arXiv:2302.12393v1 [cs.MM] 24 Feb 2023

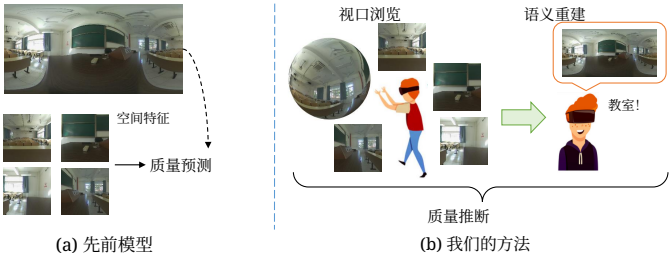
无参考全景图像质量评价：融合局部统计特征与全局语义

Wei Zhou 和 *Zhou Wang*

加拿大滑铁卢大学电气与计算机工程系，电子邮箱： {wei.zhou,zhou.wang}@uwaterloo.ca

摘要

全景图像质量评价（OIQA）旨在预测覆盖视觉环境整个 180×360° 视域范围的全景图像的感知质量。本文提出一种名为 S² 的无参考全景图像质量评价方法，弥合全景图像低层统计特征与高层语义之间的差距。具体而言，该方法分别从多个局部视口和幻构的全局全景图像中提取统计与语义特征。随后，通过质量回归和加权过程，将提取的质量感知特征映射为感知质量预测。实验结果表明，与最先进的方法相比，所提出的 S² 方法具有极具竞争力的性能。



(a) 先前模型 (b) 我们的方法

图 1。全景图像质量评估中的感知线索。现有模型从不同视口中提取空间信息，并可能借助全局投影图；而所提方法则结合局部图像统计与全局语义重建。

关键词— 全景图像、无参考图像质量评价、低层统计特征、高层语义

1. 引言

虚拟现实（VR）技术的飞速发展，使得为终端用户打造沉浸式多媒体体验质量（QoE）成为可能。作为虚拟现实的一种代表形式，全向内容日益融入我们的日常生活。为评估和优化全向内容的感知体验质量，客观的全景图像质量评价（OIQA）模型在现代虚拟现实系统的发展中发挥着关键作用。

在相关文献中，已经出现了遵循全参考（FR）和无参考（NR）框架的客观全景图像质量评价模型。全参考全景图像质量评价模型假设可以完整获取参考图像的信息，通常是针对常规矩形二维图像质量评价（IQA）开发的传统全参考方法的直接扩展。例如，Yu 等人 [1] 提出了球面峰值信噪比（S-PSNR）算法，该算法以球面上均匀分布的点为对象计算峰值信噪比，而非使用投影后的矩形图像。在 [2], 中，介绍了加权球面均匀峰值信噪比（WS-PSNR）方法，该方法通过考虑拉伸程度来构建权重图。Zakharchenko 等人 [3]提出了克拉斯特抛物线投影峰值信噪比（CPP-PSNR）方法，该方法将参考图像和失真全向图像映射到克拉斯特抛物线投影上，然后计算峰值信噪比。

无参考全向图像质量评价方法不需要访问参考图像，因此在许多应用场景中更具实用性。现有的无参考全向图像质量评价方法通常可分为两类，具体取决于它们在质量预测中采用的是传统手工设计特征还是学习得到的深层特征。MFILGN模型[4]利用多频信息和局部—全局自然度。较新的模型则采用深度卷积神经网络（卷积神经网络）或图卷积网络。这些模型表现出良好的性能，包括用于盲式 360 度图像质量评价的多通道卷积神经网络（MC360IQA） [5], 、视口定向图卷积网络（VGCN） [6], 及其名为自适应超图卷积网络（AHGCN）的变体 [7]。

在 360度观看环境中，例如使用头戴式设备时，观察者无法同时看到整个全景内容，因此，人类主观观看体验中的一个重要环节，是通过浏览多个视口并整合其中的信息，建立或重建对全局语义的感知。在图像质量评价过程中，这种全局语义会与关于图像保真度、自然度和/或伪影的局部观察相结合，从而得出

