智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

双线性特征融合和门控循环单元质量聚合的图像质量评价

王亚茹, 杨春旺, 屈卓, 赵顺, 张诗吟, 翟永杰

引用本文:
王亚茹, 杨春旺, 屈卓, 等. 双线性特征融合和门控循环单元质量聚合的图像质量评价[J]. 智能系统学报, 2025, 20(4): 946-957.
WANG Yaru, YANG Chunwang, QU Zhuo, et al. Image quality assessment based on bilinear feature fusion and gate recurrent unit quality polymerization[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(4): 946-957.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202407028

DOI: 10.11992/tis.202407028
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20250221.0912.006

双线性特征融合和门控循环单元质量聚合的图像质量评价

王亚茹, 杨春旺, 屈卓, 赵顺, 张诗吟, 翟永杰
(华北电力大学 自动化系, 河北 保定 071003)

摘 要: 目前图像质量评价方法存在特征融合方式简单、质量信息提取和利用不充分以及忽略图像不同区域间相关性的问题, 本文提出双线性特征融合和门控循环单元(gate recurrent unit, GRU)质量聚合的图像质量评价方法。提取图像的全局和局部特征, 并对局部特征进行基于可变形卷积的筛选操作, 在语义和上下文信息的引导作用下, 滤除与失真无关的信息; 构建双线性特征融合模块, 加强全局−局部特征的信息交互, 捕捉图像质量在空间关系和上下文信息上的变化; 构建基于 GRU 的质量聚合模块, 将逐图像块质量预测和全局依赖性建模相结合, 动态调整各图像块的权重比例, 最后通过聚合各图像块的质量信息生成整张图像的质量分数。在不同失真类型、不同场景的 CSIQ、TID2013 和 PIPAL 数据集上, 本文方法的皮尔逊线性相关系数和斯皮尔曼等级相关系数均为最优值, 尤其在 PIPAL 数据集中, 相比于次优方法, 皮尔逊线性相关系数提高了 3.9%, 斯皮尔曼等级相关系数提高了 3.1%。

关键词: 深度学习; 图像质量; 双线性池化; 门控循环单元; 可变形卷积; 特征提取; 特征选择; 特征融合
中图分类号: TP391.4 **文献标志码**: A **文章编号**: 1673−4785(2025)04−0946−12

中文引用格式: 王亚茹, 杨春旺, 屈卓, 等. 双线性特征融合和门控循环单元质量聚合的图像质量评价 [J]. 智能系统学报, 2025, 20(4): 946–957.
英文引用格式: WANG Yaru, YANG Chunwang, QU Zhuo, et al. Image quality assessment based on bilinear feature fusion and gate recurrent unit quality polymerization[J]. CAAI transactions on intelligent systems, 2025, 20(4): 946–957.

Image quality assessment based on bilinear feature fusion and gate recurrent unit quality polymerization

WANG Yaru, YANG Chunwang, QU Zhuo, ZHAO Shun, ZHANG Shiyin, ZHAI Yongjie
(Department of Automation, North China Electric Power University, Baoding 071003, China)

Abstract: Current image quality assessment methods suffer from simple feature fusion strategies, insufficient extraction and utilization of quality information, and neglect of the correlation between different image regions. This paper proposes an image quality assessment method based on bilinear feature fusion and gate recurrent unit (GRU) quality aggregation. We extract global and local features of images and perform selection operations on local features based on deformable convolution. Under the guidance of semantic and contextual information, information unrelated to distortion is filtered out. A bilinear feature fusion module is constructed to enhance the interaction between global and local features, capturing changes in image quality in terms of spatial relationships and contextual information. A quality aggregation module based on GRU is constructed, combining block-wise quality prediction and global dependency modeling. This dynamically adjusts the weight proportion of each image block, ultimately aggregating the quality information of all blocks to generate a quality score for the entire image. For the CSIQ, TID2013, and PIPAL datasets across different distortion types and various scenarios, the proposed method achieved optimal Pearson linear cor-relation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) metrics. Notably, on the PIPAL dataset, the PLCC improved by 3.9% and the SROCC improved by 3.1% compared with the second-best method.

Keywords: deep learning; image quality; bilinear pooling; gate recurrent unit; deformable convolution; feature extraction; feature selection; features fusion

随着信息时代的不断发展, 每天都会产生数以亿计的图片。这些图片不仅在日常网络交流中扮演着重要角色, 也是图像处理研究不可或缺的数据来源。因此, 图片质量逐渐成为一个不容忽视的关键因素。图像的质量对于信息传递的完整

收稿日期: 2024−07−24. **网络出版日期**: 2025−02−21.
基金项目: 国家自然科学基金青年科学基金项目 (62303184); 国家自然科学基金联合基金项目重点支持项目 (U21A20486); 国家自然科学基金面上项目 (62373151); 河北省自然科学基金青年科学基金项目 (2024502006); 中央高校基本科研业务费专项 (2023JC006, 2024MS136, 2024MS138).
通信作者: 张诗吟. E-mail: shiyinzhang@ncepu.edu.cn.

性和准确性至关重要, 然而, 图像在获取、压缩、传输和显示的过程中, 往往会出现各种形式的失真[1]。因此, 图像质量评价(image quality assessment, IQA)[2]在许多图像处理任务中变得至关重要。

图像质量评价旨在使用计算机模拟人类视觉系统, 实现对图像满足人类需求的质量进行评价。尽管人类使用肉眼对图像质量进行评价不存在困难, 但是对于机器来说, 完成指定要求的图像质量评价有着很大的困难。目前, IQA 可划分为两大类: 主观评价方法和客观评价方法[3]。主观评价方法是基于人类视觉系统对图像的感知实现对图像的质量评价。这些主观评价方法能够将图像直观表现出的质量反馈出来, 并且不需要任何其他的技术即可完成。但是, 这些方法存在着不可忽视的缺陷。比如, 为确保结果的稳定性, 需要进行多次重复实验, 这增加了实验成本, 并且实时质量评价难以实现[4]。主观评价方法同时受到了不同观察评价者之间存在的差异以及不确定性的影响, 例如观察者的观察目的、知识背景以及观察环境。与主观评价方法不同, 客观评价方法不完全依靠人类感官, 而是依靠数学模型计算得来, 是基于人类视觉系统对主观评价方法的模拟。相比于主观评价方法, 客观评价方法[5]可以减小主观意识上的影响, 能够更快地完成对图像的评价处理, 具有批量处理的能力和可重复性。客观评价方法根据在评价图像质量时对参考图像的需要程度分为 3 类: 全参考图像质量评价(full reference, FR)[6]、半参考图像质量评价(reduced-reference, RR)[7]和无参考图像质量评价(no-reference, NR)[8]。

