

# 基于大核卷积分解和多尺度注意力的图像质量评价

林重成, 王同罕\*, 贾惠珍

(东华理工大学 信息工程学院, 江西 南昌 330013)

**摘要:**近年来,计算机视觉领域广泛采用注意力机制来模拟人类视觉系统(HVS)的感知过程。然而,自注意力模型(如Transformer)在图像处理中的应用面临计算复杂度高和信息丢失的挑战。为此,该文提出了一种基于大核卷积分解和多尺度注意力机制的图像质量评价方法(LMA-BIQA)。该方法采用双分支结构网络设计,其中一个分支基于注意力机制,通过大核卷积捕获图像中远距离特征之间的依赖关系,以生成多尺度注意力特征,并通过分解大核卷积实现模型的轻量化,用于提取图像的高级内容特征。另一个分支利用自然场景统计(NSS)方法提取图像的低级质量特征,以补偿大核卷积在裁剪缩放过程中导致的信息丢失。该方法在LIVE、CISQ、TID2013和LIVE-C四个标准数据集上均表现出优异的性能,加权平均后的斯皮尔曼等级相关系数和皮尔森线性相关系数分别达到0.9285和0.9375。此外,LMA-BIQA显著减少了计算时间,参数数量和计算时间对比Re-IQA模型降低了65%以上,且与人类主观评价有较高的一致性。

**关键词:**大核卷积分解;无参考图像质量评价;自然场景统计;多尺度注意力机制;深度学习

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2025)06-0175-07

doi:10.20165/j.cnki.ISSN1673-629X.2025.0011

## Image Quality Assessment Based on Decomposed Large Kernel Convolution and Multi-scale Attention

LIN Chong-cheng, WANG Tong-han\*, JIA Hui-zhen

(School of Information Engineering, East China University of Technology, Nanchang 330013, China)

**Abstract:** In recent years, attention mechanisms have been widely used in the field of computer vision to simulate the perceptual process of the human visual system (HVS). However, the application of self-attention models, such as Transformers, in image processing faces challenges of high computational complexity and information loss. To address this, an image quality assessment method based on decomposed large kernel convolution and multi-scale attention mechanisms (LMA-BIQA) is proposed. This method employs a dual-branch network design. One of the branches is based on the attention mechanism. Through large kernel convolution, the dependency relationships between long-range features in the image are captured to generate multi-scale attention features. The lightweight of the model is achieved by decomposing large kernel convolution, which is used to extract advanced content features of the image. Another branch uses the Natural Scene Statistics (NSS) method to extract the low-level quality features of the image to compensate for the information loss caused by large kernel convolution during the cropping and scaling process. The proposed method demonstrates excellent performance across four standard datasets: LIVE, CISQ, TID2013, and LIVE-C. The weighted average Spearman rank correlation coefficient and Pearson linear correlation coefficient reach 0.9285 and 0.9375, respectively. Additionally, LMA-BIQA significantly reduces computation time, with parameter count and computational time reduced by over 65% compared to the Re-IQA model, while maintaining a high consistency with human subjective evaluations.

**Key words:** large-kernel convolutional decomposition; no-reference image quality assessment; natural scene statistics; multi-scale attention mechanism; deep learning

## 0 引言

图像质量评价(Image Quality Assessment, IQA)在计算机视觉和图像处理领域中扮演着重要角色。社交

平台上每天都有数百万张图片被传播,图像在采集和传输过程中可能会产生失真,影响了人们获取视觉信息的能力<sup>[1]</sup>。因此,准确符合人类感知的IQA对于提

收稿日期:2024-11-01

修回日期:2025-03-05

基金项目:国家自然科学基金资助项目(62266001,62261001)

作者简介:林重成(2000-),男,硕士研究生,研究方向为计算机视觉;通信作者:王同罕(1984-),男,副教授,博士,CCF会员号(84709M),研究方向为计算机视觉。

升视觉体验至关重要。

根据评价主体的不同, IQA 可以分为客观质量评价和主观质量评价。然而, 由于主观质量评价通常需要大量的人力和财力, 难以在大规模场景中广泛应用, 客观质量评价体系在该领域得以迅速发展。客观质量评价通过科研人员提供的多维评分来构建数据集, 从而中立地评估方法的优劣。