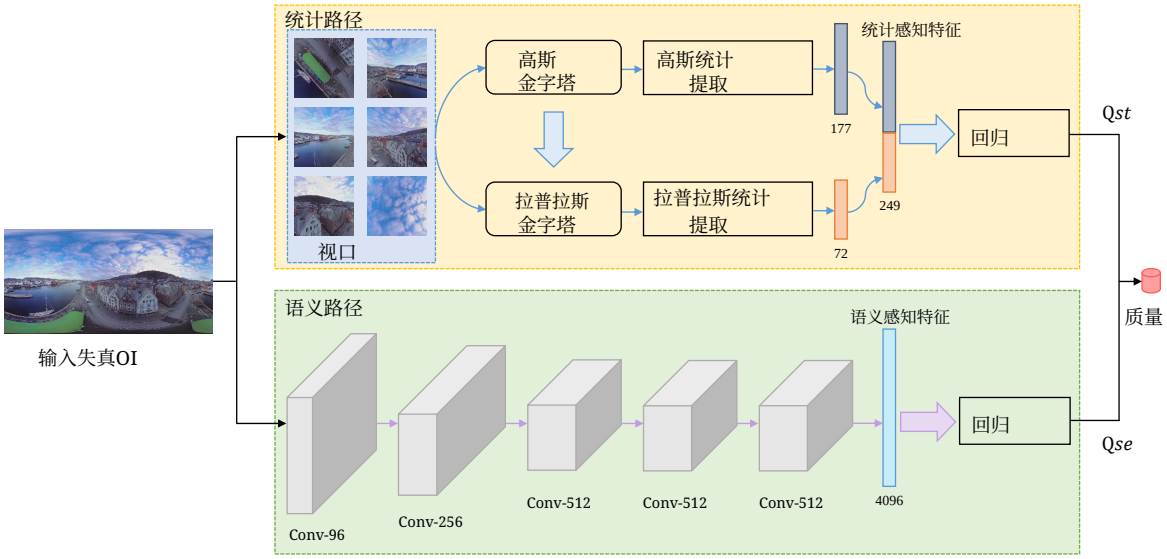


图 2。用于无参考全景图像质量评价的所提 S^2 方法框架。

整体质量评价。受这一观察启发，我们提出了一种面向统计与语义的质量预测框架 S^2，用于盲全景图像质量评价（OIQA），如图 1 所示。该框架整合了从多个局部视口的低层图像统计特征，以及从生成的全局全景图像中提取的高层语义特征。随后，利用质量回归模块，将从两条独立路径中提取的质量敏感特征集合映射为整体主观质量评分预测。大量实验结果表明，所提方法优于许多最先进的质量评估模型。此外，我们还通过消融研究，观察了语义置信度与图像失真之间的关系，以及各个组成部分如何影响最终的质量预测性能。

2. 所提方法

所提出的 S^2 方法整体框架如图 2 所示，由统计路径、语义路径和最终的质量回归步骤组成。

由于观看者会浏览各种视口，我们首先将失真全向图像（OI）转换为多个视口。给定每个输入失真 OI（记为 D），我们采用非均匀视口采样策略[8, 9]，获得 N 个视口 V_n，$n = 1, 2, \ldots\ldots, N$。

为捕捉人类视觉系统的多尺度特征 [10]，，我们构建多个局部视口的金字塔表示[11, 12] 。具体而言，通过迭代进行高斯滤波、下采样和相减，生成多层拉普拉斯金字塔 [13]；在此过程中也会生成高斯金字塔。对于特定的视口 V_n，高斯金字塔的各层计算如下：

$$G_n^i(x,y) = \begin{cases} V_n, i = 1 \\ \sum_{u=-2}^{2}\sum_{v=-2}^{2} k(u,v)G_n^{i-1}(2x+u, 2y+v), i > 1 \end{cases}, \tag{1}$$

其中，i 是高斯金字塔的层索引，$x \in [0, X)$ 和 $y \in [0, Y)$ 是像素位置索引，其中 X 和Y分别为图像的尺寸；$k(u, v)$ 表示生成核，通常由低通滤波器（例如二维高斯滤波器）的系数定义。

然后，我们对高斯金字塔的每一层进行如下插值：

$$\hat{G}_n^i(x,y) = 4\sum_{u=-2}^{2}\sum_{v=-2}^{2} k(u,v)G_n^i\left(\frac{u+x}{2}, \frac{v+y}{2}\right). \tag{2}$$