FR-IQA 通过失真图像和参考图像对比做差[9], 获取两者之间的差异信息, 通过对差异信息的分析处理获得失真图像的质量, 其评价结果更符合人类视觉系统。传统的 FR-IQA 方法, 如均方误差(mean squared error, MSE)和峰值信噪比(peak signal-to-noise ratio, PSNR)[10], 虽然被广泛采用, 但它们仅通过像素级别的比较来量化图像质量, 忽略了相邻像素之间的相关性以及人类视觉系统的非线性特性。为了更准确地反映人类感知, 文献 [11] 提出了结构相似性指数(structure similarity index measure, SSIM), 综合考虑图像的强度、对比度和结构等方面, 模拟人类视觉系统对图像结构的敏感性。PSNR 的局限性激发了后续研究的发展, 在 SSIM 的基础上先后提出了多尺度结构相似性指数(multi-scale SSIM, MS-SSIM)[12]、梯度幅值相似性偏差(gradient magnitude similarity deviation, GMSD)[13]、视觉显著性指数(visual saliency index,VSI)[14]等方法, 通过采用更加复杂的特征提取方式和分析方法, 提高了图像质量评价的准确性。然而, 这些方法依旧依赖于手工提取的特征, 这使得对图像失真的全面表达有着较大局限性, 无法更好地符合人类主观感受。因此, 研究者们提出了基于学习的方法以克服手工提取特征的不足, 如 DeepQA[15]、PieAPP[16]、LPIPS-VGG[17]、DISTS[18]、JND-SalCAR[19]等方法。但是这些方法采用传统的学习方法, 对特征的提取能力有限, 特征过于低级往往难以充分表达复杂的图像失真。

随着计算机硬件及人工智能技术的发展, 出现了许多基于深度学习的 IQA 方法。文献 [20] 提出了深度相似性, 通过深度神经网络提取图像失真特征, 计算参考图像和待评估图像的特征相似性, 将特征相似性输入到一个回归网络, 预测图像的质量评分。分级退化级联卷积神经网络[21]考虑了人类视觉系统中分层感知机制, 利用卷积神经网络(convolutional neural network, CNN)学习退化特征, 实现质量的预测。WaDIQaM-FR 方法[22]通过堆叠多个卷积层和池化层自动提取与畸变有关的特征。此外, Cheon 等[23]提出了 IQT 方法, 利用 Transformer 网络作为特征提取器, 获得包含更多语义信息和图像长距离关系的特征。深度学习网络增强了对图像特征的学习能力, 相对于传统学习方式, 提高了图像质量评价性能。但以上方法通常只能处理失真图像的全局特征, 而无法注意到局部特征。而图像失真往往发生在局部区域, 并且人类视觉系统对局部失真具有较高的敏感性, 更容易关注到局部区域的失真[24], 因此图像局部特征对于图像质量评价具有重要作用[25]。Lao 等[26]提出了基于注意力的混合质量评价方法 (attention-based hybrid image quality assessment, AHIQ), 除采用 Vison Transformer(Vit)进行全局特征提取外, 同时使用 CNN 网络提取图像中的局部特征, 对全局特征进行补充, 进一步提高了模型的图像质量评价性能。然而, 该方法将局部特征和全局特征简单地拼接或堆叠在一起, 可能会导致信息冗余和特征冲突, 并不能充分利用特征中的图像质量信息。

此外, 文献 [22] 在采用深度学习网络提取图像特征的基础上, 将图像分成若干图像块, 对图像质量进行分块评价, 最后进行加权求和得到整幅图像的质量分数。类似地, 文献 [27] 提出的方法, 同样通过逐块评价方法分析图像中局部区域的

特征和失真，提高了图像质量评价的准确性和鲁棒性。这种逐图像块评价的方法可识别图像不同区域的显著差异，并且更加适应不同分辨率和场景的失真图像。但现有方法大多利用卷积网络计算各图像块的权重，通过全连接层回归质量分数，无法有效捕捉图像块序列中的动态变化和序列相关性，不能根据图像内容自适应调整，对一些复杂的空间变化处理能力有限，且会造成信息损失。

对于上述问题，本文提出了一种双线性特征融合和 GRU 质量聚合（bilinear feature fusion and GRU quality aggregation, BFFGQA）的图像质量评价方法，主要贡献如下：

1）对局部特征进行基于可变形卷积的筛选操作，采用包含语义信息和上下文信息的全局特征进行可变形卷积偏移量学习，使其更好地适应图像中的空间变化和局部细节，引导卷积特征关注图像的失真区域，滤除与失真无关的信息。

2）提出双线性全局−局部特征融合（bilinear global-local feature fusion, BFF）模块，对失真−参考图像对的全局特征和局部特征进行双线性池化，使两种特征之间的信息交互更加紧密，更好地捕捉图像质量在空间关系和上下文信息上的变化，有助于消除冗余信息和减少特征冲突。

3）提出基于门控循环单元的质量聚合（quality aggregation based on GRU, GRU-QA）模块，将逐图像块质量预测和全局依赖性建模相结合，建立各图像块之间的联系，综合考虑图像特征的空间维度长距离和短距离特征关系上的变化，动态调整各像素块的权重比例，增强模型的灵活性和适应性。

4）在 4 个通用基准的图像质量评价数据集上进行对比实验，本文方法获得具有显著优势的评价结果。

1　相关技术和理论

1.1　Vision Transformer

Transformer[28] 起初是一个为了解决自然语言处理任务的经典模型。Image Transformer 首次将 Transformer 应用在图像处理任务中，算法中加入的自注意力机制[29] 使得模型在处理图像时关注图像的全局信息，在各局部特征间建立依赖关系，充分利用特征中的上下文关系。Vison Transformer[30] 的出现使 Transformer 更加适用于图像任务。Vit 模型将图像分成若干图像块转化成 Transformer 能够处理的序列数据，并使用自注意力机制，能够同时捕捉图像的全局结构和局部细节，从而提高图像质量评估的准确性。其自动特征提取和灵活的模型架构使其能够适应不同的失真类型和分辨率，进一步增强了评估效果。

1.2　可变形卷积

可变形卷积[31] 是卷积神经网络的一种改进形式，引入了可学习的偏移量参数，使网络可以自适应图像中的空间变化，更好地捕捉到图像中的形变和位置关系，提高了卷积神经网络的鲁棒性和空间变换建模能力。可变形卷积通常被应用在目标检测、图像分割等领域，在视觉任务中发挥出强大的性能。鉴于此，Shi 等[32] 提出了区域自适应变形网络（region-adaptive deformable network, RADN），将可变形卷积引入到图像质量评价中，充分利用其对空间变形的自适应能力来提升评价精度。然而，RADN 仅利用局部特征学习可变形卷积中的偏移量，虽然在一定程度上提升了图像质量评价性能，但仅依赖局部特征的学习方法可能会对局部噪声或异常敏感，影响评价的鲁棒性。

1.3　双线性池化

双线性池化[33] 是一种高级特征聚合方法，用于从图像或其他输入数据中提取复杂的特征交互。与传统的平均池化或最大池化不同，双线性池化捕捉输入特征之间的二阶统计信息，它通过对两个输入张量计算外积来捕获它们之间的相关性信息并进行池化操作进行汇总。在处理图像时，双线性池化可以更好地保留图像的全局特征，帮助模型更好地理解图像中不同位置像素之间的关系，从而提高图像处理任务的性能[34]。在许多视觉任务中，如图像分类、目标检测和图像分割等，双线性池化已被广泛应用，并在提高模型性能方面取得了显著的成果。

1.4　门控循环单元

门控循环单元（gated recurrent unit, GRU）[35] 是循环神经网络结构的一种改进形式，和长短时记忆网络 (long-short term memory, LSTM) 一样，也是为了解决长期记忆和反向传播中的梯度问题提出来的。相比于 LSTM 和 Transformer 网络，其结构更加简单，参数量更少，但有着同样优秀的学习长期依赖信息的能力。GRU 由更新门和重置门两部分组成。其中更新门用于控制当前时间步骤的候选隐状态（candidate hidden state）对最终的隐状态更新的贡献程度，可以更好地获取序列数据中的长期依赖性的信息。更新门的输出接近于 0 表示较少地更新当前时间步骤的隐状态，而接近于 1 表示较多地更新当前时间步骤的隐状态。而重置门用于控制前一个隐状态在当前时间

