降级。预测的音频质量与所应用的测试材料的主观质量评级显示出良好的相关性。PEMO-Q 具有更高的泛化能力，但是它不适合预测线性失真对感知质量的影响。

Hines 等人[87]评估客观质量指标是否可以预测音质通过将客观预测与听众测试的结果进行比较，以低比特率编码的音乐。对三个客观指标进行了基准测试：PEAQ、POLQA 和 VISQOLAudio。结果表明，为语音质量评估设计的客观指标在低比特率音频编解码器的质量评估方面具有强大的潜力。

Barbedo 等人[88]引入新的认知模型提高了 PEAQ 的准确性。Moore 等人[89-90]对 PEAQ 方法进行改进，提出针对线性失真和非线性失真的音频感知质量进行评分的方法，并通过实验证明具有良好的效果。Zheng 等人[91]提出了一种与 PEAQ（音频质量的感知评估）兼容的改进的客观音频质量评估系统。基于计算听觉模型，使用了一种新的心理声学模型来评估严重受损音频的质量。并使用线性 MOA 和 Minmax MOA 进行计算，对感知模型做出估计，该方法能够适用于质量高度受损的音频。Zhu 等人[92]提出了一种改进的结构相似性的音频质量评价方法，即调制结构相似性分析（Structural Similarity Analysis of Modulation, SSAM），该方法使得 PEAQ 方法不仅能适用于损失小的音频，也能够对高度受损的音频进行评价。

然而，不管对 PEAQ 如何进行改进，全参考的方式仍然有着无法避免的问题，那就是需要原始的参考音频，然而在实际问题中，很难获得原始的参考音频，因此研究不需要参考音频的无参考音频质量评价方法是目前研究的大方向。

4.2.2　无参考音频质量评价

在图像、视频等评价领域的无参考评价方法都取得了一定的发展，而在音频领域，直到 2013 年，Li 等人[93]提出了使用机器学习的方法将音频质量评价转换为对音频质量进行排序，从而间接地得到音频的大概质量，该方法主要用于现场音乐的评价。到了 2015 年，Yang 等人[94]使用波形来进行无参考网络语音质量评估，该方法对需要评价的音频进行解码得到信号波形，综合多种失真因素来得到语音感知质量。Defraene 等人[95]在 2016 年通过应用基于嵌入式优化的预补偿算法，以减轻音频信号中可感知的线性和非线性失真，使音频质量得到改善。同时发现主观和客观 PEAQ 音频质量分数之间的正相关，验证了使用 PEAQ 预测线性和非线性失真对感知音频质量的影响的有效性。Sun 等人[96]利用 SSIM 算法得到待评价音频的质量分数并结合 PEAQ 方法得到的质量分数对主观评价的分数进行拟合。Kang[97]在 2017 年结合了多种无参考评价方案提出了一套无参考的音频质量评价方法，其核心

是通过提取音频关键指标信息来作为评价依据。结合了音频静音、音频爆音、音频反相、音频响度这四个指标进行质量评价，同时佐以音频采样率、声道数等参数进行辅助测量，在不依托原始质量音频作为参考的情况下，比较好的解决了音频质量参数的测量问题，同时做到了边测量边输出，可以实时的观察当前音频的质量状况。到了2020年，Min等人[98]提供了一种基于自然音频统计特性的无参考音频质量评价方法，通过将相关的自然图像统计特性推广至自然音频统计，从而实现基于自然音频统计的无参考音频质量评价。

传统的音频质量评价方法已经基本成熟，当前仅针对音频的质量评价也基本能满足人们的日常需求，但是在现实生活中音频往往伴随着各类其他形式的信息出现，尤其是音频与视频之间的联系。例如，视频中往往伴随着音频，当视频或音频的失真可能导致音画不同步的现象也会极大的影响用户的体验。因此在之后的研究当中，如何将音频信息与视频信息相结合，提出视听联合的质量评价方法也是非常有必要的。

5 文本质量评价

文本质量评价方法也可以分为人工评估和自动评估的方式，人工评价是指人工阅读和查看内容的过程，并在此基础上人工编码进行分析，最终做出判断。人工评价的过程会更加的灵活，但是人工评价往往过程繁琐，耗费时间较长并且容易受到实验者的个人主观因素的影响。自动评估又可以分为无训练的和基于神经网络的方法。本文所介绍的文本质量评价内容框架如图15所示。

5.1 自动文本评分(AES)

自动文本评分(AES)是指一套统计和自然语言处理技术，用于在评分等级上自动给文本评分。一个典型的AES系统将一篇关于特定主题的文章作为输入。然后，系统会根据文章的内容、语法、组织和上面讨论的其他因素，给文章分配一个反映其质量的数字分数。

5.1.1 基于手工选择特征

由于多种因素影响文本的质量，自动测试系统通常利用大量的文本特征，这些特征对应于文本的不同属性，如语法、词汇、风格、主题相关性以及语篇连贯和衔接。除了词汇和词性标注之外，语言上更深层次的特征，如句法结构的类型、语法关系和句子复杂性的度量，也是构成自动测试系统内部标记标准的一些属性。文本的最终表示通常由特征向量组成，这些特征被手动选择和调整以预测评分等级上的分数。简单地说基于手工选择特征的AES方法通常都是通过人工设计提取相关文本特征，再使用分类回归或者排序的方法对文本内容进行评测。

90年代即初代所提出的自动文本评分技术主要是从文本中提取多个文本特征，通过多元回归的方法以人工评分作参考进行分析。2006年Attali和Burstein提出的E-rater[99]被教育考试服务用于自动论文评分，具有小而有意义的特征集和简单直观的组合特征的方式。这些特征允许用户对评分过程进行更大程度的判断控制，例如确定由系统测量的不同书写尺寸的相对重要性。它还允许评分更加标准化，特别是允许为程序或评估的所有提示开发单一评分模型。这些方面有助于e-rater的有效性，允许更好地理解和控制自动评分。写作文本的质量分析主要包括语义、词汇、语言准确性、结构质量等多个方面，但e-rater衡量标准显然没有涵盖写作文本质量的所有重要方面，也没有完美地衡量它所涵盖的维度。对文本结构的分析只拘泥于文本的表层特征，关注文本语句的多样性。

2014年Somasundaran等人[100]提出将词汇链特征与话语要素的互动特征相结合，从语篇连贯看文本的质量，可提高系统语篇特征：语法特征、词汇用法特征、机制错误特征等，这里使用词汇链特征来训练一个语篇连贯分类器，词汇衔接是有助于词汇意义连续性的相关词汇链的结果。这些序列的特点是单词之间的关系，以及它们在给定范围内的距离和密度。词汇链不受句子边缘的限制，它们可以将相邻的单词连接起来，也可以遍及到整个文本的范围。