根据所需参考信息的多少, 客观质量评价可以简单分为全参考图像质量评价 (Full Reference IQA, FR-IQA) 和无参考图像质量评价 (No Reference IQA, NR-IQA)<sup>[2]</sup>。许多 FR-IQA 算法 (如 LPIPS<sup>[3]</sup> 等) 在性能上优于同一阶段的 NR-IQA 方法, 但在现实的应用中, 几乎没有完美的未失真图像可用于量化 FR-IQA 算法的优劣。相反, 像 BRISQUE<sup>[4]</sup>、PaQ-2-PiQ<sup>[5]</sup> 和 CONTRIQUE<sup>[6]</sup> 等的 NR-IQA 算法不需要参考图像, 凭借其便利性和快捷性, 已成为 IQA 领域的研究重点。

近年来, 随着深度学习技术的发展, 众多新模型和方法不断被引入到 IQA 领域。在基于卷积神经网络 (Convolutional Neural Network, CNN) 的无参考图像质量评价 (NR-IQA) 中, 研究者们提出了多种改进方案。例如, 文献[7]基于合成失真和真实失真图像的不同, 设计了两个卷积神经网络, 并通过双线性池化 (bilinear pooling) 结合这两种情况。文献[8]提出了利用卷积自动编码器 (Convolutional Auto Encoder, CAE) 生成畸变图, 证明了所提出的失真图生成算法在各种失真类型上具有鲁棒性。Re-IQA<sup>[9]</sup> 模型通过双分支结构分别提取高级内容特征和低级质量特征, 最终得出这两者是互补的。然而, 这些算法通常要求固定大小的图像输入, 这就要求在预处理时对图像尺寸进行调整, 这种调整在 IQA 任务中可能影响图像质量, 从而导致预测偏差。

当人类视觉系统聚焦于一张图像时, 会对不同区域施加不同的注意力。神经学研究表明, 人类视觉系统 (Human Vision System, HVS) 在聚焦图像内容时, 对关注区域赋予更高的视觉权重。因此, 引入注意力机制可以模拟这一特性, 为图像的不同区域设置不同比例的注意力权重。许多 NR-IQA 方法采用自注意力模型 (Transformer) 来实现这一机制。TRIQ<sup>[10]</sup> 模型利用 CNN 的归纳能力进行质量特征推导, 并通过 Transformer 编码器实现注意力机制的聚合表示, 从而处理不同分辨率的图像输入。TReS<sup>[11]</sup> 模型则采用了混合 CNN 和 Transformer 的自注意力机制, 提取图像的局部和非局部特征。这些算法均充分利用了 CNN 和 Transformer 的优势, 取得了相对优异的性能。

尽管 Transformer 模型在性能上表现优异, 但仍存

在诸多问题, 许多研究者对它为何产生如此优异的结果而感到困惑。文献[12]尝试用池化层或空间多层感知机 (Spatial Multilayer Perceptron) 替换 Transformer 模型中的自注意力模块, 发现 Transformer 的宏观架构仍能保持良好的效果。文献[13]指出: 一维结构用于处理二维图像是不合理的, 这种强行将图像视为一维序列的做法忽视了其二维特性。而自注意机制在处理图像时又会产生二次时间复杂度, 对于高分辨率图像而言, 既耗时又昂贵, 难以在轻量级设备中实现<sup>[14]</sup>。同时, 相较于 CNN, Transformer 仅捕捉了空间适应性, 而忽略了通道适应性。针对这些缺陷, 研究者们提出了一些改进方案。例如, TECNet<sup>[15]</sup> 在解码器部分采用残差神经网络替代 Transformer, 从而降低了整体计算复杂度。为更好地利用注意力机制处理图像, 同时避免 Transformer 模型中的信息损失, 许多研究者将目光转向大核卷积。实际上, 大型卷积核的使用并非新鲜事物, 早在 AlexNet<sup>[16]</sup>、早期的 ResNet<sup>[17]</sup> 以及 GoogleNet Inception V1<sup>[18]</sup> 模型中就得到了广泛应用。近年来, 随着自注意力机制的流行, 文献[17]提出了大核注意力机制, 通过大核卷积捕获图像的远距离依赖。其独特之处在于无需深度堆叠即可获得大感受野, 避免了深度增加带来的边际递减问题, 并能实现类似于 Transformer 的全局尺度窗口效果。文献[19]提出的 UniRepLKNNet 模型指出: 在现代 CNN 设计优化下, 大核卷积可以显著提升有效感受野而不依赖深度堆叠。然而, 由于大核卷积带来了大量计算和参数, 同时降低了与各种硬件平台的兼容性, 导致在轻量级应用中仍面临挑战。文献[13]对 21×21 的卷积进行了分解, 使用较小的卷积核和空洞卷积在保持感受野不变的同时大幅降低了计算量。文献[20]在此基础上进一步拆分空洞卷积, 从而再次降低了模型的计算量。此外, Shift-ConvNets<sup>[21]</sup> 模型通过平移标准卷积实现大卷积核效果, 并通过修剪获得稀疏组卷积。