步骤的计算中保留多少历史信息，可以更好地捕捉序列数据中的短期相关信息。重置门的输出接近于 0 表示较多地忽略前一个隐状态的信息，而接近于 1 表示较多地保留前一个隐状态的信息。

2　本文方法

本文以文献 [26] 的方法为基线模型，构建双线性全局−局部特征融合（BFF）模块和基于 GRU 的质量聚合（GRU-QA）模块。提出的 BFFGQA 图像质量评价方法的网络框架如图 1 所示。通过基于 CNN 和 Vit 的双分支特征提取模块，分别从参考图像和失真图像中提取特征；然后通过双线性全局−局部特征融合模块对全局−局部特征进行双线性池化融合；最后由基于 GRU 的质量聚合模块进行图像质量评价，输出失真图像的质量分数。

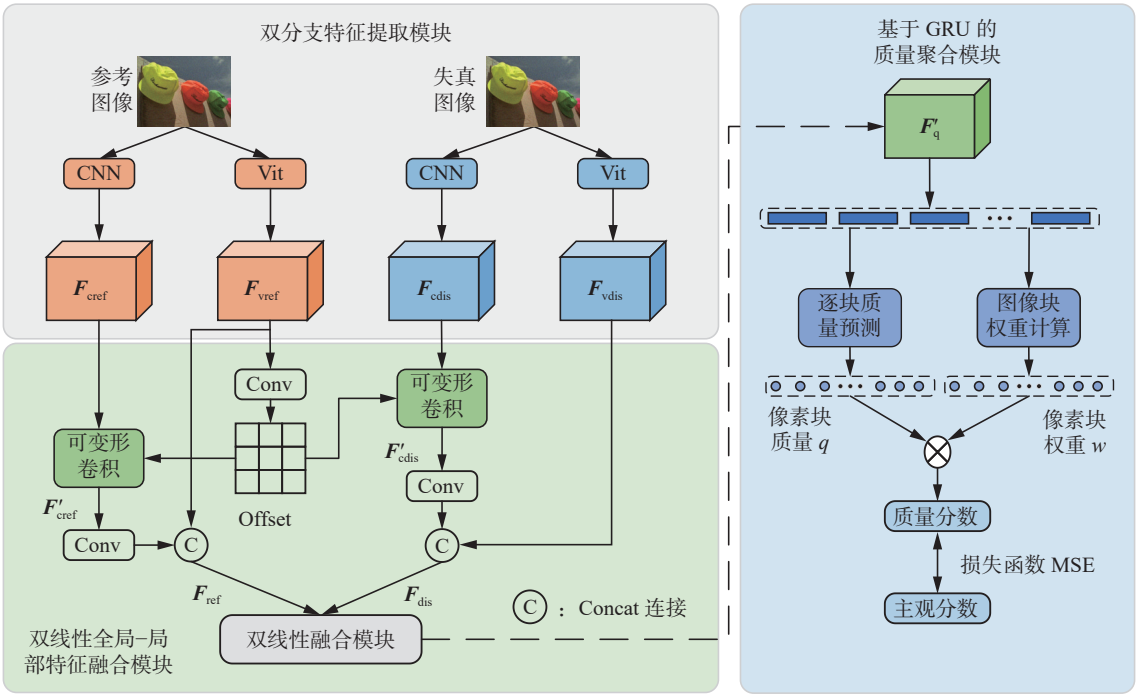


图 1　BFFGQA 网络框架

Fig. 1　BFFGQA network framework

2.1　双分支特征提取模块

该模块设计双分支结构，对输入的失真和参考图像对进行全局和局部特征提取。

全局特征提取器采用 Vit 网络，将原 Vit 输出特征图进行重塑，剔除分类编号，保留图像的全局特征，结构如图 2 所示。利用自注意力机制在全局范围内建立图像中各个图像块之间的依赖关系信息，捕捉全局语义特征。这种全局关注机制有助于提取图像中的长距离依赖性，从而更好地理解图像质量信息。失真和参考图像对经 Vit 网络后，分别获得失真全局特征 $\boldsymbol{F}_{\mathrm{vdis}}$ 和参考全局特征 $\boldsymbol{F}_{\mathrm{vref}}$。

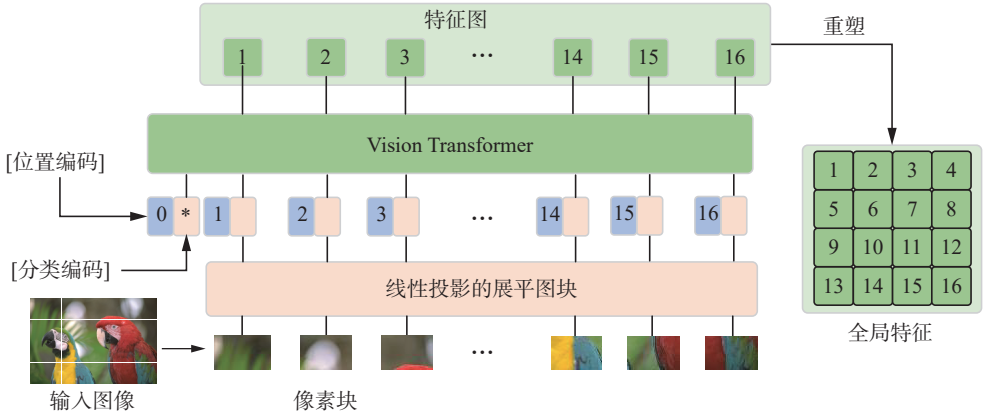


图 2　Vit 特征提取器结构

Fig. 2　Architecture of Vit feature extractor

由于在训练过程中通过多层的自注意力机制来学习不同尺度的特征表示，Vit 能够有效地处理不同大小和分辨率的图像，但对局部图像质量的特征关注较少。而 CNN 网络在捕捉局部特征和细节方面具有优势，能够对图像质量的细微差异进行更精细的分析。因此，采用 ResNet50 网络作为局部特征提取器，移除 ResNet50 最后的平均池化层和完全连接层，识别图像局部结构和纹理特征。失真和参考图像对经 ResNet50 网络后，获得包含空间信息的局部特征 $\boldsymbol{F}_{\mathrm{cdis}}$ 和 $\boldsymbol{F}_{\mathrm{cref}}$，对全局特征进行补充，获得更多的图像质量信息。

2.2　双线性全局−局部特征融合模块

该模块能够在保留图像局部和全局特征的基础上，通过细粒度的信息融合，使模型更加关注图像纹理、噪声等信息，提高模型对图像质量的敏感度。

由于对图像进行逐图像块处理，ResNet50 网络在提取图像局部特征的同时会带来与失真不相关或冗余的信息，对图像质量评价造成干扰。采用可变形卷积对 $\boldsymbol{F}_{\mathrm{cdis}}$ 和 $\boldsymbol{F}_{\mathrm{cref}}$ 自适应地调整感受野，基于特征对全局信息的贡献度以及与参考特征的相似性，筛除掉不相关和冗余的信息，从而更加精确地捕捉到重要特征 $\boldsymbol{F}'_{\mathrm{cdis}}$ 和 $\boldsymbol{F}'_{\mathrm{cref}}$。可变形卷积使用偏移量 Offset 进行动态采样，可对图像不同区域的特征响应进行加权。偏移量 Offset 通过 Vit 提取的参考图像全局特征 $\boldsymbol{F}_{\mathrm{vref}}$ 自适应学习得到，用于指导采样位置的偏移值。$\boldsymbol{F}_{\mathrm{vref}}$ 中包含了关于图像整体结构和关键区域的信息，能够反映出哪些区域对整张图像的理解和表达较为重要。可变形卷积通过 Offset 对卷积核采样位置的调整，使其不再局限于标准的网格采样，而是能够移动到与 $\boldsymbol{F}_{\mathrm{vref}}$ 包含的全局上下文一致的关键区域，同时忽略不相关或冗余区域。