高斯金字塔当前层与下一层插值结果之间的残差定义为拉普拉斯金字塔的当前层：

$$L_n^i = G_n^i - \hat{G}_n^{i+1}. \tag{3}$$

由于拉普拉斯金字塔的第 i层计算需要高斯金字塔的第 $i+1$层，因此拉普拉斯金字塔的层数比高斯金字塔少一层。

为从高斯金字塔中提取特征，我们计算默认的均匀局部二值模式（LBP）描述符，从而为每个高斯层得到 59 个统计量。使用 3 层高斯金字塔时，可得到 177 个高斯金字塔特征，记为 f_{GP}。对于拉普拉斯金字塔，受自然场景统计（NSS）在图像质量评价研究中的成功启发 [14, 15, 16]，，我们提取均值去除并对比度归一化后的系数，从而得到

个特征。使用 2 层拉普拉斯金字塔时，可得到 72 个拉普拉斯金字塔特征，记为 f_{LP}。完整的统计特征集 f_{st}（每个视口对应一组）通过拼接从高斯金字塔和拉普拉斯金字塔中提取的统计特征得到，如下所示：

$$f_{st} = [f_{GP}, f_{LP}]. \tag{4}$$

我们采用在大型 ImageNet 数据集上训练的 VGGNet [17]作为语义特征提取骨干网络，主要是因为其结构简单，并且能够捕捉与图像失真相关的表征 [18]。在 [19], 中，VGGNet 提出了三种不同的结构，以在复杂性和准确率之间取得平衡，分别为快速 VGG（VGG-F）、中速 VGG（VGG-M）和慢速 VGG（VGG-S）。每种结构都包含 5 个卷积（Conv）层和 3 个全连接（FC）层。前两个 FC 层各有 4,096 个神经元，而最后一层有 1,000 个节点，表示用于图像识别的 1,000 个类别。在当前实现中，我们选择 VGG-M 第一层 FC 层的深层特征作为语义特征集 f_{se}：

$$f_{se} = FC_1(D). \tag{5}$$

为了学习从特征到质量标签的映射关系，我们将统计特征和语义特征分别输入支持向量回归（SVR）模型 [20], ，并将回归得到的统计质量分数和语义质量分数分别记为 Q_{st} 和Q_{se}。总体质量分数通过加权平均计算：

$$Q_{\text{overall}} = wQ_{st} + (1-w)Q_{\text{se}}, \tag{6}$$

其中， w 是一个加权因子，用于确定统计特征预测器和语义特征预测器的相对重要性。

3. 验证

3.1. 实验设置与性能比较

我们在 CVIQD 主观数据库上评估所提出的方法。该数据库是目前规模较大且广泛采用的数据库之一，包含全景图像及其由人类受试者给出的相应质量标签。它由 16 张原始图像和 528 张失真图像组成，这些失真图像由三种经典图像或视频编码技术生成，分别是 JPEG、AVC 和 HEVC。主观质量评分以平均意见分（MOS）的形式给出，并重新缩放至 [0, 100], 范围；较高的 MOS 表示更好的图像感知质量。

为比较各种图像质量评价模型的性能，我们采用斯皮尔曼等级相关系数（SROCC）、皮尔逊线性相关系数（PLCC）和均方根误差（RMSE）作为评估指标。在计算 PLCC 和 RMSE 之前，采用五参数逻辑非线性拟合方法 [22] 将预测质量映射到主观质量空间。

表 1。客观模型的性能比较。

类型	方法	SROCC	PLCC	RMSE
全参考图像质量评价	PSNR	0.6239	0.7008	9.9599
	结构相似性 [23]	0.8842	0.9002	6.0793
	多尺度结构相似性 [10]	0.8222	0.8521	7.3072
	特征相似性 [24]	0.9152	0.9340	4.9864
	DeepQA [25]	0.9292	0.9375	4.8574
全参考全景图像质量评估	球面峰值信噪比 [1]	0.6449	0.7083	9.8564
	加权球面均匀峰值信噪比 [2]	0.6107	0.6729	10.3283
	CPP-PSNR [3]	0.6265	0.6871	10.1448
无参考图像质量评价	自然图像统计质量评估 [26]	0.8180	0.8376	7.6271
	BMPRI [27]	0.7470	0.7919	8.5258
	DB-CNN [28]	0.9308	0.9356	4.9311
NR-OIQA	MFILGN [4]	0.9670	0.9751	3.1036
	MC360IQA [5]	0.9428	0.9429	4.6506
	VGCN [6]	0.9639	0.9651	3.6573
	AHGCN [7]	0.9623	0.9643	3.6990
	所提 S^2	**0.9710**	**0.9781**	**2.8945**