2015年McNamara等人[101]使用了一种层次分类方法进行自动评分，与以往依赖回归模型的研究不同，该方法使用类似于递阶分类增量算法的分层算法来计算文本分数，利用语义和修辞特征等，分析中包含的特征是使用自动化工具Coh-Metrix、写作评估工具(WAT)以及语言查询和字数统计

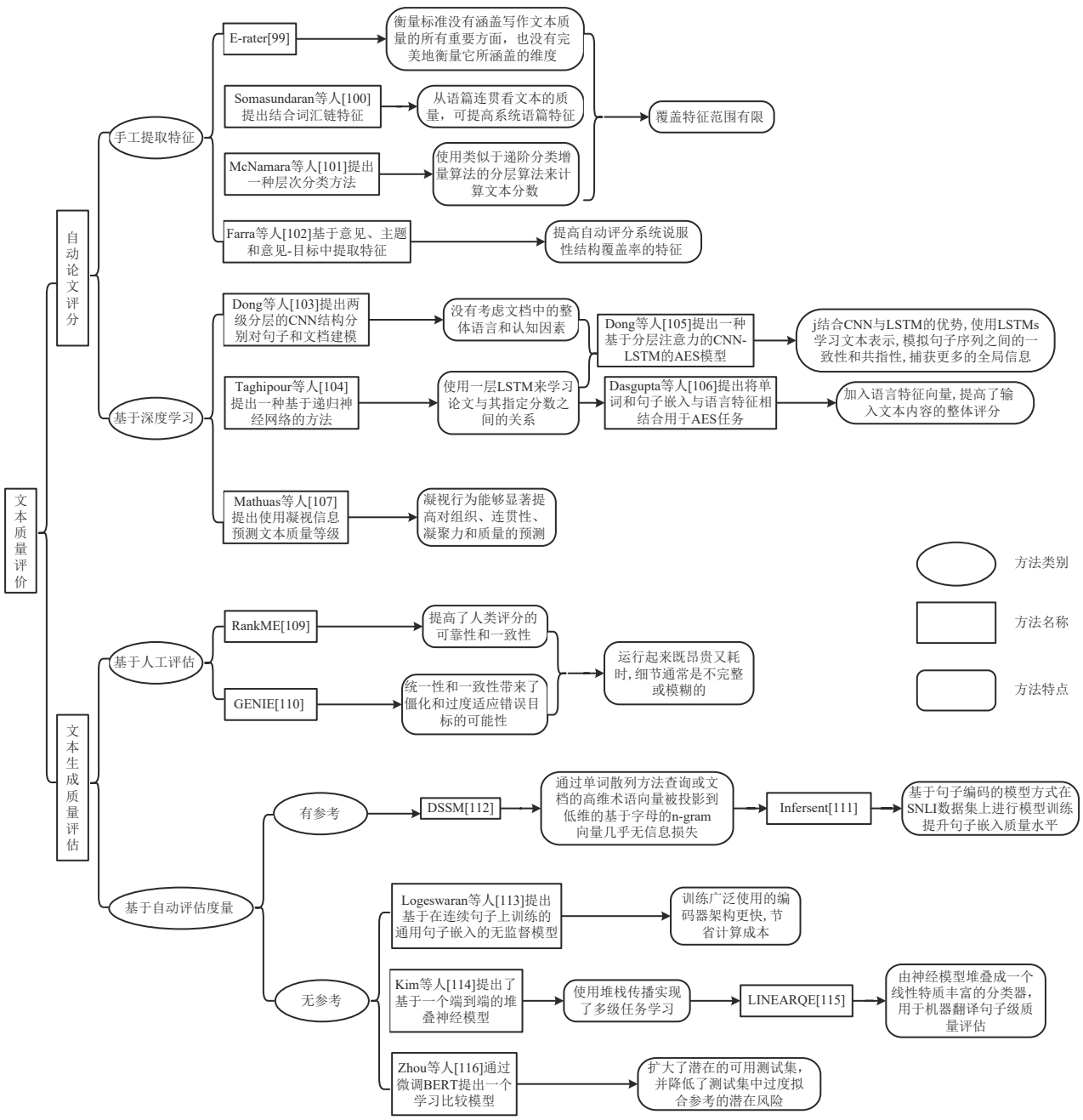

图 15　文本质量评价方法框架图

Fig. 15　Framework diagram of text quality evaluation method

(LIWC)计算得出的。

以上自动评分系统具有强大的预测能力,但它们覆盖范围有限,将说服力因素的知识纳入文本评分模型可以增加与评分结构和写作文本任务直接相关的特征。2015年Farra等人[102]开发了基于从意见、主题和意见-目标中提取的特征并结合了逻辑推理和线性回归的变体的模型系统来评估文本。为此构建了三个独立系统:1)意见—系统仅使用基于意见表达的特征,并测试表达意见是否影响文章分数。2)主题—系统仅使用基于主题表达的特征,并且测试唤起与提示相关联的相关主题是否影响文章分数。3)意见-目标—系统使用基于意见及其目标的组合的特征,目的是测量意见的相关性和一致性,这个系统测试了根据观点和目标的相互作用来

预测文章分数的效果。

但是传统的文章评分工作集中在自动手工制作的特征上，这些特征很昂贵，且很稀少。而且手工选择的特征具有针对性，每一种方法所涵盖的文本特征有限，系统的泛化性不强。

5.1.2 基于深度学习

神经模型提供了一种自动学习句法和语义特征的方法，不用依赖于特征的手工工程，以端到端的方式可更好改善离散特征。大多数现有的基于深度学习的作品使用卷积神经网络（CNN）对输入文本进行建模。2016年Dong等人[103]为AES任务建立了一个高层次的CNN模型，CNN模型将短文评分作为回归任务，采用两层CNN模型，有一个较低层次的表示句子结构和一个基于句子表示的较高层的表示文章结构。

相比较于CNN，LSTM在模拟长期历史方面很强大，2016年Taghipour等人[104]提出一种基于递归神经网络的方法在单词序列上使用一层LSTM来学习论文与其指定分数之间的关系，而无需任何特征工程。

2017年Dong等人[105]继续提出了一种基于分层注意力的CNN-LSTM自动作文评分模型，结合CNN和LSTM两种网络优势通过构建一个层次化的句子-文档模型来表示短文，使用注意机制来自动决定单词和句子的相对权重。神经模型使用LSTMs学习文本表示，这可以模拟句子序列之间的一致性和共指性（与CNNs相比，捕获更多的全局信息）。此外，注意力集中在单词和句子上，旨在捕捉更多有助于论文最终质量的相关单词和句子。