然后采用卷积网络分别对 $\boldsymbol{F}'_{\mathrm{cdis}}$ 和 $\boldsymbol{F}_{\mathrm{vdis}}$、$\boldsymbol{F}'_{\mathrm{cref}}$ 和 $\boldsymbol{F}_{\mathrm{vref}}$ 进行特征对齐处理，并经特征融合后分别得到包括全局和局部信息的失真特征 $\boldsymbol{F}_{\mathrm{dis}}$ 和参考特征 $\boldsymbol{F}_{\mathrm{ref}}$。接下来对 $\boldsymbol{F}_{\mathrm{dis}}$ 和 $\boldsymbol{F}_{\mathrm{ref}}$ 进行双线性融合，具体的操作流程如图 3 所示，不仅可以捕捉失真特征与参考特征之间的细微差异，还可以增强不同通道间的交互信息。

首先采用卷积计算分别对输入特征 $\boldsymbol{F}_{\mathrm{dis}}$ 和 $\boldsymbol{F}_{\mathrm{ref}}$ 进行压缩，在保留质量信息的基础上减少计算量。然后对压缩后特征 $\boldsymbol{F}'_{\mathrm{dis}}$ 和 $\boldsymbol{F}'_{\mathrm{ref}}$ 进行差值运算，得到差值特征 $\boldsymbol{F}_{\mathrm{d}}$，以突出失真特征与参考特征之间的差异，使模型能够更敏锐地感知输入图像失真区域的变化。再通过双线性池化点乘运算对 $\boldsymbol{F}_{\mathrm{d}}$ 和 $\boldsymbol{F}'_{\mathrm{dis}}$ 进行进一步整合和提炼，得到融合后的特征向量 $\boldsymbol{F}_{\mathrm{q}}$，以实现跨特征通道的交互，增强特征表示的丰富度，计算公式为

$$\boldsymbol{F}_{\mathrm{q}} = \mathrm{AAP}\left(\boldsymbol{F}_{\mathrm{d}}^{\mathrm{T}} \otimes \boldsymbol{F}'_{\mathrm{dis}}\right)$$

式中：⊗为外积计算，AAP(·) 表示对特征张量进行自适应平均池化处理。

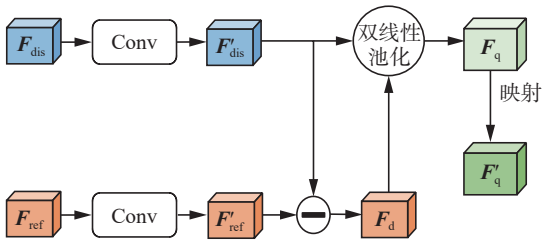


图 3　双线性特征融合操作

Fig. 3　Bilinear feature fusion operations

$\boldsymbol{F}_{\mathrm{q}}$ 具有较强的表达能力，可为后续的图像质量回归提供丰富和准确的特征表示，但它是黎曼流形，为了简化后续的计算和处理，将 $\boldsymbol{F}_{\mathrm{q}}$ 映射到欧氏空间，记为 $\boldsymbol{F}'_{\mathrm{q}}$，转换公式为

$$\boldsymbol{F}'_{\mathrm{q}} = \frac{\mathrm{sign}(\boldsymbol{F}_{\mathrm{q}}) \odot \sqrt{|\boldsymbol{F}_{\mathrm{q}}|}}{\left\| \mathrm{sign}(\boldsymbol{F}_{\mathrm{q}}) \odot \sqrt{|\boldsymbol{F}_{\mathrm{q}}|} \right\|}$$

式中⊙表示逐元素相乘。

2.3　基于 GRU 的质量聚合模块

该模块通过两个分支一方面提取每个图像块的质量分数，一方面根据图像内容的变化动态调整每个图像块的权重，最后将图像块的质量分数聚合得到整张失真图像的质量分数，如图 4 所示。

GRU 质量聚合模块逐块质量预测分支使用 3 个全连接（FC）层对图像中每个图像块的特征进行逐块处理，对图像块质量进行独立预测，输出其对应的质量分数向量 $\boldsymbol{q}$。图像块权重计算分支由 FC 层、GRU 层和 FC 层组成，通过对所有图像块之间的时序关系和长期依赖性进行建模和提取，使网络理解图像块之间的相互影响和全局联系，识别出在图像质量评价中较重要的图像块，并赋予其较大的权重，更准确地反映图像的整体质量；对语义信息含量较少的图像块则赋予较低权重，减少其对最终质量分数的影响。最终该分支输出各图像块的权重向量 $\boldsymbol{w}$。首先由 FC 层将展平的图像质量特征 $\boldsymbol{F}'_{\mathrm{q}}$ 转换为适合 GRU 处理的序列数据，即多个图像块的特征，记为 $\boldsymbol{x}$，并输入到 GRU 层。对 $\boldsymbol{x}$ 进行处理：

$$\boldsymbol{z}_i = \sigma(\boldsymbol{W}_z \cdot [\boldsymbol{h}_{i-1}, \boldsymbol{x}_i])$$

$$\boldsymbol{r}_i = \sigma(\boldsymbol{W}_r \cdot [\boldsymbol{h}_{i-1}, \boldsymbol{x}_i])$$

$$\boldsymbol{h}'_i = \tanh(\boldsymbol{W}_z \cdot [\boldsymbol{h}_{i-1} \cdot \boldsymbol{r}_i, \boldsymbol{x}_i])$$

$$\boldsymbol{h}_i = (1 - \boldsymbol{z}_i)\boldsymbol{h}_{i-1} + \boldsymbol{z}_i \boldsymbol{h}'_i$$

式中: z_i 是更新门; r_i 是重置门; x_i 是第 i 个图像块的特征; h_i 是第 i 个图像块的初始权重, i=1, 2, …, N_p、N_p 为图像块的数量; σ 是 Sigmoid 函数; W、W_z、W_r 分别是候选隐状态、更新门和重置门的权重矩阵。最终, GRU 层输出所有图像块的初始权重序列, 记为 h。h 经过后续 FC 层的归一化处理, 预测出所有图像块的最终注意力权重向量 w, 动态调整各个图像块在整张图像质量分数中的重要性。归一化公式为

$$w_i = \frac{\alpha_i}{\sum_{i=1}^{N_p} \alpha_i}$$

式中: i 是每个图像块的索引, α_i 和 w_i 分别是第 i 个图像块的预测注意力权重和归一化注意力权重。

将图像块质量分数向量 q 和图像块权重向量 w 进行点积运算, 得到整张失真图像的质量 Q, 计算公式为

$$Q = \sum_{i=1}^{N_p} w_i q_i = \frac{\sum_{i=1}^{N_p} a_i q_i}{\sum_{i=1}^{N_p} a_i}$$