为了解决上述问题并降低基于 Re-IQA 模型的无参考图像质量评价 (NR-IQA) 方法的计算复杂度, 本算法在现有优秀方法中引入分解式大核卷积, 以取代原有的高级内容特征提取器。通过将大卷积核分解为深度卷积、深度空洞卷积和通道卷积的方式, 能够在较低计算开销下实现 Transformer 的注意力机制, 同时支持多尺度特征提取, 模拟人类视觉系统在不同尺度下观察图像的行为。此外, 本算法用传统的自然场景统计 (Natural Scene Statistics, NSS) 特征提取器提取图像的低级质量特征。为此, 该文提出了一种新的基于大核卷积分解的多尺度注意力机制的无参考图像质量评价算法 (Large-kernel decomposition and Multi-scale Attention Mechanism based BIQA, LMA-BIQA), 旨在

降低编码器的计算负担,提升模型性能,方便部署。模型采用双分支结构,分别使用传统的自然场景统计方法和深度学习中的大核卷积分解算法来提取图像的低级质量特征和高级内容特征。同时,设计了多级权重调节机制,以便对图像中高低级特征的总权重进行调整,使得模型更好地适应不同失真类型的应用场景。

## 1 LMA-BIQA 方法

### 1.1 LMA-BIQA 总体框架

Re-IQA 原模型主要通过对比学习训练两组编码器,分别提取图像的高级内容特征和低级质量特征,并利用这些互补特征来训练线性回归模型。本实验的工作旨在降低 Re-IQA 的计算消耗,同时保留其基本框架。在高级内容特征提取器中,采用分解大核多尺度注意力模块替代原有编码器。这样既能保持模拟人类视觉系统的注意力机制,又能大幅降低计算要求。在低级质量特征提取器方面,选择自然场景统计(NSS)中的特征提取器,以提取图像的低级特征,弥补深度学

习过程中因图像裁剪和缩放所造成的信息丢失。

图1展示了LMA-BIQA的总体框架。该网络结构包括一个高级内容特征分支和一个低级质量特征分支。在高级内容特征分支中,通过大核多尺度注意力模块提取多个尺度的特征,随后进行全局平均池化(Global Average Pooling, GAP)以获得多尺度特征。在低级质量特征分支中,采用NIQE模型的NSS特征提取器。首先对图像进行正则化,得到均值减损对比归一化(Modified Structural Similarity Coefficient, MSCN)系数,因原始质量图像的MSCN系数符合高斯分布,失真图像则不符合该统计规律<sup>[22]</sup>。通过量化两者之间的差异提取图像特征,然后利用广义高斯分布(Generalized Gaussian Distribution, GGD)进行拟合,并通过非对称广义高斯分布(Asymmetric Generalized Gaussian Distribution, AGGD)求取图像块的16个特征。从1和1/2尺度中提取NSS特征,并将其输入到两个全连接层中,这些层使用PReLU作为激活函数,以加速训练过程。

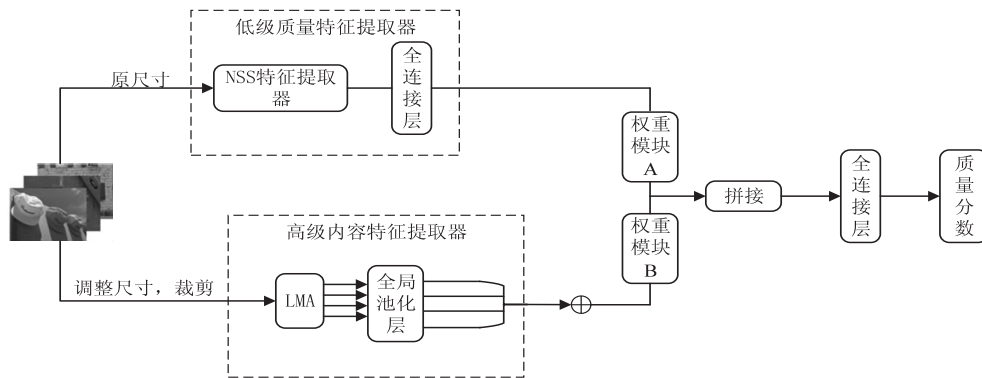


图1 LMA-BIQA 整体架构

该算法采用三层全连接的多层感知机层(MLP)来融合低级质量特征和高级内容特征,使用均方误差损失函数(Mean Squared Error, MSE)最小化回归损失,以稳定训练并防止梯度爆炸<sup>[23]</sup>。

$$f_{\text{MSE}} = \frac{\sum_i^n (y_i - y_i^p)^2}{n} \quad (1)$$

其中,  $f_{\text{MSE}}$  为均方误差损失函数。

LMA-BIQA 模型主要包括三个步骤:

(1) 使用真实失真图像或合成失真图像作为模型输入,经过不同处理后进入双分支进行不同尺度的特征提取。

(2) 分支一通过分解大核卷积提取多尺度注意力特征,以模拟人眼的聚焦效果,从而提取高级内容特征;分支二采用传统自然场景统计中的特征提取器来提取低级内容特征,以弥补因图像缩放导致的底层信息丢失。

(3) 在测试阶段,将测试数据集中的图像输入到

训练好的模型中,以获得质量评价分数。

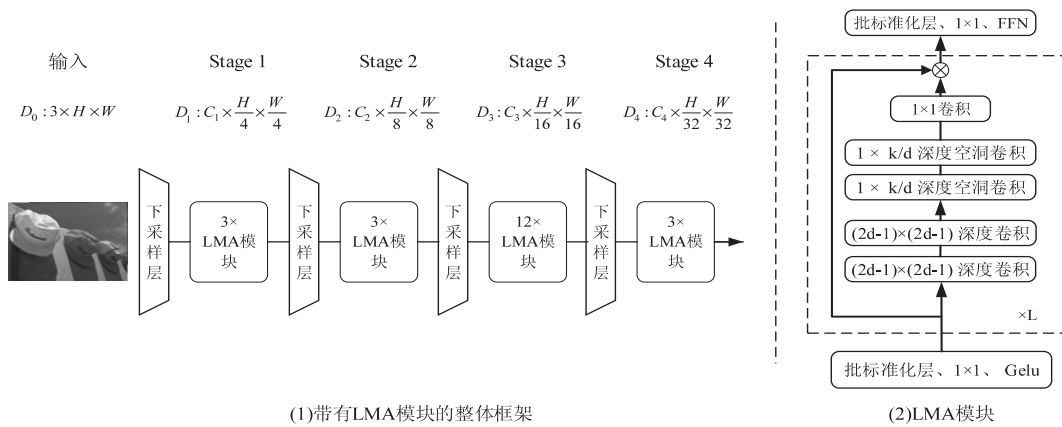
### 1.2 LMA 模块具体实现

在计算机视觉领域,通常使用注意力机制模拟人眼视觉系统中的聚焦效果。注意力过程是一个自适应的调节过程,能够根据输入信息动态调整输出<sup>[24]</sup>。这个过程大致可以分为两个步骤:首先生成注意力图(Attention Map),然后通过该图反馈空间中每个点的重要性。然而,仅依赖单个点在缺乏高级信息的情况下难以判断周围点的关系,因此目前常用两种方法来捕获广泛的关系依赖点。一种是通过自注意力模型获得全局信息以实现远距离依赖;另一种则是使用足够大的卷积核来建立远距离依赖关系。然而,方法一存在前文提到的缺陷,方法二直接使用大卷积核会导致参数量和计算量显著增加。为了解决这一问题,本算法采用分离卷积的方式,将大卷积核划分为三个方向的分量,进而将其分解为五个部分,包括两个深度空洞卷积、两个深度卷积和一个通道卷积。通过这种分解处理,模型在保持相同依赖关系的情况下,显著降低了

计算成本并减少了参数量。

图 2(b)展示了该算法设计的 LMA 模块。模块以视觉注意力网络<sup>[13]</sup>为主干,对大核进行拆解。具体而言,将  $K \times K$  的卷积核分解为两个分别为  $1 \times (2d - 1)$  和  $(2d - 1) \times 1$  的深度卷积,以及两个分别为  $1 \times \frac{K}{d}$  和  $\frac{K}{d} \times 1$  的深度空洞卷积和一个  $1 \times 1$  的卷积。这样可以在降低大核的计算成本和参数的同时,获得长距离依赖关系,从而实现注意力机制。LMA 模块可以表示为:

$$f_{DW-C} = f_{DW-C2}(f_{DW-C1}(F)) \quad (2)$$



(1)带有LMA模块的整体框架

(2)LMA模块

图 2 LMA 模块及其多层结构

### 1.3 多级特征权重调节机制

根据文献[9]的研究,通过对比高级内容特征和低级质量特征在真实失真图像与合成失真图像中的表现,可以得出结论:对于真实失真的图像,高级内容特征是主要特征;对于合成失真的图像,低级质量特征占据主导地位。