深层多层神经网络可以从数据中自动提取有用的特征，下层学习基本的特征检测器，上层学习更高级的抽象特征。尽管基于神经网络的方法比传统的统计方法表现更好，然而深度神经网络模型不允许识别和提取网络识别为有区别的文本属性，特别是没考虑文档中的整体语言和认知因素。2018年Dasgupta等人[106]提出了一种定性增强的深卷积递归神经网络计算文本质量的方法，提出的系统考虑了单词和句子层面的嵌入。该论文使用的基于语言的卷积递归神经网络结构如图16所示，这里不仅依赖于文本的预先训练的单词或句子表示，而且考虑了质量增强的特征，例如，词汇多样性、信息性、衔接性、良好形式等，利用层次卷积递归神经网络框架增强了不同的复杂语言，与文本相关的认知和心理特征。语言特征向量的加入确实提高了输入文章的整体评分。

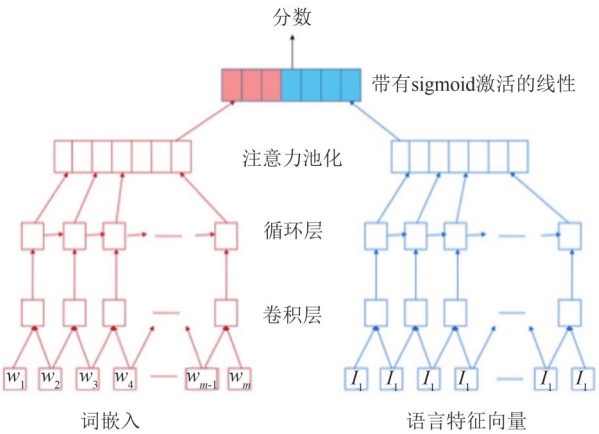

图16 用于AES的定性增强卷积递归神经网络概述[106]

Fig. 16 Overview of the qualitatively enhanced convolutional recurrent neural network for AES[106]

首先构建了一个预先训练好的句子向量，来自每个输入文章的句子向量被附加上由该特定句子的语言特征形成的向量。将每个生成的单词$X_1, X_2, \cdots X_h$嵌入馈送到卷积层进行连接形成长度为h的向量$\boldsymbol{X}$，对输出向量进行如下卷积操作：

$$\mathrm{conv}(\boldsymbol{X}) = \boldsymbol{w}(\boldsymbol{X}) + \boldsymbol{b} \tag{18}$$

其中$\boldsymbol{w}$和$\boldsymbol{b}$是网络学习的权重。

使用双向LSTMs网络连接以便可以检查未来和过去的序列上下文（即前面和后面的元素）。从双向LSTMs层中获得中间隐藏层之后，在激活层中又在句子表示上使用了注意力集中层。集中注意力有助于获得句子对文本最终质量的贡献权重。句子的注意力集中表现为：

$$\boldsymbol{a}_i = \tanh\left(\boldsymbol{W}_a \cdot \boldsymbol{h}_i + \boldsymbol{b}_a\right) \tag{19}$$

$$\boldsymbol{\alpha}_i = \frac{\mathrm{e}^{\boldsymbol{w}_\alpha \cdot \boldsymbol{a}_i}}{\sum \mathrm{e}^{\boldsymbol{w}_\alpha \cdot \boldsymbol{a}_i}} \tag{20}$$

$$\boldsymbol{O} = \sum\left(\boldsymbol{\alpha}_i \cdot \boldsymbol{h}_i\right) \tag{21}$$

其中$\boldsymbol{W}_a$和$\boldsymbol{w}_\alpha$分别为权重矩阵和向量，$\boldsymbol{b}_a$是偏向向量，$\boldsymbol{a}_i$为第i句的注意向量，$\boldsymbol{\alpha}_i$为第i句的注意权重。$\boldsymbol{O}$是最终的文本表示，它是所有句子向量的加权和。

线性图层对输入向量执行线性变换，将其映射为连续的标量值：

$$s(\boldsymbol{X}) = \mathrm{sigmoid}(\boldsymbol{w}.\boldsymbol{X} + \boldsymbol{b}) \tag{22}$$

其中$\boldsymbol{X}$为输入向量，$\boldsymbol{w}$为权重向量，$\boldsymbol{b}$为偏差值。

2018年Mathias等人[107]证明注视行为有助于有效预测文本质量的等级，使用从读者的注视行为中获得认知信息，通过将注视特征添加到传统的文本特征中来预测分数。该方法基于三个属性来评估整体质量——组织性、连贯性和凝聚力，将文本质量建模为三个属性的函数——组织、连贯和衔接，使用李克特量表，范围从1到4，用于测量这些属性中的每一个；分数越高，就该属性而言，文本越好。这里使用这三个分数作为输入，在1到10的范围内对文本质量评级进行建模：

$$\mathrm{Quality}(T)=\mathrm{Org}(T)+\mathrm{Chr}(T)+\mathrm{Chs}(T)-2 \quad (23)$$

其中$\mathrm{Quality}(T)$是文本T的文本质量等级，$\mathrm{Org}(T)$，$\mathrm{Chr}(T)$和$\mathrm{Chs}(T)$分别对应于文本的组织、连贯和衔接分数。这里减去2，将分数从3~12分为1~10。

5.1.3　自动文本评分系统实验结果分析

本节的实验主要是在Automated Student Assessment Prize（ASAP）数据集上进行的。ASAP数据集由8种不同类型的提示组成，每一种提示都围绕一个主题展开。一些提示依赖于主题信息，另一些则是自由发挥。同时使用ASAP竞赛官方标准所使用的评估标准QWK。QWK统计或其变体被广泛用于衡量注释者或专家的评分者之间的一致性，QWK由kappa（一种衡量分类精度的指标）修改而来，kappa采用二次权重。

多篇方法实验结果如下表5。

表中前两行的模型是一个开源的论文评分系统EASE（增强型人工智能评分引擎），该系统是参加ASAP竞赛的最佳开源系统，EASE基于手工制作的语言特征和回归方法，包括支持向量回归（SVR）和贝叶斯线性岭回归（BLRR）。在EASE上与SVR模型相比，BLRR模型的性能有所提高。而从表中数据对比得出，基于深度学习的方法在ASAP数据集上的性能是比基于手工选择特征的方法EASE更优异，侧面也是体现了深度学习方法在自动文本评分系统应用前景广阔。

从表中可以得到LSTM-CNN-att模型比CNN-CNN-MoT模型在平均QWK指标上好3.0%，比LSTM-MoT模型好0.24%，这很大程度上在于LSTM-CNN-att模型结合了LSTM和CNN两个网络各自的优势，探索CNN的句子表示和LSTM的文本表示，这表明，句子-文档模式对长文章更有效。

对于提示8篇文章，其中Qe-C-LSTM模型平均QWK长度最大，传统的应用深度神经网络（如CNN、LSTM等）的方法无法识别评估文本质量所涉及的不同因素之间的相互联系。Qe-C-LSTM模型在某些情况下取得了显著的改进是因为该方法不仅依赖于文本的预先训练的单词或句子表示还考虑了定性增强的特征。