其中 q_i 是第 i 个图像块的质量预测分数。

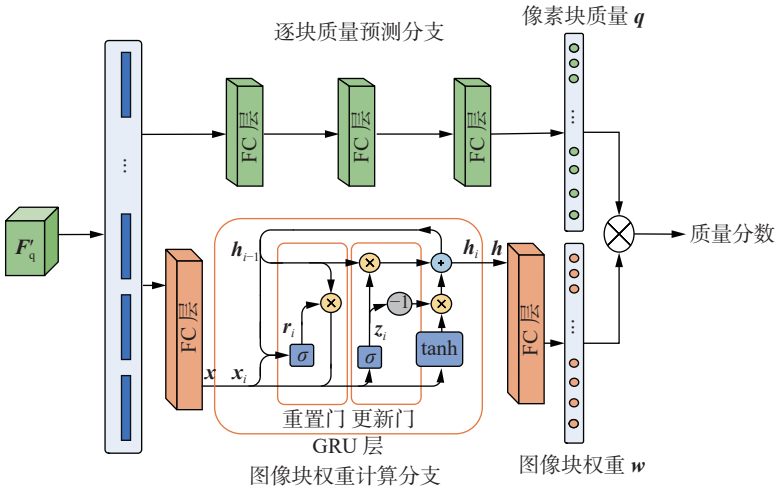


图 4　GRU 质量聚合模块

Fig. 4　Quality aggregation based on GRU

3　实验与结果分析

3.1　数据集

为了验证本文所提方法的性能, 分别在 4 个常用的公共数据集上进行实验, 包括 LIVE、CSIQ、TID2013 和 PIPAL[36], 详细信息见表 1。

表 1　数据集信息

Table 1　Statistics information of datasets

数据集名称	参考图像数	失真图像数	失真类型数
LIVE	29	982	5
CSIQ	30	866	6
TID2013	25	3 000	25
PIPAL	81	10 125	40

LIVE 数据集包含 982 张失真图像和 29 张参考图像, 具有 5 种失真类型, 每张图像质量评级的差分平均意见分数(differential mean opinion score, DMO)取值范围在 [0,100], 较高的 DMO 意味着较低的质量。CSIQ 数据集包含 866 张失真图像和 30 张参考图像, 具有 6 种失真类型, 具有 5 000 个 DMOs 评价数据, 取值范围为 [0,1]。TID2013 数据集包含 3 000 张失真图像和 25 张参考图像, 具有 25 种失真类型, 平均意见分数(mean opinion score, MOS)取值范围为 [0, 9], MOS 越大表示视觉质量越好。PIPAL 数据集包含了 10 125 张失真图像和 81 张参考图像, 具有 40 种失真类型, 其中还包括了生成网络所生成的图像中存在的失真。

3.2　实验参数

本文在 NVIDIA 3 090 GPU 上进行实验, 采用 PyTorch 架构搭建模型。按照 6∶2∶2 的比例将数据集中的失真图像划分为训练集、验证集和测试集。将每张训练集图像进行归一化和随机水平翻转, 并随机裁剪为 224×224 的大小, 对模型进行训练。使用验证集来调整网络参数使其性能最佳, 使用测试集测试网络的图像质量评价性能。在本文方法的双分支特征提取模块中, CNN 特征提取分支采用在 Imagenet 上预训练好的 Resnet50; Vit 特征提取分支采用在 Imagenet 中预训练好的 VIT-B-8 作为主干网络。采用均方误差损失函数

(MSE)计算模型损失，使用 AdamW 优化器，初始学习率设置为 10^{-4}，权重衰减率设置为 10^{-5}。使用余弦退火调度设置每个参数组的学习率，其中 $\eta_{\max}$ 设置为初始学习率，T_{cur} 的周期数设为 50。模型训练过程迭代次数设置为 200。

3.3　评价指标

采用目前常用的两个评价指标对模型性能进行评价：皮尔逊线性相关系数 (Pearson linear correlation coefficient, PLCC) 和斯皮尔曼等级相关系数 (Spearman rank-order correlation coefficient, SROCC)。

PLCC 通过计算模型的评价结果和主观评价结果之间的相对距离来衡量二者在非线性回归后的线性相关性，PLCC 的值越接近 1，代表二者之间的线性相关性越强，模型性能越好，计算公式为

$$P_{\text{LCC}}=\frac{\sum_{i=1}^{N}(s_i-\overline{s})(p_i-\overline{p})}{\sqrt{\sum_{i=1}^{N}(s_i-\overline{s})^2\sum_{i=1}^{N}(p_i-\overline{p})^2}}$$

式中：s_i 和 p_i 分别表示第 i 个图像的模型预测分数和主观评价分数，$\overline{s}$ 和 $\overline{p}$ 分别表示它们的平均值，N 代表测试图像的数量。

SROCC 用于衡量模型的评价结果和主观评价结果之间的相关性。SROCC 的值越相近于 1 表示二者的正相关性越高，模型性能越好，计算公式为

$$S_{\text{ROCC}}=1-\frac{6\sum_{i=1}^{N}d_i^2}{N(N^2-1)}$$

式中：d_i 代表第 i 个测试图像的模型预测分数和主观评价分数的等级差异。

3.4　消融实验

3.4.1　不同模块的消融

为了验证提出的 BFF 模块和 GRU-QA 模块的有效性，作者在 CSIQ 数据集上进行消融实验，实验结果如表 2 所示。

表 2　不同模块的消融实验结果

Table 2　Ablation experimental results of different modules

AHIQ	BFF	GRU-QA	PLCC	SROCC
√			0.978	0.975
√	√		0.985	0.981
√		√	0.986	0.980
√	√	√	**0.988**	**0.982**

注：加黑为本列中最好效果。

AHIQ 模型添加 BFF 模块后，PLCC 提高了 0.72%，SROCC 提高了 0.62%；添加 GRU-QA 模块后，PLCC 提高 0.82%，SROCC 提高了 0.5%；同时添加 BFF 模块和 GRU-QA 模块，模型性能达到最优，PLCC 提高了 1.0%，SROCC 提高了 0.72%。实验结果表明，BFF 和 GRU-QA 模块均可有效提升模型的图像质量评价性能，两个模块同时使用，其作用可相互促进。

3.4.2　不同特征融合方式的消融

为了进一步验证本文提出的双线性特征融合方式的优势，将其与多种常见的特征融合方式进行对比，包括拼接、做差、求和以及相乘。在 CSIQ 数据集上进行实验，实验结果见表 3。

表 3　不同特征融合方式对比

Table 3　Comparison of different feature fusion methods

特征融合方式	PLCC	SROCC
拼接	0.981	0.978
做差	0.987	0.980
求和	0.985	0.973
乘法	0.980	0.970
双线性融合	**0.988**	**0.982**

双线性融合方式通过将全局特征和局部特征进行交互融合，更好地捕捉图像质量在空间关系和上下文信息上的变化，使模型的图像质量评价性能最优，较其他特征融合方式，PLCC 至少提高 0.1%，SROCC 至少提高 0.2%。

3.4.3　特征重要性的消融

为了评估不同特征在图像质量评价中的贡献，本文分别进行了全局特征、局部特征及二者融合的消融实验，实验结果如表 4 所示。

表 4　不同特征消融实验的性能对比

Table 4　Performance comparison of different feature ablation experiments

特征类型	PLCC	SROCC
global	0.886	0.861
local	0.864	0.856
global + local	**0.988**	**0.982**