因此,可以通过调整两分支上的权重调节模块,使模型适应不同失真类型的情况。权重计算公式如下:

$$F_c = wf_{AD} + (1 - w)f_{SD} \quad (6)$$

其中,  $w$  为特征权重。当使用真实失真数据集时,权重将增加对真实失真的占比,以确保图像中的高级内容特征占据主要地位。相反,当使用合成失真数据集时,权重将增加对合成失真的占比,以使低级质量特征占据主导,从而提升实验效果。

## 2 实验结果分析

### 2.1 模型参数设置

在高级内容特征分支中,将输入图像随机裁剪为  $224 \times 224$  的尺寸,并通过随机水平翻转和垂直翻转来增强训练样本。对于低级质量特征分支,使用原始尺寸的图像作为输入。所提出的 LMA-BIQA 模型采用 Pytorch 2.1.2 实现,并在 Ubuntu 22.04 操作系统下,使用 NVIDIA A10 GPU 进行训练。模型训练使用

$$f_{DW-D-C} = f_{DW-D-C2}(f_{DW-D-C1}(f_{DW-C})) \quad (3)$$

$$P_{Attention} = f_{1 \times 1}(f_{DW-D-C}) \quad (4)$$

$$y_{Output} = F \otimes P_{Attention} \quad (5)$$

式中,  $F$  表示输入特征,  $P_{Attention}$  表示注意力图,  $\otimes$  代表哈达玛积 (Hadamard Product)。  $f_{1 \times 1}$  表示通道卷积,  $f_{DW-D-C1}$  表示深度空洞卷积,  $f_{DW-C1}$  表示深度卷积。  $y_{Output}$  表示输出。

同时,如图 2(a)所示,算法采用了类似 Swin Transformer<sup>[24]</sup>的层次化结构,通过堆叠  $L$  层 LMA 模块以提取特征,其中  $L$  在每组分别为  $\{3, 3, 12, 3\}$ ,并在每组的最后进行层归一化。

Adam 优化器,权重设置为  $5 \times 10^{-4}$ ,每批次大小为 64,学习率设置为  $2 \times 10^{-5}$ ,训练集与测试集的比例为 8 : 2。通过图 3 的实验结果可看出,在真实失真数据集中,特征权重  $w$  的最佳设置为 0.87,而在合成失真数据集中  $w$  设置为 0.65。

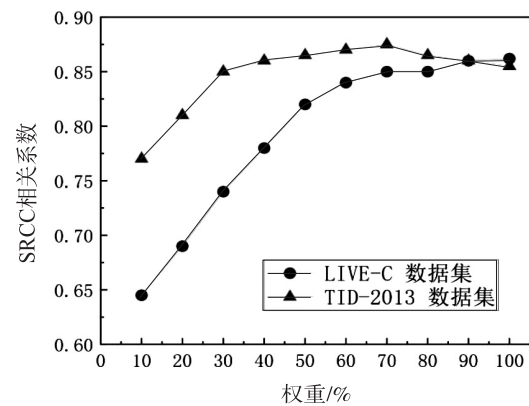


图 3 权重模块参数实验

### 2.2 评价数据集及评价指标

为了更加有效地验证图像评价方法的可靠性,并研究高级内容特征与低级质量特征对不同失真类型的偏好,在 4 个公开的 IQA 数据集上进行了实验。这些数据集包括 3 个合成失真数据集 (LIVE II<sup>[25]</sup>、TID2013<sup>[26]</sup>、CSIQ<sup>[27]</sup>) 以及 1 个真实失真数据集 (LIVE-C<sup>[28]</sup>)。LIVE 数据集总计有 25 幅参考图像和

982 幅失真图像,涵盖白噪声、高斯模糊、锐利衰减、JPEG2000 和 JPEG 等 5 种失真类型,主观分数范围为  $[0,100]$  的 DMOS 值。CSIQ 数据集由 30 个参考图像生成 886 幅失真图像,包含 6 种失真类型。TID2013 数据集包含 3 000 幅图像,有 25 幅参考图像和 24 种失真类型,源自不同的加性高斯噪声和稀疏采样重构等,主观实验评分人数为 931,质量分数用范围为  $[0,9]$  的平均主观分值 (Mean Opinion Score, MOS) 表示。LIVE-C 数据集包含 1 162 幅图像,由 8 100 多名测试者评估,收集到超过 35 万份主观意见得分, MOS 的取值范围为  $[3.42,92.43]$ 。