5.2　文本生成质量评估

许多自然语言处理任务旨在响应某些输入生成人类可读文本，文本生成是语言翻译、聊天机器人、问答、摘要和人们日常交互的其他几个应用程序的关键组成部分。这里讲NLG评估方法分为基于人工评估和基于自动度量评估两类。

5.2.1　基于人工评估

NLG的最终目标都是生成对人们有价值的文本。因此，人工评估通常被视为开发新自动指标的黄金标准。这里将使用人类判断评估生成的文本的方法分为内在评估和外在评估。外在的人类评估通常用于评估对话系统的性能（Deriu等人[108]），并对对话建模系统的发展产生了影响。相对来说，内在评价比外在评价更常见。内在评估要求人们

表5　自动文本评分系统在ASAP数据集上的性能

Tab. 5　Performance of the automatic text scoring system on ASAP datasets

Models/Prompts	1	2	3	4	5	6	7	8	AVG QWK
EASE(SVR)	0.781	0.621	0.630	0.749	0.782	0.771	0.727	0.534	0.699
EASE(BLRR)	0.761	0.606	0.621	0.742	0.784	0.775	0.730	0.617	0.705
LSTM-MoT（Taghipour等[104]，2016）	0.818	0.688	0.679	0.805	0.808	0.817	0.797	0.527	0.742
CNN-CNN-MoT (Dong等[103]，2016)	0.805	0.613	0.662	0.778	0.800	0.809	0.758	0.644	0.734
LSTM-CNN-att (Dong等[105]，2017)	0.822	0.682	0.672	0.814	0.803	0.811	0.801	0.705	0.764
Qe-C-LSTM（Dasgupta等[106]，2018）	0.799	0.631	0.712	0.711	0.801	0.831	0.815	0.695	**0.786**

评估生成的文本的质量，无论是整体的还是沿着某个特定的维度（例如，流畅性、连贯性、正确性等）通常是通过从模型中生成几个文本样本，并要求人类评估者对它们的质量进行评分来完成的。

2018 年 Novikova 等人[109]提出了一种新的基于秩的幅度估计方法（RankME），它结合了连续标度和相对评估的使用。这里将幅度估计添加到排名任务中，要求评估者指出他们选择的文本比备选文本好多少，给出了生成文本的绝对质量。

2018 年 Khashabi 等人[110]引入了 GENIE，这是一个用于评估生成性自然语言处理模型的新基准，它使得人类能够对模型进行大规模评分，为生成性自然语言处理任务发布了一个公共排行榜。可以将人群工作者反应转换成模型性能估计和置信区间的方法形式化。这里包括了标准化的人工评估，但是需要注意的是统一性和一致性带来了僵化和过度适应错误目标的可能性。

人工评估最能洞察模型在任务中的表现，但人工评估运行起来既昂贵又耗时，而且关于如何进行人工评估的细节通常是不完整或模糊的。

5.2.2 基于自动度量评估

未经训练的自动度量评估方法基于相同的输入数据将机器生成的文本与人类生成的文本（参考文本）进行比较，并使用不需要机器学习的指标，而只是基于字符串重叠、内容重叠、字符串距离或词汇多样性。但许多未经训练的评估指标假设生成的文本与真实文本有显著的单词（或 n-gram）重叠。通过基于机器学习的方法可以避免出现这一问题。构建机器学习模型可以（基于人类判断数据训练）来模仿人类判断，以测量输出的许多质量指标，例如事实正确性、流畅性、相似性等。

评价机器生成的文本质量时，根据有无参考文本又分为两种方向。有参考的是比较待评文本和参考文本的相似程度作评分，这种研究居多；不需要参考文本的评分又称为质量估计，这种被视为二分类问题。

在比较与参考文本间的相似度的方法其实是多样的，就比如基于句子语义相似度的评估的方法，Conneau 等人[111]扩展 Dssm 模型（Huang 等人[112]）提出一种有效的模型（Infersent），它使用基于 LSTM 的暹罗网络，对词序进行编码，通过基于句子编码的模型方式在 SNLI 数据集上进行模型训练，一定程度上提升了句子嵌入质量的水平。如图 17 所示，这种类型的典型架构使用共享句子编码器，可以输出前提和假设的表示，这里前提即为 $\boldsymbol{u}$，假设为 $\boldsymbol{v}$。当生成句子向量时，就会立刻匹配三种不同的方法来提取 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的关系：1. 两个表示（$\boldsymbol{u}$，$\boldsymbol{v}$）的串联；2. 元素积 $\boldsymbol{u}*\boldsymbol{v}$；和 3. 绝对元素差异 $|\boldsymbol{u}-\boldsymbol{v}|$。从前提和假设中获取信息的结果向量被输入到一个 3 类分类器中，该分类器由多个完全连接的层组成，最终形成一个软最大值层。

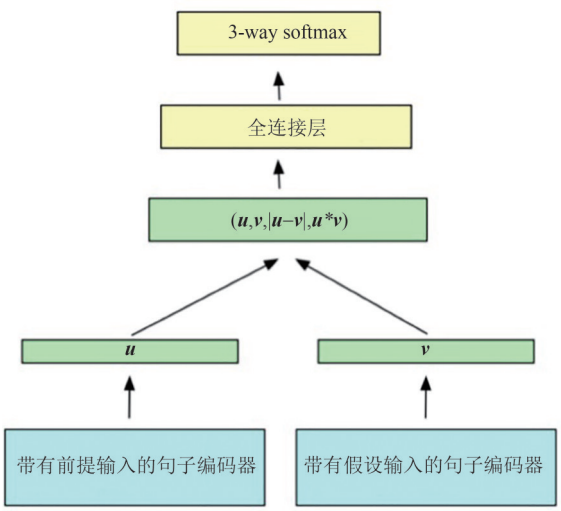

图 17 NLI 通用训练体制[111]

Fig. 17 NLI general training system[111]

在基于回归分类的方法上，Logeswaran 等人[113]从分布假设和学习句子表征的工作中得到启发，将预测句子中所出现的上下文的问题进行重新表述作为一种分类问题，提出基于在连续句子上训练的通用句子嵌入的无监督模型。上下文句子与其他对比句子由分类器在给定的一个句子及其出现的上下文语境时根据其向量来进行区分，这能够有效地学习不同类型的编码函数，可以通过使用句子表征作为下游 NLP 任务的特征表征来评估句子表征。Kim 等人[114]提出了基于一个端到端的堆叠神经模型，称为预测-估计器，其结构如图 18，它有两个阶段，包括神经单词预测模型和神经翻译质量评估模型，该模型采用多级任务学习，对句子、单词和短语级别评估翻译质量（QE）。Martins 等人[115]