单独使用全局特征时，PLCC 和 SROCC 分别达到 0.886 和 0.861，明显高于单独使用局部特征 (PLCC 为 0.864，SROCC 为 0.856)，表明亮度分布、对比度等全局统计特征在整体图像质量评价中起到了基础作用。全局特征具有较强的稳定性，能够较好地反映低频失真的影响，如模糊、对

比度下降等退化现象。然而, 在椒盐噪声等高频失真场景下, 全局特征的表现受到一定限制。相比之下, 局部特征能够有效识别噪声点的空间分布特性, 因此在某些特定场景下(如高频噪声、纹理丢失)贡献更大。

当全局特征与局部特征融合后, PLCC 和 SROCC 分别达到 0.988 和 0.982, 相较于单一特征均有显著提升(PLCC 提升 10.3% ~ 14.4%, SROCC 提升 12.3% ~ 14.7%)。这一结果证明, 全局特征和局部特征在图像质量评价中存在互补作用, 共同提升了模型的预测精度。

3.5　对比试验

在表 1 所示的 4 个基准数据集上进行对比实验。对比方法包含 18 种先进方法, 实验结果如表 5 所示, 其中对比方法的指标数据采用相应参考文献中的数据。

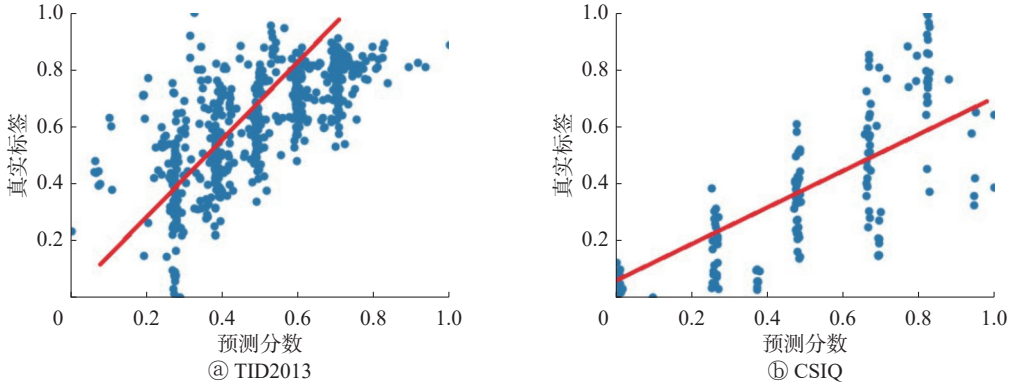
表 5　在公共数据集上的对比实验结果

Table 5　Performance comparison results on public datasets

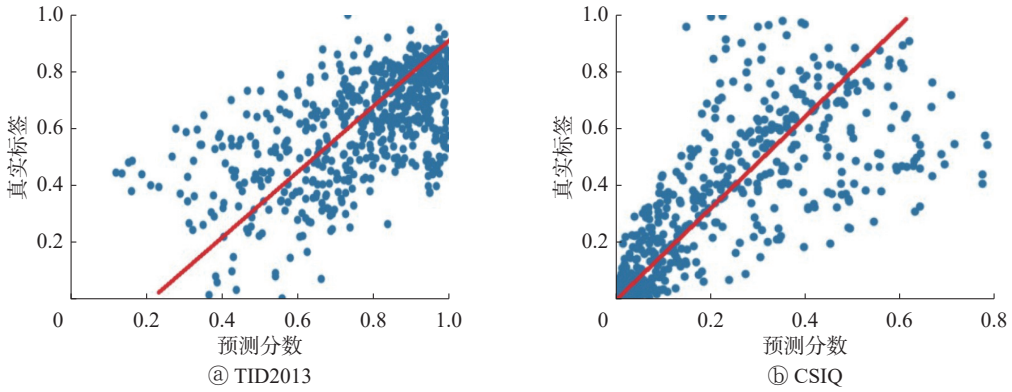
算法	LIVE		CSIQ		TID2013		PIPAL	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
PSNR[10]	0.865	0.873	0.819	0.810	0.677	0.687	0.277	0.249
SSIM[11]	0.937	0.948	0.852	0.865	0.777	0.727	0.391	0.361
MS-SSIM[12]	0.940	0.951	0.889	0.906	0.830	0.786	0.163	0.369
GMSD[13]	0.957	0.960	0.945	0.950	0.855	0.804	0.608	0.537
VSI[14]	0.948	0.952	0.928	0.942	0.900	0.897	0.517	0.458
NLPD[37]	0.932	0.937	0.923	0.932	0.839	0.800	0.401	0.355
MAD[38]	0.968	0.967	0.950	0.947	0.827	0.781	0.580	0.543
VIF[39]	0.960	0.964	0.913	0.911	0.771	0.677	0.479	0.397
FSIMc[40]	0.961	0.965	0.919	0.931	0.877	0.851	—	—
DeepQA[15]	0.982	0.981	0.965	0.961	0.947	0.939	—	—
WaDIQaM-FR[16]	0.980	0.970	—	—	0.946	0.940	0.548	0.553
PieAPP[17]	0.986	0.977	0.975	0.973	0.946	0.945	0.597	0.607
LPIPS-VGG[18]	0.978	0.972	0.970	0.967	0.944	0.936	0.633	0.595
DISTS[19]	0.980	0.975	0.973	0.965	0.947	0.943	0.687	0.655
JND-SalCAR[21]	0.987	0.984	0.977	0.976	0.956	0.949	—	—
IQT[23]	—	—	—	—	—	—	0.790	0.799
AHIQ[26]	**0.989**	**0.984**	0.978	0.975	0.968	0.962	0.823	0.813
TOPIQ[41]	0.989	0.984	0.980	0.978	0.958	0.954	0.830	0.813
本文BFFGQA	0.987	0.980	**0.988**	**0.982**	**0.972**	**0.965**	**0.863**	**0.838**

在 LIVE 失真图像数据集上, 本文方法仅次于 AHIQ, 获得次优的评价结果; 在 CSIQ、TID2013 和 PIPAL 等 3 个失真图像数据集上, 本文方法获得最优的评价结果。具体地, 在 LIVE 数据集上, 与 AHIQ 相比, 本文方法的 PLCC 仅落后 0.2%, SROCC 落后 0.4%; 相比于其他方法, 本文方法的 PLCC 至少提高了 0.9%, SROCC 至少提高 0.3%。在 CSIQ 数据集上, 相比于性能排名第 2 的方法, 本文方法的 PLCC 提高 0.8%, SROCC 提高了 0.4%; 在 TID2013 数据集上, 相比于性能第 2 的模型, 本文方法的 PLCC 提高了 0.4%, SROCC 提高了 0.3%; 在 PIPAL 数据集上, 相比于性能第 2 的模型, 本文方法的 PLCC 提高了 3.9%, SROCC 提高了 3.1%。以上实验结果表明, 本文方法可对多种场景、多种失真进行有效识别, 其图像质量评价性能具有明显优势。