IQA 算法的评价指标应具备 2 个特性:单调性和准确性。与之相对应的评价指标有 2 个:SRCC (Spearman Rank Order Correlation Coefficient, 斯皮尔曼秩相关系数) 和 PLCC (Pearson Linear Correlation Coefficient, 皮尔森线性相关系数)。SRCC 用来衡量两组数据之间秩的相关性,取值范围为  $[0,1]$ , 当值为 1 时,表示两组数据完全一致。计算公式为:

$$c_{\text{SRCC}} = 1 - \frac{6 \sum_{i=1}^N (x_i - y_i)^2}{N(N^2 - 1)} \quad (7)$$

式中:  $c_{\text{SRCC}}$  表示斯皮尔曼秩相关系数;  $N$  表示图像的数量;  $x_i$  和  $y_i$  表示主观分数和客观分在数据集中的排名。

PLCC 用于描述 MOS 与预测分数之间的线性相

关性,值越接近于 1,则表示算法的性能越高。计算公式为:

$$c_{\text{PLCC}} = \frac{\sum_{i=1}^N (s_i - \bar{s})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^N (s_i - \bar{s})^2} \sqrt{\sum_{i=1}^N (q_i - \bar{q})^2}} \quad (8)$$

式中:  $c_{\text{PLCC}}$  表示皮尔森线性相关系数;  $s_i$  和  $q_i$  分别表示第  $i$  幅图像的主观分数和客观分数;  $\bar{s}$  和  $\bar{q}$  分别表示主观分数和客观分数的平均值。通过这两个评价指标对算法进行评估,来衡量人类评分与系统评分之间的一致性。

### 2.3 实验结果分析

将文中算法与几种具有代表性的 NR-IQA 方法在 3 个合成失真数据集和 1 个真实失真数据集上进行了比较,表中加粗数字表示该组数据最优值,下划线标记数字表示该组数据次优值。从表 1 可以看出, LMA-BIQA 算法在 4 个数据集上在与其他同类方法对比中均取得了较为优异的成绩。加权平均后的斯皮尔曼等级相关系数和皮尔森线性相关系数分别达到 0.928 5 和 0.937 5。这主要得益于其双分支结构,能够提取互补的高级内容特征和低级质量特征。实验结果表明,无论是在人工合成的数据集还是自然失真数据集中,该方法的预测分数与主观分数之间具有较高的相关性,表明 LMA-BIQA 能够很好地反映人类的主观评价。

表 1 不同方法的性能对比

方法	LIVE		CSIQ		TID2013		LIVEC	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
DIIVINE <sup>[29]</sup>	0.892	0.908	0.804	0.776	0.643	0.567	—	—
BRISQUE <sup>[4]</sup>	0.938	0.941	0.750	0.573	0.651	0.829	0.629	0.628
DeepIQA <sup>[30]</sup>	0.955	0.964	0.906	0.911	0.786	0.813	—	—
MetalQA <sup>[31]</sup>	0.960	0.958	0.899	0.908	0.856	0.868	0.835	0.802
TRIQ <sup>[10]</sup>	0.949	0.965	0.825	0.838	0.846	0.858	0.845	0.861
DBCNN <sup>[7]</sup>	0.968	0.971	0.943	0.959	0.816	0.865	0.851	0.869
TReS <sup>[11]</sup>	0.966	0.965	0.923	0.944	0.863	0.884	0.845	0.878
MAM <sup>[32]</sup>	<u>0.975</u>	<u>0.981</u>	<b>0.971</b>	<u>0.962</u>	<b>0.937</b>	<b>0.928</b>	<u>0.857</u>	<u>0.874</u>
Re-IQA <sup>[9]</sup>	0.970	0.971	0.947	0.960	0.844	0.880	0.840	0.854
LMA-BIQA	<b>0.979</b>	<b>0.982</b>	<u>0.956</u>	<b>0.966</b>	<u>0.908</u>	<u>0.919</u>	<b>0.871</b>	<b>0.883</b>

表 2 展示了在交叉数据集上的实验结果,所有预测方法均在某一数据集上训练后,在其他 3 个数据集上进行测试,未进行任何微调或参数调整。实验结果显示, LMA-BIQA 方法在 12 个测试案例中有 6 个取得了最佳结果,其余 6 个也表现出很强的竞争力,表明该方法具有良好的泛化能力。

此外,进一步比较了文中模型与三种 NR-IQA 方

法的复杂度,选取的方法包括基于 CNN 的 ResNet-50、基于 Transformer 的 TReS 以及作为改进目标的 Re-IQA 原代码。在此采用浮点运算数 (GFlops) 和参数量 (Params) 作为运行复杂度的评定标准。