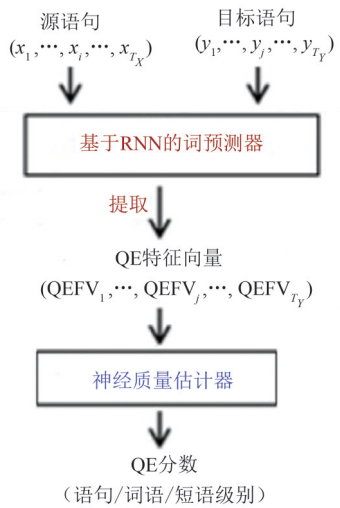

图 18　两级预测估计器结构[114]

Fig. 18　Two-stage predictive estimator structure[114]

提出的QE系统由一个神经模型(NEURALQE)堆叠成一个线性特征丰富的分类器(LINEARQE)，训练自动后期编辑APE系统(使用大量的人工“往返翻译”)，并调整预测句子级质量评分和单词级质量标签。

Bidirectional Encoder Representation from Transformers(BERT)是一个预训练的语言表征模型，已经被证明具有良好的自然语言理解能力，2020年，Zhou等人[116]提出了一个“学习比较”模型，以更好地评估基于成对比较的NLG模型生成的文本的质量。该模型能够以自我监督的方式通过微调从BERT传递自然语言理解知识，同时还能够通过人类偏好注释进一步微调。一旦经过训练，该模型能够在不需要黄金参考的情况下进行模型间比较，这极大地扩大了潜在的可用测试集，并降低了测试集中过度拟合参考的潜在风险，该方法与人类评价有更好的相关性。

6　结论

全媒体内容包涵多种信息模态，本文重点介绍了四种常见模态下质量评价方法的发展情况，如图像、声音、文本、视频，这其中每种信息模态均各自形成了一套、或者多套较为完善质量评价方法。对媒体的内容进行质量评价是提升用户信息交互体验感的核心环节，各种模态下的质量评价方法都经过了长期的理论研究和实践的积累。其中主观质量评价的方法以及规范流程已经较为完善，而客观评价方法在以主观感受为参考的约束下也在逐渐成熟。这些方法相互独立，自成体系，难以使用一套准则进行统一，不利于在多种信息模态融合的情况下进行主观一致性评价。无论是基于传统方法的还是基于深度学习方法，在不同应用场景下，都有其适用的特点。传统方法泛化能力受限但计算量小，深度学习结构复杂但评价精度高，在实际应用场景中我们可以根据需要，根据不同方法的特点进行选择。但是就目前来看深度学习方法的应用仍是质量评价领域的主流。如今面对现实生活中失真多样化的情况，单一算法还有所欠缺不能同时识别多种失真。因此，需要研究混合多种信息模态失真的多任务学习深度神经网络框架，在不同层面上对各个模态信息的失真予以度量。

在这个全媒体充斥着我们生活方方面面的时代，人们用来信息交互的手段越来越多样化，媒体用来呈现内容的方式越来越丰富，这些因素都使得人们传输的信息受到影响。由于传输的方式和手段不当，信息在传播过程当中产生失真、模糊和噪声等问题，使得用户无法有效的接收到信息甚至对信息产生迷惑和反感。但是，目前针对这些问题的全媒体内容质量评价方法仅仅停留在“流量思维”的阶段，并不能客观合理的对其质量进行评价，也无法有效的评价传播的效能。因此，发展能够以用户的体验为中心、以用户的需求为导向，并将多种模态下的质量评价方法进行凝炼和融合的全媒体质量评价方法是十分有必要的。

参考文献

[1]　谢皓，张健，倪江群. 数字图像操作取证综述[J]. 信号处理，2021，37(12)：2323-2337.
XIE Hao，ZHANG Jian，NI Jiangqun. A survey of digital image operation forensics [J]. Journal of Signal Processing，2021，37(12)：2323-2337.(in Chinese)

[2]　谭舜泉，黎思力，陈保营，等. 面向图像视频取证的机器学习综述[J]. 信号处理，2021，37(12)：2235-2250.
TAN Shunquan，LI Sili，CHEN Baoying，et al. A survey of deep learning in image and video forensics[J]. Journal of Signal Processing，2021，37(12)：2235-2250.(in Chinese)

[3]　周琳娜，杨震，储贝林，等. 多媒体认知安全综述[J].

信号处理, 2021, 37(12): 2440-2456.
ZHOU Linna, YANG Zhen, CHU Beilin, et al. Overview of multimedia cognition security[J]. Journal of Signal Processing, 2021, 37(12): 2440-2456.(in Chinese)

[4] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.

[5] LARSON E C, CHANDLER D M. Most apparent distortion: Full-reference image quality assessment and the role of strategy[J]. Journal of Electronic Imaging, 2010, 19(1):011006.

[6] PONOMARENKO N, LUKIN V, ZELENSKY A, et al. TID2008-A database for evaluation of full-reference visual quality assessment metrics[EB/OL]. 2009, 10(1): 30-45. https://videoclarity.com/PDF/mre2009tid.pdf.

[7] PONOMARENKO N, JIN Lina, IEREMEIEV O, et al. Image database TID2013: Peculiarities, results and perspectives[J]. Signal Processing: Image Communication, 2015, 30: 57-77.

[8] HOSU Vlad, LIN Hanhe, SZIRANYI Tamas, et al. KonIQ-10k: an ecologically valid database for deep learning of blind image quality assessment[J]. IEEE Transactions on Image Processing, 2020,29: 4041-4056.

[9] FANG Yuming, ZHU Hanwei, ZENG Yan, et al. Perceptual quality assessment of smartphone photography [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 3674-3683.

[10] AVCIBAS I, SANKUR B, SAYOOD K. Statistical evaluation of image quality measures[J]. Journal of Electronic Imaging, 2002, 11: 206-223.

[11] SHEIKH H R, BOVIK A C, VECIANA G. An information fidelity criterion for image quality assessment using natural scene statistics[J]. IEEE Transactions on Image Processing, 2005, 14(12): 2117-2128.

[12] LIU Anmin, LIN Weisi, NARWARIA M. Image quality assessment based on gradient similarity[J]. IEEE Transactions on Image Processing, 2012, 21(4): 1500-1512.

[13] LIU Yongjin, LUO Xi, XUAN Yuming, et al. Image retargeting quality assessment[J]. Computer Graphics Forum, 2011, 30(2): 583-592.

[14] LIN Weisi, NARWARIA M. Perceptual image quality assessment: Recent progress and trends[C]//Proc SPIE 7744, Visual Communications and Image Processing 2010, 2010, 7744: 33-41.

[15] CHANDLER D M, HEMAMI S S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images[J]. IEEE Transactions on Image Processing, 2007, 16(9): 2284-2298.

[16] SAAD M A, BOVIK A C, CHARRIER C. Blind image quality assessment: A natural scene statistics approach in the DCT domain[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3339-3352.