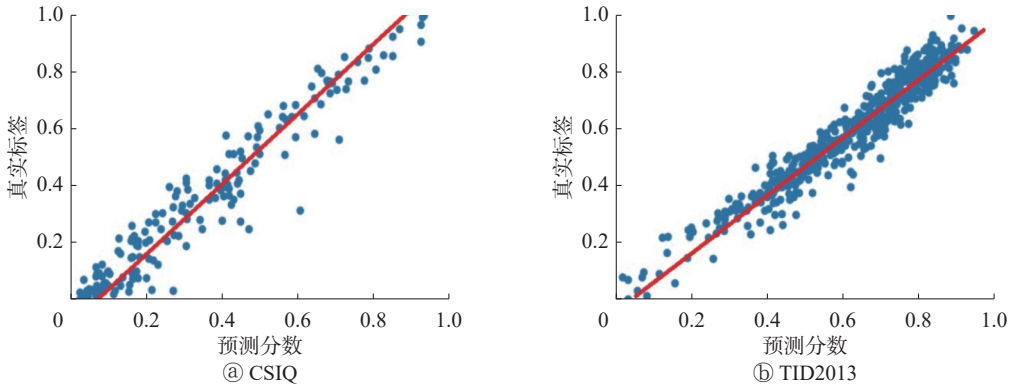
对 TID2013 和 CSIQ 数据集上的 PSNR、SSMI、AHIQ 和 BFFGQA 等 4 种方法的图像评价得分进行可视化, 如图 5 所示。其中, 横坐标是评价方法对失真图像的评价得分数值, 纵坐标是失真图像数据集中自带的主观评价分数值(真实标签)。红色直线表示理想评价结果, 即评价方法的评价结果与真实标签完全一致。将测试样本的评价分数和真实标签以散点的形式绘于图中。通过散点分布可以更加直观地显示出评价方法的预测分数和真实标签之间的关系, 方法的评价性能越好, 散点分布就越集中于红色直线。



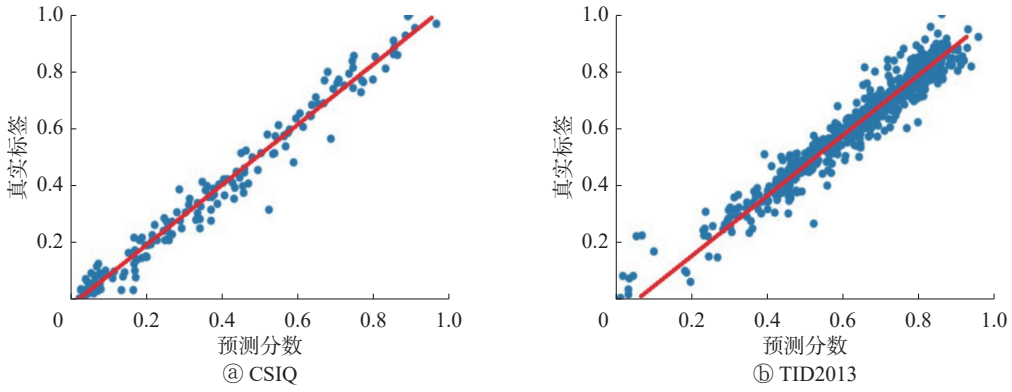
(a) PSNR 评价分数与主观分数散点图



(b) SSMI 评价分数与主观分数散点图



(c) AHIQ 评价分数与主观分数散点图



(d) BFFGQA 评价分数与主观分数散点图

图 5　不同方法的图像质量评价结果可视化对比

Fig. 5　Visual comparison of image quality evaluation results of different methods

从图 5 可以看出, PSNR 和 SSIM 方法对应的散点分布较为分散, 说明其评价分数与真实标签的相关性较弱, 评价性能较差; AHIQ 和本文 BFFGQA 方法对应的散点分布集中度较高, 且分布于红色直线附近, 说明这两种方法的评价性能较好; 进一步对比 AHIQ 和本文 BFFGQA 方法, BFFGQA 对应的散点分布更加接近红色直线, 在 CSIQ 数据集上体现的较为明显。

3.6　模型计算复杂度分析

为了分析模型的计算复杂度和效率, 我们在实验中对比了参数量、计算量(FLOPs)、推理时间和显存占用情况。实验结果如表 6 所示。

表 6　参数量、FLOPs、运行时间及显存占用对比

Table 6　Comparison of parameters, FLOPs, inference time, and memory usage

方法	参数量 /10^6	FLOPs /10^{12}	运行时间 /s	最大显存占用 /MB
BFFGQA	29.96	1.73	0.801 3	2 373.6
AHIQ	44.96	1.65	0.786 8	2 943.5

可以看出, 本文方法在计算效率方面与对比方法接近, 推理时间仅增加 1.8%, 但通过优化特征提取和融合策略, 参数量减少了约 33%, 最大显存占用降低约 19%。这表明, 本文方法在保证计算性能的同时, 提升了资源利用效率, 更适用于计算资源受限的应用场景。

4　结束语

为提高对真实失真图像质量评价的准确性, 提出了基于双线性特征融合(BFF)和 GRU 质量聚合(GRU-QA)的全参考图像质量评价方法。BFF 模块先通过可变形卷积对局部特征进行可变形卷积筛选处理, 去除特征中与图像质量无关的信息, 引导卷积特征关注图像失真区域; 然后基于双线性池化融合全局和局部特征, 充分保留和利用质量信息。GRU-QA 模块通过 GRU 建立各图像块间的序列相关性, 动态分配不同图像块的权重。实验基于 LIVE、CSIQ、TID2013 和 PIPAL 等 4 个常用的基准图像质量评价数据集开展, 采用 PLCC 和 SROCC 两种常用的性能评价指标。分别进行了不同模块的消融实验和不同特征融合方式的消融实验, 实验结果验证了 BFF 和 GRU-QA 模块对提升模型图像质量评价性能的有效性及二者的协同作用。与现有 18 种典型方法进行了对比实验, 实验结果验证了本文方法具有良好的图像质量评价性能和良好的泛化性。

尽管本文方法在图像质量评价任务中取得了一定的效果, 但仍然存在一些局限性。首先, 由于本文采用基于图像块的质量评价策略, 将图像划分为多个区域并分别评估其质量, 这种方法可能在一定程度上引入无关的标签信息干扰, 影响整体质量评价的准确性。未来的研究将重点关注如何有效降低此类干扰, 提高模型的鲁棒性和泛化能力。

参考文献:

[1] 方玉明, 眭相杰, 鄢杰斌, 等. 无参考图像质量评价研究进展[J]. 中国图象图形学报, 2021, 26(2): 265−286.
FANG Yuming, SUI Xiangjie, YAN Jiebin, et al. Progress in no-reference image quality assessment[J]. Journal of image and graphics, 2021, 26(2): 265−286.

[2] 曹玉东, 刘海燕, 贾旭, 等. 基于深度学习的图像质量评价方法综述[J]. 计算机工程与应用, 2021, 57(23): 27−36.
CAO Yudong, LIU Haiyan, JIA Xu, et al. Overview of image quality assessment method based on deep learning [J]. Computer engineering and applications, 2021, 57(23): 27−36.

[3] HU Runze, LIU Yutao, GU Ke, et al. Toward a No-reference quality metric for camera-captured images[J]. IEEE transactions on cybernetics, 2023, 53(6): 3651−3664.

[4] 秦小倩, 杜浩. 基于自然场景统计的图像质量评价算法[J]. 现代电子技术, 2023, 46(23): 36−42.
QIN Xiaoqian, DU Hao. Image quality assessment algorithm based on natural scene statistics[J]. Modern electronics technique, 2023, 46(23): 36−42.