数据库被随机划分为 80% 的数据用于训练，其余 20% 用于交叉验证。为降低训练集/测试集划分带来的不确定性，我们重复这一随机划分与交叉验证过程 100 次，并报告性能中位数。

我们将所提出算法的性能与当前先进的质量评估模型进行了比较，包括五种全参考图像质量评价方法、三种全参考全景图像质量评估方法、三种无参考图像质量评价方法和四种无参考全景图像质量评估方法。结果如表 1 所示。对于全参考图像质量评价指标，我们发现，基于峰值信噪比的模型不如结构相似性、多尺度结构相似性、特征相似性等更先进的方法，以及 DeepQA 等深度学习模型。令人有些意外的是，全参考全景图像质量评估方法并未在全参考图像质量评价方法的基础上进一步提升性能。相比之下，无参考全景图像质量评估模型明显优于无参考图像质量评价方法。这可能是因为这些模型专门针对全景图像的特性进行了设计。在所有测试指标中，所提出的 S^2 方法表现出很强的竞争力。

3.2. 语义置信度与图像失真

由于所提方法包含一条语义路径，因此观察语义置信度与图像失真之间的关系很有意思。图 3 展示了不同 JPEG、AVC 和 HEVC 失真级别下全景图像的失真示例。从第一列到第三列可以观察到，随着失真程度增加，语义置信度会降低。这表明语义信息可能与图像感知质量密切相关。另一个有趣的现象是，语义置信度对不同失真类型表现出不同的敏感度。具体而言，在较先进的图像/视频编码方法

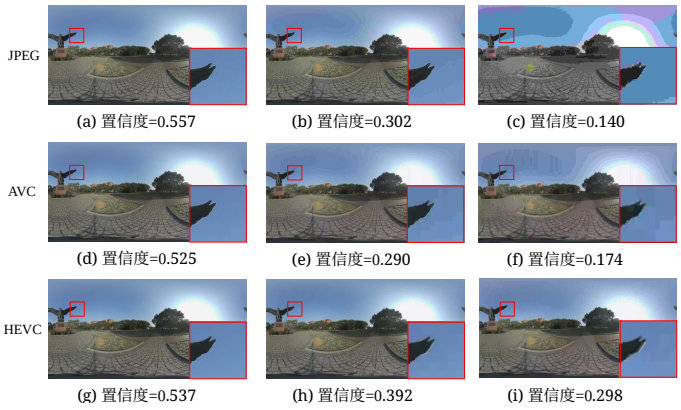


图 3。语义置信度与图像失真之间的关系。第一、二、三行分别对应三种失真类型（JPEG、AVC 和 HEVC 压缩）。对于每种失真类型，第一、二、三列分别对应逐渐增加的失真级别（低、中和高）。

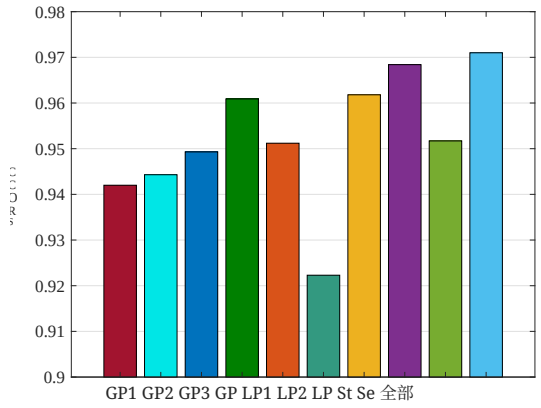


图 4。消融实验的性能结果。

HEVC 中，语义置信度的下降幅度远小于较早的 JPEG 和 AVC 编码器。

3.3. 消融与参数敏感性测试

我们通过消融实验评估统计路径和语义路径的贡献，结果如图 4 所示。其中，GP1、GP2 和 GP3 分别表示仅使用高斯金字塔统计的第一层、第二层和第三层的情况。GP 表示使用高斯金字塔统计三层的情况。我们发现，性能逐步提升。类似地，LP1 和 LP2 分别表示仅使用拉普拉斯金字塔统计第一层和第二层的情况，而 LP 表示使用 2 层拉普拉斯金字塔统计的情况。结果表明，在这三种情况下，LP 的性能最佳。仅采用统计路径和仅采用语义路径的情况分别记为 St 和 Se。可以