[17] 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.

[18] 董宏平, 刘利雄. 互信息域中的无参考图像质量评价[J]. 中国图象图形学报, 2014, 19(3): 484-492.
DONG Hongping, LIU Lixiong. No-reference image quality assessment in mutual information domain[J]. Journal of Image and Graphics, 2014, 19(3): 484-492.(in Chinese)

[19] XU Jingtao, LI Qiaohong, YE Peng, et al. Local feature aggregation for blind image quality assessment[C]//2015 Visual Communications and Image Processing (VCIP). Singapore. IEEE, 2015: 1-4.

[20] XU Jingtao, YE Peng, LI Qiaohong, et al. Blind image quality assessment based on high order statistics aggregation[J]. IEEE Transactions on Image Processing, 2016, 25(9): 4444-4457.

[21] Li Z, TANG J. Unsupervised feature selection via nonnegative spectral analysis and redundancy control [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5343-5355.

[22] Li Z, TANG J. Semi-supervised local feature selection for data classification[J]. Science China Information Sciences, 2021, 64(9): 1-12.

[23] LIU Yutao, GU Ke, ZHANG Yongbing, et al. Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(4): 929-943.

[24] MIN Xiongkuo, GU Ke, ZHAI Guangtao, et al. Blind quality assessment based on pseudo-reference image[J]. IEEE Transactions on Multimedia, 2018, 20(8): 2049-2062.

[25] ZHANG Yi, CHANDLER D M. Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation[J]. IEEE Transactions on Image Processing, 2018, 27(11): 5433-5448.

[26] KANG Le, YE Peng, LI Yi, et al. Convolutional neural networks for no-reference image quality assessment[C]//

2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. IEEE, 2014: 1733-1740.

[27] KANG Le, YE Peng, LI Yi, et al. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks[C]//2015 IEEE International Conference on Image Processing (ICIP). Quebec City, QC, Canada. IEEE, 2015: 2791-2795.

[28] PAN Da, SHI Ping, HOU Ming, et al. Blind predicting similar quality map for image quality assessment[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 6373-6382.

[29] YAN Bo, BARE B, TAN Weimin. Naturalness-aware deep no-reference image quality assessment[J]. IEEE Transactions on Multimedia, 2019, 21(10): 2603-2615.

[30] LIU Xialei, VAN DE WEIJER J, BAGDANOV A D. RankIQA: learning from rankings for no-reference image quality assessment[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 1040-1049.

[31] KIM J, LEE S. Fully deep blind image quality predictor[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(1): 206-220.

[32] MA Kede, LIU Xuelin, FANG Yuming, et al. Blind image quality assessment by learning from multiple annotators[C]//2019 IEEE International Conference on Image Processing (ICIP). Taipei, Taiwan, China. IEEE, 2019: 2344-2348.

[33] MA Kede, LIU Wentao, ZHANG Kai, et al. End-to-end blind image quality assessment using deep neural networks[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1202-1213.

[34] GAO Fei, YU Jun, ZHU Suguo, et al. Blind image quality prediction by exploiting multi-level deep representations[J]. Pattern Recognition, 2018, 81: 432-442.

[35] SU S L, YAN Q S, ZHU Y, et al. Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Seattle, WA, USA:IEEE:3664-3673.

[36] SUCHOW J W, ALVAREZ G A. Motion silences awareness of visual change[J]. Current Biology, 2011, 21(2): 140-143.

[37] QIAN Jiansheng, WU Dong, LI Leida, et al. Image quality assessment based on multi-scale representation of structure[J]. Digital Signal Processing, 2014, 33(1): 125-133.

[38] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C, et al. A subjective study to evaluate video quality assessment algorithms[C]//Human Vision and Electronic Imaging XV. SPIE, 2010, 7527: 128-137.

[39] SESHADRINATHAN K, BOVIK A C. Motion tuned spatio-temporal quality assessment of natural videos[J]. IEEE Transactions on Image Processing, 2010, 19(2): 335-350.

[40] PARK J, SESHADRINATHAN K, LEE S, et al. Video quality pooling adaptive to perceptual distortion severity[J]. IEEE Transactions on Image Processing, 2012, 22(2): 610-620.

[41] GALKANDAGE C, CALIC J, DOGAN S, et al. Stereoscopic video quality assessment using binocular energy[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(1): 102-112.

[42] GALKANDAGE C, CALIC J, DOGAN S, et al. Full-reference stereoscopic video quality assessment using a motion sensitive HVS model[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(2): 452-466.

[43] WANG Z, LI Q. Video quality assessment using a statistical model of human visual speed perception[J]. JOSA A, 2007, 24(12): B61-B69.

[44] 陈义如.基于人眼视觉特性的视频质量评价方法研究[D].西安:西安电子科技大学, 2014.
CHEN Yiru. Research on Video Quality Evaluation Method Based on Human Visual Characteristics [D]. Xi'an: Xidian University, 2014. (in Chinese)

[45] VU P V, VU C T, CHANDLER D M. A spatiotemporal most-apparent-distortion model for video quality assessment[C]//2011 18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE, 2011: 2505-2508.

[46] WANG Y, JIANG T, MA S, et al. Novel spatio-temporal structural information based video quality metric[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(7): 989-998.

[47] LI X, GUO Q, LU X. Spatiotemporal statistics for video quality assessment[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3329-3342.

[48] MITTAL A, SAAD M, BOVIK A C. Assessment of video naturalness using time-frequency statistics [C] //2014 IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE, 2014: 571-574.

[49] MITTAL A, SAAD M A, BOVIK A C. A completely blind video integrity oracle[J]. IEEE Transactions on Im-

age Processing, 2015, 25(1): 289-300.

[50] SAAD M A, BOVIK A C, CHARRIER C. Blind prediction of natural video quality[J]. IEEE Transactions on Image Processing, 2014, 23(3): 1352-1365.

[51] LE CALLET P, VIARD-GAUDIN C, BARBA D. A convolutional neural network approach for objective video quality assessment[J]. IEEE Transactions on Neural Networks, 2006, 17(5): 1316-1327.

[52] ZHANG Yu, GAO Xinbo, HE Lihuo, et al. Objective video quality assessment combining transfer learning with CNN[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8): 2716-2730.

[53] KIM W, KIM J, AHN S, et al. Deep video quality assessor: From spatio-temporal visual sensitivity to a convolutional neural aggregation network[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 219-234.

[54] VINYALS O, BENGIO S, KUDLUR M. Order Matters: Sequence to sequence for sets[EB/OL]. https://arxiv.org/abs/1511.06391, 2015.

[55] YANG Jiaolong, REN Peiran, ZHANG Dongqing, et al. Neural aggregation network for video face recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4362-4371.