在 IQA 领域常常使用 GFlops (G floating point operations, 浮点数计算量) 和 Params (参数量) 作为指标以衡量模型的计算性能及轻量程度。从表 3 可以看

出,使用 ResNet-50 来取代所提出模型的 LMA 模块,所提出模型的浮点数仅为 ResNet-50 替代模型的一半,而参数数量更是远低于 ResNet-50 模型。与基于 Transformer 的模型 (TReS) 相比,尽管 TReS 方法在测

试中具有较低的参数量,但该文提出的方法在计算复杂度上比其降低了约 35%,并且在性能测试中取得了更好的效果。比起 Re-IQA 模型不仅降低了参数量和计算量,同时也提升了性能。

表 2 交叉数据集实验性能对比 (SRCC)

方法	LIVE			TID2013			CSIQ			LIVEC		
	TID2013	CSIQ	LIVEC	LIVE	CSIQ	LIVEC	LIVE	TID2013	LIVEC	LIVE	TID2013	CSIQ
DIIVINE	0.342	0.602	0.296	0.714	0.583	0.235	0.817	0.417	0.366	0.354	0.327	0.419
BRISQUE	0.354	0.572	0.327	0.724	0.568	0.109	0.823	0.433	0.106	0.244	0.275	0.236
DBCNN	0.536	0.762	0.552	0.852	0.703	0.412	0.833	0.491	0.107	—	—	—
MAM	0.587	0.753	0.559	0.885	<b>0.785</b>	0.319	<b>0.934</b>	<b>0.566</b>	0.437	0.713	0.386	0.589
Re-IQA	0.522	<b>0.766</b>	<b>0.615</b>	0.822	0.721	0.307	0.889	0.523	0.459	<b>0.742</b>	0.403	0.538
LMA-BIQA	<b>0.593</b>	0.759	0.582	<b>0.865</b>	0.783	<b>0.359</b>	0.913	0.552	<b>0.471</b>	0.713	<b>0.411</b>	<b>0.594</b>

表 3 不同方法的复杂度对比

方法	GFlops	Params/M
ResNet50+NSS+权重调节机制	10.22	189.71
TReS	8.37	<b>34.40</b>
MAM	<u>5.45</u>	38.68
Re-IQA	29.53	153.36
LMA-BIQA	<b>5.42</b>	<u>37.46</u>

### 2.4 消融实验

为了分析使用 NSS 方法和权重调节机制的有效

性,进行了消融实验,以验证该模型中各模块所带来的影响。在 LMA 单分支的基础上,构建了以下几种模型:(1)仅包含 LMA 模块的单分支网络;(2)包含 LMA 模块和 NSS 的双分支网络;(3)完整的 LMA-BIQA 模型。表 4 给出了 3 个模型在 LIVE、CSIQ、TID2013 和 LIVEC 数据集上的性能。结果表明对于合成失真图像和真实失真图像,添加 NSS 或权重调节机制的任一模块均能提升性能,这进一步验证了 LMA-BIQA 模型的有效性。

表 4 消融实验结果

方法	LIVE		CSIQ		TID2013		LIVEC	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
LMA 单分支	0.966	0.959	0.935	0.944	0.834	0.861	0.843	0.858
LMA+NSS	0.975	0.976	0.937	0.960	0.873	0.878	0.864	0.875
LMA+NSS+权重调节机制	<b>0.979</b>	<b>0.982</b>	<b>0.956</b>	<b>0.966</b>	<b>0.908</b>	<b>0.919</b>	<b>0.871</b>	<b>0.883</b>

### 3 结束语

针对已有注意力机制相关方法存在计算量过大以及造成的信息丢失等问题,从大核卷积分解多尺度注意力机制、低级图像特征提取及权重管理模块 3 个方面进行了研究,并通过 LMA-BIQA 模型建立特征到质量分数的映射,实现对多种失真图像的质量评价。实验结果表明,LMA-BIQA 模型在与人类主观评价的一致性和所需的算力上明显优于其他相关方法。由于目前在低级特征提取过程中的复杂度较高,下一步将深入研究更高效的特征提取及多尺度特征拟合方式,并且对已有方法进行优化,设立多尺寸的模型,拓展该方法的应用场景。

#### 参考文献:

[1] ZHANG X, XIONG R, FAN X, et al. Compression artifact

reduction by overlapped-block transform coefficient estimation with block similarity [J]. IEEE Transactions on Image Processing, 2013, 22 (12): 4613-4626.