表 2。统计路径中不同视口数量的性能比较。

数量 SROCC PLCC			RMSE
6	0.9684	0.9769	3.0083
20	0.9686	0.9777	2.9626
80	0.9683	0.9771	2.9501

表 3。语义路径中不同神经网络架构的性能比较。

架构	SROCC	PLCC	RMSE
VGG-F	0.9497	0.9537	4.2107
VGG-M	**0.9517**	**0.9576**	**4.0329**
VGG-S	0.9451	0.9486	4.4345

观察到，单独采用任一路径都能取得较好的质量预测性能，但同时采用两条路径（即 All 情况）可获得最佳性能。相对而言，统计路径似乎是更主要的因素。这或许并不令人意外，因为统计特征直接来自人类受试者所观察的不同视口，而全局语义则为质量评估提供了额外线索等互补信息。

由于所提出框架的具体实现可能采用不同的参数设置，因此我们在此测试模型对不同视口数量和语义架构的敏感性。结果分别见表 2 和表 3。可以看出，所提出的模型对视口数量并不敏感。因此，在实际应用中，我们可以减少视口数量（例如，减少至 6 个），以降低计算复杂度。结果还表明，在语义路径中，VGG-M 的表现优于其他神经网络架构。可能的原因是，VGG-M 在算法复杂度和准确率之间实现了较为理想的折中，因此适合作为深度语义骨干网络。

4. 结论

我们提出了一种新颖的 S^2 框架，用于无参考全景图像质量评价，将局部低层统计特征与全局高层语义特征相结合。大量实验表明，所提方法达到了最先进的性能。关于语义置信度与图像失真之间关系的观察，以及消融／敏感性测试，提供了其他有用的见解。在相同框架下，未来可以采用更先进的统计分析和语义分析模型，以构建更准确的体验质量评估模型，从而推动沉浸式多媒体系统的发展。