[56] LIU Wentao, DUANMU Zhengfang, WANG Zhou. End-to-end blind quality assessment of compressed videos using deep neural networks[C]//Proceedings of the 26th ACM International Conference on Multimedia. Seoul Republic of Korea. New York, NY, USA: ACM, 2018: 546-554.

[57] XU Munan, CHEN Junming, WANG Haiqiang, et al. C3DVQA: full-reference video quality assessment with 3D convolutional neural network[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain. IEEE, 2020: 4447-4451.

[58] LI Yuming, PO Laiman, XU Xuyuan, et al. No-reference image quality assessment using statistical characterization in the shearlet domain[J]. Signal Processing: Image Communication, 2014, 29(7): 748-759.

[59] WANG Chunfeng, SU Li, HUANG Qingming. CNN-MR for no reference video quality assessment[C]//2017 4th International Conference on Information Science and Control Engineering (ICISCE). Changsha, China: IEEE, 2017: 224-228.

[60] AHN S, LEE S. Deep blind video quality assessment based on temporal human perception[C]//2018 25th IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE, 2018: 619-623.

[61] VARGA D. No-reference video quality assessment based on the temporal pooling of deep features[J]. Neural Processing Letters, 2019, 50(3): 2595-2608.

[62] LOMOTIN K, MAKAROV I. Automated image and video quality assessment for computational video editing[C]// International Conference on Analysis of Images, Social Networks and Texts. Springer: Cham, 2020: 243-256.

[63] LI D, JIANG T, JIANG M. Quality assessment of in-the-wild videos[C]//Proceedings of the 27th ACM International Conference on Multimedia, 2019: 2351-2359.

[64] VARGA D, SZIRÁNYI T. No-reference video quality assessment via pretrained CNN and LSTM networks[J]. Signal, Image and Video Processing, 2019, 13(8): 1569-1576.

[65] HOSU V, HAHN F, JENADELEH M, et al. The Konstanz natural video database (KoNViD-1k)[C]//2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX). Erfurt, Germany: IEEE, 2017: 1-6.

[66] CHIU T Y, ZHAO Yinan, GURARI D. Assessing image quality issues for real-world problems[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 3643-3653.

[67] YING Zhenqiang, NIU Haoran, GUPTA P, et al. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 3572-3582.

[68] NUUTINEN M, VIRTANEN T, VAAHTERANOKSA M, et al. CVD2014—A database for evaluating no-reference video quality assessment algorithms[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3073-3086.

[69] GHADIYARAM D, PAN J, BOVIK A C, et al. In-capture mobile video distortions: A study of subjective behavior and objective algorithms[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(9): 2061-2077.

[70] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C, et al. Study of subjective and objective quality assessment of video[J]. IEEE Transactions on Image Processing, 2010, 19(6): 1427-1441.

[71] BAMPIS C G, LI Zhi, 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.

[72] GHADIYARAM D，PAN J，BOVIK A C. A subjective and objective study of stalling events in mobile streaming videos[J]. IEEE Transactions on Circuits and Systems for Video Technology，2019，29(1)：183-197.

[73] ABU-EL-HAIJA S，KOTHARI N，LEE J，et al. YouTube-8M：A large-scale video classification benchmark[EB/OL]. arXiv preprint arXiv：1609.08675，2016. https://arxiv.org/abs/1609.08675.

[74] GU Chunhui，SUN Chen，ROSS D A，et al. AVA：A video dataset of spatio-temporally localized atomic visual actions[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA：IEEE，2018：6047-6056.

[75] HOSU V，HAHN F，JENADELEH M，et al. The Konstanz natural video database (KoNViD-1k)[J]. 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)，2017，1(1)：1-6.

[76] SINNO Z，BOVIK A C. Large scale subjective video quality study[C]//2018 25th IEEE International Conference on Image Processing (ICIP). Athens，Greece：IEEE，2018：276-280.

[77] WANG Yilin，INGUVA S，ADSUMILLI B. YouTube UGC dataset for video compression research[C]//2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP). IEEE，2019：1-5.

[78] TU Zhengzhong，WANG Yilin，BIRKBECK N，et al. UGC-VQA：Benchmarking blind video quality assessment for user generated content[J]. IEEE Transactions on Image Processing，2021，30(1)：4449-4464.

[79] WINKLER S. Analysis of public image and video databases for quality assessment[J]. IEEE Journal of Selected Topics in Signal Processing，2012，6(6)：616-625.

[80] TU Zhengzhong，CHEN C J，CHEN Liheng，et al. Regression or classification? New methods to evaluate no-reference picture and video quality models[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP). Toronto，ON，Canada：IEEE，2021：2085-2089.

[81] TU Zhengzhong，CHEN C J，WANG Yilin，et al. Efficient user-generated video quality prediction[C]//2021 Picture Coding Symposium (PCS). Bristol，United Kingdom：IEEE，2021：1-5.

[82] WANG Yilin，KE Junjie，TALEBI Hossein，et al. Rich features for perceptual quality assessment of UGC videos[J]. Proceedings of the IEEECVF Conference on Computer Vision and Pattern Recognition，2021，1(1)：13435-13444.

[83] RAZEQ S N S A，KHRAISAT Y S H，IBAHIM M M A. Mobile station positioning using time difference of arrival and received signal strength[J]. International Journal of Mobile Communications，2012，10(6)：637.

[84] Recommendation I T U. 13871—2001，Methods for objective measurements of perceived audio quality[S]. Geneva：ITU-R BS，2001.

[85] CAVE C R.Perceptual modelling for low-rate audio coding[J]. eScholarship，2002，1(1).

[86] HUBER R，KOLLMEIER B. PEMO-Q—A new method for objective audio quality assessment using a model of auditory perception[J]. IEEE Transactions on Audio，Speech，and Language Processing，2006，14(6)：1902-1911.

[87] HINES A，GILLEN E，KELLY D，et al. ViSQOLAudio：An objective audio quality metric for low bitrate codecs[J]. The Journal of the Acoustical Society of America，2015，137(6)：EL449-EL455. DOI：10.1121/1.4921674.

[88] BARBEDO J G A，Lopes A.A new cognitive model for objective assessment of audio quality[J].Journal of the Audio Engineering Society，2005，53(12)：22-31.

[89] MOORE B C J，TAN C T，ZACHAROV N，et al.Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion[J].Journal of the Audio Engineering Society，2004，52(12)：1228-1244.

[90] MOORE B C，TAN C T. Perceived naturalness of spectrally distorted speech and music[J]. The Journal of the Acoustical Society of America，2003，114(1)：408-419.

[91] ZHENG Jia，ZHU Mengyao，HE Junwei，et al. PEAQ compatible audio quality estimation using computational auditory model[C]//Neural Information Processing. Berlin，Heidelberg：Springer Berlin Heidelberg，2012：83-90.