[5] 李沛钊, 王同罕, 贾惠珍, 等. USformer-Net: 基于 U-Net 和 Swin Transformer 的脑部 MRI 图像质量评价方法[J]. 现代电子技术, 2024, 47(7): 1−7.
LI Peizhao, WANG Tonghan, JIA Huizhen, et al. USformer-Net: brain MRI image quality assessment fusing U-Net and Swin Transformer[J]. Modern electronics technique, 2024, 47(7): 1−7.

[6] 江本赤, 卞仕磊, 史晨阳, 等. 基于色貌尺度相位一致性的全参考图像质量评价[J]. 光学精密工程, 2023, 31(10): 1509−1521.
JIANG Benchi, BIAN Shilei, SHI Chenyang, et al. Full reference image quality assessment based on color appearance-based phase consistency[J]. Optics and precision engineering, 2023, 31(10): 1509−1521.

[7] 赵文清, 许丽娇, 陈昊阳, 等. 多层特征融合与语义增强的盲图像质量评价[J]. 智能系统学报, 2024, 19(1): 132−141.
ZHAO Wenqing, XU Lijiao, CHEN Haoyang, et al. Blind

image quality assessment based on multi-level feature fusion and semantic enhancement[J]. CAAI transactions on intelligent systems, 2024, 19(1): 132−141.

[8] 王伟, 刘辉, 杨俊安. 一种特征字典映射的图像盲评价方法研究[J]. 智能系统学报, 2018, 13(6): 989−993.
WANG Wei, LIU Hui, YANG Jun'an. Blind quality evaluation with image features codebook mapping[J]. CAAI transactions on intelligent systems, 2018, 13(6): 989−993.

[9] 王成, 刘坤, 杜砾. 全参考图像质量指标评价分析[J]. 现代电子技术, 2023, 46(21): 39−43.
WANG Cheng, LIU Kun, DU Li. Evaluation and analysis of full reference image quality indicators[J]. Modern electronics technique, 2023, 46(21): 39−43.

[10] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600−612.

[11] SAMPAT M P, WANG Zhou, GUPTA S, et al. Complex wavelet structural similarity: a new image similarity index[J]. IEEE transactions on image processing, 2009, 18(11): 2385−2401.

[12] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers. Pacific Grove: IEEE, 2003: 1398−1402.

[13] XUE Wufeng, ZHANG Lei, MOU Xuanqin, et al. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index[J]. IEEE transactions on image processing, 2014, 23(2): 684−695.

[14] ZHANG Lin, SHEN Ying, LI Hongyu. VSI: a visual saliency-induced index for perceptual image quality assessment[J]. IEEE transactions on image processing, 2014, 23(10): 4270−4281.

[15] KIM J, LEE S. Deep learning of human visual sensitivity in image quality assessment framework[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1969−1977.

[16] PRASHNANI E, CAI Hong, MOSTOFI Y, et al. PieAPP: perceptual image-error assessment through pairwise preference[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1808−1817.

[17] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 586−595.

[18] DING Keyan, MA Kede, WANG Shiqi, et al. Image quality assessment: unifying structure and texture similarity [J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 44(5): 2567−2581.

[19] SEO S, KI S, KIM M. A novel just-noticeable-difference-based saliency-channel attention residual network for full-reference image quality predictions[J]. IEEE transactions on circuits and systems for video technology, 2021, 31(7): 2602−2616.

[20] GAO Fei, WANG Yi, LI Panpeng, et al. DeepSim: Deep similarity for image quality assessment[J]. Neurocomputing, 2017, 257: 104−114.

[21] WU Jinjian, MA Jupo, LIANG Fuhu, et al. End-to-end blind image quality prediction with cascaded deep neural network[J]. IEEE transactions on image processing, 2020, 29: 7414−7426.

[22] 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, 2018, 27(1): 206−219.

[23] CHEON M, YOON S J, KANG B, et al. Perceptual image quality assessment with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE, 2021: 433−442.

[24] VARGA D. No-reference image quality assessment using the statistics of global and local image features[J]. Electronics, 2023, 12(7): 1615.

[25] VARGA D. No-reference quality assessment of authentically distorted images based on local and global features [J]. Journal of imaging, 2022, 8(6): 173.

[26] LAO Shanshan, GONG Yuan, SHI Shuwei, et al. Attentions help CNNs see better: attention-based hybrid image quality assessment network[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022: 1139−1148.

[27] 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.

[28] YUAN Li, CHEN Yunpeng, WANG Tao, et al. Tokens-to-token ViT: training vision transformers from scratch on ImageNet[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 538−547.

[29] 毛明毅, 吴晨, 钟义信, 等. 加入自注意力机制的 BERT 命名实体识别模型[J]. 智能系统学报, 2020, 15(4): 772−779.
MAO Mingyi, WU Chen, ZHONG Yixin, et al. BERT named entity recognition model with self-attention mechanism[J]. CAAI transactions on intelligent systems, 2020, 15(4): 772−779.

[30] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×6 words: Transformers for image recognition at scale[EB/OL]. (2020−10−22)[2024−07−24]. https://arxiv.org/abs/2010.11929.

[31] WANG Hao, ZHANG Yue, LIU Chao, et al. sEMG based hand gesture recognition with deformable convolutional network[J]. International journal of machine learning and cybernetics, 2022, 13(6): 1729−1738.

[32] SHI Shuwei, BAI Qingyan, CAO Mingdeng, et al. Region-adaptive deformable network for image quality assessment[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE, 2021: 324−333.

[33] ZHANG Haochen, LIU Dong, XIONG Zhiwei. Convolutional neural network-based video super-resolution for action recognition[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018: 746−750.

[34] ZHANG Weixia, MA Kede, YAN Jia, et al. Blind image quality assessment using a deep bilinear convolutional neural network[J]. IEEE transactions on circuits and systems for video technology, 2020, 30(1): 36−47.

[35] 刘扬, 王立虎, 杨礼波, 等. 改进 EEMD-GRU 混合模型在径流预报中的应用[J]. 智能系统学报, 2022, 17(3): 480−487.
LIU Yang, WANG Lihu, YANG Libo, et al. Application of improved EMD-GRU hybrid model in runoff forecasting[J]. CAAI transactions on intelligent systems, 2022, 17(3): 480−487.

[36] GU Jinjin, CAI Haoming, CHEN Haoyu, et al. PIPAL: a large-scale image quality assessment dataset for perceptual image restoration[M]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 633−651.

[37] LAPARRA V, BALLÉ J, BERARDINO A, et al. Perceptual image quality assessment using a normalized Laplacian pyramid[J]. Electronic imaging, 2016, 28(16): 1−6.

[38] 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.

[39] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE transactions on image processing, 2006, 15(2): 430−444.

[40] ZHANG Lin, ZHANG Lei, MOU Xuanqin, et al. FSIM: a feature similarity index for image quality assessment[J]. IEEE transactions on image processing, 2011, 20(8): 2378−2386.

[41] CHEN Chaofeng, MO Jiadi, HOU Jingwen, et al. TOPIQ: a top-down approach from semantics to distortions for image quality assessment[J]. IEEE transactions on image processing, 2024, 33: 2404−2418.

作者简介：

王亚茹，博士，讲师，主要研究方向为模式识别与计算机视觉、数据挖掘和电力视觉。发表学术论文 10 余篇。E-mail: wangyaru@ncepu.edu.cn

杨春旺，硕士研究生，主要研究方向为图像质量评价。E-mail: 2312661795@qq.com

张诗吟，讲师，博士，主要研究方向为计算机视觉和图像处理。发表学术论文 5 篇。E-mail: shiyinzhang@ncepu.edu.cn。