[2] 俞梦婷, 贾惠珍, 王同罕. 基于对抗和梯度的无参考图像质量评价算法 [J]. 现代电子技术, 2023, 46 (11): 60-65.

[3] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 586-595.

[4] 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.

[5] YING Z, NIU H, GUPTA P, et al. From patches to pictures (PaQ-2-PiQ): mapping the perceptual space of picture quality [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 3575-3585.

- [6] MADHUSUDANA P C, BIRKBECK N, WANG Y, et al. Image quality assessment using contrastive learning [J]. *IEEE Transactions on Image Processing*, 2022, 31: 4149–4161.
- [7] ZHANG W, MA K, YAN J, et al. Blind image quality assessment using a deep bilinear convolutional neural network [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 30(1): 36–47.
- [8] DENDI S V R, DEV C, KOTHARI N, et al. Generating image distortion maps using convolutional autoencoders with application to no reference image quality assessment [J]. *IEEE Signal Processing Letters*, 2018, 26(1): 89–93.
- [9] SAHA A, MISHRA S, BOVIK A C. Re-iqa: unsupervised learning for image quality assessment in the wild [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Vancouver: IEEE, 2023: 5846–5855.
- [10] YOU J, KORHONEN J. Transformer for image quality assessment [C]//*2021 IEEE international conference on image processing (ICIP)*. Anchorage: IEEE, 2021: 1389–1393.
- [11] GOLESTANEH S A, DADSETAN S, KITANI K M. No-reference image quality assessment via transformers, relative ranking, and self-consistency [C]//*Proceedings of the IEEE/CVF winter conference on applications of computer vision*. Waikoloa: IEEE, 2022: 1220–1230.
- [12] YU W, LUO M, ZHOU P, et al. Metaformer is actually what you need for vision [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans: IEEE, 2022: 10819–10829.
- [13] GUO M H, LU C Z, LIU Z N, et al. Visual attention network [J]. *Computational Visual Media*, 2023, 9(4): 733–752.
- [14] MEI Z, WANG Y C, HE X, et al. Greenbiqu: a lightweight blind image quality assessment method [C]//*2022 IEEE 24th international workshop on multimedia signal processing (MMSP)*. Bologna: IEEE, 2022: 1–6.
- [15] ZHAO M, CAO G, HUANG X, et al. Hybrid transformer-CNN for real image denoising [J]. *IEEE Signal Processing Letters*, 2022, 29: 1252–1256.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas: IEEE, 2016: 770–778.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston: IEEE, 2015: 1–9.
- [19] DING X, ZHANG Y, GE Y, et al. UniRepLkNet: a universal perception large-kernel convnet for audio video point cloud time-series and image recognition [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seattle: IEEE, 2024: 5513–5524.
- [20] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN [J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [21] LI D, LI L, CHEN Z, et al. Shift-ConvNets: small convolutional kernel with large kernel effects [J]. *arXiv*: 2401.12736, 2024.
- [22] RUDERMAN D, BIALEK W. Statistics of natural images: scaling in the woods [J]. *Advances in Neural Information Processing Systems*, 1994, 6: 517–548.
- [23] QI J, DU J, SINISCALCHI S M, et al. On mean absolute error for deep neural network based vector-to-vector regression [J]. *IEEE Signal Processing Letters*, 2020, 27: 1485–1489.
- [24] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey [J]. *Computational Visual Media*, 2022, 8(3): 331–368.
- [25] SHEIKH H R, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms [J]. *IEEE Transactions on Image Processing*, 2006, 15(11): 3440–3451.
- [26] PONOMARENKO N, JIN L, IEREMEIEV O, et al. Image database TID2013: peculiarities, results and perspectives [J]. *Signal Processing: Image Communication*, 2015, 30: 57–77.
- [27] 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.
- [28] GHADIYARAM D, BOVIK A C. Massive online crowdsourced study of subjective and objective picture quality [J]. *IEEE Transactions on Image Processing*, 2015, 25(1): 372–387.
- [29] 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.
- [30] BOSSE S, MANIRY D, MÜLLER K R, et al. Deep neural networks for no-reference and full-reference image quality assessment [J]. *IEEE Transactions on Image Processing*, 2017, 27(1): 206–219.
- [31] ZHU H, LI L, WU J, et al. MetaQA: deep meta-learning for no-reference image quality assessment [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Virtual: IEEE, 2020: 14143–14152.
- [32] 厉俊阳. 基于视觉感知特性与注意力机制的无参考图像质量评价研究 [D]. 南京: 南京信息工程大学, 2024.