5. 参考文献

[1] Matt Yu、Haricharan Lakshman 和 Bernd Girod，“一种评估全向视频编码方案的框架”，载于*IEEE*国际混合与增强现实研讨会，2015年，第31–36页。[2] Yule Sun、Ang Lu 和 Lu Yu，“面向全向视频的加权球面均匀质量评价”，*IEEE*信号处理快报，第24卷，第9期，第1408–1412页，2017年。[3] Vladyslav Zakharchenko、Kwang Pyo Choi 和 Jeong Hoon Park，“ 球形全景视频质量指标 ”，载于信息处理光学与光子学*X*。国际光学与光子学学会，2016年，第9970卷，第99700C页。[4] Wei Zhou、Jiahua Xu、Qiuping Jiang 和 Zhibo Chen，“通过分析多频信息和局部—全局自然度对360度图像进行无参考质量评价”，*IEEE*视频技术电路与系统汇刊，2021年。[5] Wei Sun、Xiongkuo Min、Guangtao Zhai、Ke Gu、Huiyu Duan 和 Siwei Ma，“MC360IQA：用于无参考360度图像质量评价的多通道卷积神经网络”，*IEEE*信号处理精选主题期刊，第14卷，77，2019 1，64第。页–。[6] Jiahua Xu、Wei Zhou 和 Zhibo Chen，“基于视口定向图卷积网络的无参考全景图像质量评价”，*IEEE*视频技术电路与系统汇刊，第31卷，第5期，第1724–1737页，2020年。[7] J un Fu、Chen Hou、Wei Zhou、Jiahua Xu 和 Zhibo Chen，“用于无参考360度图像质量评价的自适应超图卷积网络”，载于第*30*届*ACM*多媒体国际会议论文集，2022年，第961–969页。[8] Jiahua Xu、Ziyuan Luo、Wei Zhou、Wenyuan Zhang 和 Zhibo Chen，“基于多视口的立体360度图像质量评价”，载于*IEEE*图像编码研讨会，2019年，第1–5页。[9] Zhibo Chen、Jiahua Xu、Chaoyi Lin 和 Wei Zhou，“基于预测编码理论的立体全景图像质量评价”，*IEEE*信号处理精选主题期刊，第14卷，第1期，第103–117页，2020年。[10] Zhou Wang、Eero P Simoncelli 和 Alan C Bovik，“用于图像质量评价的多尺度结构相似性”，载于第三十七届阿西洛马会议：信号、系统与计算机，*2003*。IEEE，2003年，第2卷，第1398–1402页。[11] Zhou Wang 和 Eero P Simoncelli，“利用小波域自然图像统计模型进行简化参考图像质量评价”，载于人类视觉与电子成像*X*。国际光学与光子学学会，2005年，第5666卷，第149–159页。[12] Valero Laparra、Jo hannes Ball´e、Alexander Berardino 和 Eero P Simoncelli，“利用归一化拉普拉斯金字塔进行感知图像质量评价”，电子成像，第2016卷，第16期，第1–6页，2016年。[13] Peter J Burt 和 Edward H Adelson，“拉普拉斯金字塔：一种紧凑图像编码”，载于计算机视觉读本，第671–679页。爱思唯尔，1987年。[14] Anish Mittal、Rajiv Soundararajan 和 Alan C Bovik，“构建一个‘完全盲’的图像质量分析器”，*IEEE*信号处理快报，第20卷，第3期，第209–212页，2012年。[15] Yuming Fang、Kede Ma、Zhou Wang、Weisi Lin、Zhijun Fang 和 Guangtao Zhai，《基于自然场景统计的对比度失真图像无参考质量评价》，*IEEE*信号处理快报，第 22 卷，2014 年，第 7 期，页码 838–842。[16] Zhibo Chen、Wei Zhou 和 Weiping Li，《无参考立体视频质量评价：从深度感知到整体体验》，*IEEE*图像处理汇刊，第 27 卷，第 2 期，页码 721–734，2017 年。[17] Jia Deng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei，《ImageNet：一个大规模分层图像数据库》，载于*IEEE*计算机视觉与模式识别会议论文集，2009 年，页码 248–255。[18] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang，《深度特征作为感知度量的非凡效能》，载于*IEEE*计算机视觉与模式识别会议论文集，2018 年，页码 586–595。[19] Ken Chatfield、Karen Simonyan、Andrea Vedaldi 和 Andrew Zisserman，《细节中的魔鬼归来：深入探究卷积网络》，*arXiv*预印本 *arXiv:1405 3531*，2014 年。。[20] Chih-Chung Chang 和 Chih-Jen Lin，《LIBSVM：支持向量机库》，*ACM*智能系统与技术汇刊，第 2 卷，第 3 期，页码 1–27，2011 年。[21] Wei Sun、Ke Gu、Siwei Ma、Wenhan Zhu、Ning Liu 和 Guangtao Zhai，《大型压缩 360 度球面图像数据库：从主观质量评价到客观模型比较》，载于*IEEE*第*20*届多媒体信号处理国际研讨会，2018 年，页码 1–6。[22] 视频质量专家组等，《视频质量专家组关于视频质量评估客观模型验证的最终报告，第二阶段》，*VQEG*，2003 年。[23] Zhou Wang、Alan C Bovik、Hamid R Sheikh 和 Eero P Simoncelli，《图像质量评价：从误差可见性到结构相似性》，*IEEE*图像处理汇刊，第 13 卷，第 4 期，页码 600–612，2004 年。[24] Lin Zhang 、Lei Zhang、Xuanqin Mou 和 David Zhang，《FSIM：图像质量评价的特征相似性指数》，*IEEE*图像处理汇刊，第 20 卷，第 8 期，页码 2378–2386，2011 年。[25] Jongyoo Kim 和 Sanghoon Lee，《图像质量评价框架中的人类视觉敏感度深度学习》，载于*IEEE*计算机视觉与模式识别会议论文集，2017 年，页码 1676–1684。[26]Anish Mittal、Anush Krishna Moorthy 和 Alan Conrad Bovik，《空间域中的无参考图像质量评价》，*IEEE*图像处理汇刊，第 21 卷，第 12 期，页码 4695–4708，2012 年。[27] Xiongkuo Min、Guangtao Zhai、Ke Gu、Yutao Liu 和 Xiaokang Yang，《通过失真加剧进行盲图像质量估计》，*IEEE*广播汇刊，第 64 卷，第 2 期，页码 508–517，2018 年。[28] Weixia Zhang、Kede Ma、Jia Yan、Dexiang Deng 和 Zhou Wang，《使用深度双线性卷积神经网络进行无参考图像质量评价》，*IEEE*视频技术电路与系统汇刊，第 30 卷，2018 年，第 1 期，页码 36–47。

[1]© 2023 IEEE。允许个人使用本材料。对于其他任何用途，包括在当前或未来媒体中重印/再版本材料以用于广告或宣传用途、创作新的汇编作品、转售或再分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护的组成部分，均须事先获得 IEEE 的许可。