[92] ZHU Mengyao，ZHENG Jia，JIN C，et al. Structural Similarity Analysis of Modulation for audio quality assessment[C]//2013 IEEE International Conference on Acoustics，Speech and Signal Processing. Vancouver，BC，Canada：IEEE，2013：403-407.

[93] LI Zhonghua，WANG Juchiang，CAI Jingli，et al.Non-reference audio quality assessment for online live music recordings[J].Proceedings of the 21st ACM International Conference on Multimedia，2013，1(1)：63-72.

[94] 杨佳俊.网络音频质量无参考客观评估[D].西安：西安电子科技大学，2015.
YANG Jiajun. No-reference objective quality assessment for networked audio[D]. Xi'an：Xidian University，2015.

(in Chinese)

[95] DEFRAENE B, VAN WATERSCHOOT T, DIEHL M, et al. Subjective audio quality evaluation of embedded-optimization-based distortion precompensation algorithms [J]. The Journal of the Acoustical Society of America, 2016, 140(1): EL101-EL106. DOI:10.1121/1.4955025.

[96] 孙佳婷. 低码率音频质量客观评价算法研究[J]. 黑龙江大学工程学报, 2017, 8(2): 80-87.
SUN Jiating. Research on objective evaluation of low bit rate audio quality [J]. Journal of Engineering of Heilongjiang University, 2017, 8(2): 80-87.(in Chinese)

[97] 康健. 音频质量评价和语音识别预处理技术的研究及实现[D]. 北京:北京邮电大学, 2017.
KANG Jian. Research and implementation of audio quality evaluation and speech recognition preprocessing technology [D]. Beijing: Beijing University of Posts and Telecommunications, 2017. (in Chinese)

[98] 闵雄阔, 翟广涛, 杨小康. 基于自然音频统计特性的无参考音频质量评价方法和装置[P].CN111508528A, 2020.
MIN Xiongkuo, ZHAI Guangtao, YANG Xiaokang. No-reference audio quality evaluation method and device based on natural audio statistical characteristics [P]. CN111508528A. 2020. (in Chinese)

[99] YIGAL A, JILL B. Automated essay scoring with e-Rater v20 [J]. Journal of Technology, Learning, and Assessment, 2006, 4(3): 1-30.

[100] SWAPNA S, JILL B, MARTIN C. Lexical chaining for measuring discourse coherence quality in test-taker essays[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, 2014, 1(1): 950-961.

[101] MCNAMARA D S, CROSSLEY S A, ROSCOE R D, et al. A hierarchical classification approach to automated essay scoring[J]. Assessing Writing, 2015, 23: 35-59.

[102] FARRA N, SOMASUNDARAN S, BURSTEIN J. Scoring persuasive essays using opinions and their targets[C]//Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications. Denver, Colorado. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015, 1(1): 64-74.

[103] DONG Fei, ZHANG Yue. Automatic features for essay scoring-an empirical study[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, 1(1): 1072-1077.

[104] KAVEH T, HWEE T N. A neural approach to automated essay scoring [C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, 1(1): 1882-1891.

[105] DONG Fei, ZHANG Yue, YANG Jie. Attention-based recurrent convolutional neural network for automatic essay scoring[C]//Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017. Vancouver, Canada. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, 1(1): 153-162.

[106] DASGUPTA T, NASKAR A, DEY L, et al. Augmenting textual qualitative features in deep convolution recurrent neural network for automatic essay scoring[C]//Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications. Melbourne, Australia. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, 1(1): 93-102.

[107] MATHIAS S, KANOJIA D, PATEL K, et al. Eyes are the windows to the soul: Predicting the rating of text quality using gaze behaviour [C] //Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, 1(1): 2352-2362.

[108] DERIU J, RODRIGO A, OTEGI A, et al. Survey on evaluation methods for dialogue systems[J]. Artificial Intelligence Review, 2021, 54(1): 755-810.

[109] NOVIKOVA J, DUŠEK O, RIESER V. RankME: reliable human ratings for natural language generation[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans, Louisiana. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, 2(1): 72-78.

[110] KHASHABI D, STANOVSKY G, BRAGG J, et al. GENIE: A leaderboard for human-in-the-loop evaluation of text generation [EB/OL]. 2021: arXiv: 2101.06561 [cs.CL]. https://arxiv.org/abs/2101.06561.

[111] CONNEAU A, KIELA D, SCHWENK H, et al. Supervised learning of universal sentence representations from natural language inference data[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark. Stroudsburg, PA, USA: Association for Computational Linguistics,

2017, 1(1):670-680.

[112] HUANG Posen, HE Xiaodong, GAO Jianfeng, et al. Learning deep structured semantic models for web search using clickthrough data[C]//Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13. San Francisco, California, USA. New York: ACM Press, 2013, 1(1): 2333-2338.

[113] LOGESWARAN L, LEE H. An efficient framework for learning sentence representations[EB/OL]. 2018: arXiv: 1803.02893[cs.CL]. https://arxiv.org/abs/1803.02893.

[114] KIM H, LEE J H, NA S H. Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation[C]//Proceedings of the Second Conference on Machine Translation. Copenhagen, Denmark. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, 1(1):562-568.

[115] MARTINS A F T, KEPLER F, MONTEIRO J. Unbabel's participation in the WMT17 translation quality estimation shared task[C]//Proceedings of the Second Conference on Machine Translation. Copenhagen, Denmark. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, 1(1):569-574.

[116] ZHOU W, XU Ke. Learning to compare for better training and evaluation of open domain natural language generation models[C]//Proceedings of the AAAI Conference on Artificial Intelligence, New York, USA. 2020, 34(5): 9717-9724.

作者简介

颜成钢　男，1984年生，浙江杭州人。杭州电子科技大学教授，研究方向为智能信息处理。
E-mail：cgyan@hdu.edu.cn

孙垚棋　男，1993年生，浙江杭州人。杭州电子科技大学科研助理，研究方向为多媒体信息处理。
E-mail：syq@hdu.edu.cn

钟　昊　男，1999年生，江西新余人。杭州电子科技大学硕士研究生，研究方向为智能信息处理、大规模超图聚类。
E-mail：907251797@qq.com

朱晨薇　女，1999年生，浙江诸暨人。杭州电子科技大学硕士研究生，研究方向为图像的显著性检测。
E-mail：764052159@qq.com

朱尊杰　男，1994年生，浙江乐清人。杭州电子科技大学讲师，研究方向为计算机视觉与图形学。
E-mail：zunjiezhu@hdu.edu.cn

郑博仑（通讯作者）　男，1991年生，湖北武汉人。杭州电子科技大学讲师，研究方向为视频与图像处理、多媒体信息处理、工业视觉检测等。
E-mail：blzheng@hdu.edu.cn

周晓飞　男，1988年生，安徽淮北人。杭州电子科技大学讲师，研究方向为视频与图像处理、视觉显著目标检测与分割等。
E-mail：zxforchid@outlook.com