

文章编号:1007-2780(XXXX)XX-0001-15

# 人眼视觉特性引导的频率感知无参考图像 质量评价方法

高敏娟\*, 李彦康, 宋晨烨

(太原师范学院 计算机科学与技术学院, 山西 晋中 030619)

**摘要:**针对无参考图像质量评价在复杂失真场景下频率信息利用不足及高频细节表征能力受限的问题,提出一种人眼视觉特性引导的频率感知质量评价方法。首先,以轻量级层次化视觉 Transformer 网络为骨干进行多尺度特征提取,实现从局部纹理到全局语义的信息建模。其次,引入受视觉对比敏感函数启发的频率感知门控模块,从通道、频率与空间三个维度对失真特征进行自适应调制,以增强模型对敏感频率信息的感知能力。随后,设计基于离散余弦变换的高频增强模块,通过频域中的空间与通道交互强化模型对局部高频退化的响应能力。最后,对多层级特征进行聚合,以提升特征间的协同表征能力。实验结果表明,本文方法在 LIVE 数据集上的皮尔逊线性相关系数与斯皮尔曼秩相关系数分别达到 0.981 和 0.980;在真实失真数据集 LIVEC 上分别达到 0.911 和 0.888;在跨数据集实验中的斯皮尔曼秩相关系数达到 0.907。该方法能够有效提升复杂失真场景下的质量预测精度,具有较好的泛化能力与鲁棒性,可为智能视觉系统中的感知质量优化提供支持。

**关键词:**无参考图像质量评价;视觉对比敏感函数;频率感知;离散余弦变换;特征融合

中图分类号:TP391.41 文献标识码:A doi:10.37188/CJLCD.2026-0091 CSTR:32172.14.CJLCD.2026-0091

## Frequency-aware no-reference image quality assessment method guided by human visual system characteristics

GAO Minjuan\*, LI Yankang, SONG Chenye

(School of Computer Science and Technology, Taiyuan Normal University, Jinzhong City,  
Shanxi Province, 030619, China)

**Abstract:** To address the issues of insufficient frequency information utilization and the limited capability to represent high-frequency details in complex distortion scenarios for no-reference image quality assessment (NR-IQA), a frequency-aware quality assessment method guided by human visual system (HVS) characteristics is proposed. First, a lightweight hierarchical Vision Transformer network is employed as the backbone for multi-scale feature extraction, enabling information modeling from local textures to global semantics. Second, a frequency-aware gating module inspired by the contrast sensitivity

收稿日期:2026-05-20;修订日期:2026-06-17.

基金项目:山西省基础研究计划(自由探索类)项目(No.202203021222237);山西省高等学校科技创新项目资助(No.2022L401)

Supported by Fundamental Research Programme of Shanxi Province (No.202203021222237) Supported by Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (No.2022L401)

\*通信联系人, E-mail:gaominjuan1984@tynu.edu.cn

function (CSF) is introduced. This module adaptively modulates distortion features across channel, frequency, and spatial dimensions to enhance the model's perceptual sensitivity to sensitive frequency information. Subsequently, a high-frequency enhancement module based on the Discrete Cosine Transform (DCT) is designed, which strengthens the model's response to local high-frequency degradation through spatial and channel interactions in the frequency domain. Finally, multi-level features are aggregated to improve the collaborative representation capability among features. Experimental results demonstrate that the proposed method achieves a Pearson Linear Correlation Coefficient (PLCC) of 0.981 and a Spearman Rank-order Correlation Coefficient (SRCC) of 0.980 on the LIVE dataset. On the authentic distortion dataset LIVEC, the PLCC and SRCC reach 0.911 and 0.888, respectively, while the SRCC in the cross-dataset evaluation reaches 0.907. The proposed method effectively improves quality prediction accuracy in complex distortion scenarios, exhibits strong generalization and robustness, and can provide reliable support for perceptual quality optimization in intelligent visual systems.

**Key words:** no-reference image quality assessment; contrast sensitivity function; frequency perception; discrete cosine transform; feature fusion

## 1 引言

随着成像设备、移动互联网及智能视觉系统的快速发展,加之全息显示与新型光学防伪等前沿视觉技术的不断涌现<sup>[1]</sup>,数字图像已成为信息获取与传播的重要载体。然而,图像在采集、压缩、传输及显示等过程中,往往会受到噪声、模糊、压缩伪影及曝光异常等多种因素影响,导致图像视觉质量退化。特别是在真实场景中,多种失真通常同时存在,并在不同空间尺度与频率范围内相互耦合,不仅影响用户的主观视觉体验,也会降低目标检测、图像识别等下游视觉任务的准确性。由于实际应用中通常难以获得对应的无失真参考图像,无参考图像质量评价(No-Reference Image Quality Assessment, NR-IQA)逐渐成为图像处理与视觉感知领域的重要研究方向<sup>[2-5]</sup>。

人类视觉系统(Human Visual System, HVS)在感知图像质量时具有明显的频率选择特性,对边缘、纹理以及局部结构等高频信息尤为敏感<sup>[6-7]</sup>。已有研究表明,视觉对比敏感函数(Contrast Sensitivity Function, CSF)能够有效描述人眼对不同空间频率的响应规律,并呈现出典型的带通特性<sup>[8]</sup>。当图像受到模糊、压缩或噪声等失真影响时,高频结构往往会发生明显退化,从而直接影响视觉感知质量。因此,仅依赖空间域特征或单一尺度信息,通常难以充分模拟复杂失真场景下的人眼感知机制<sup>[9-10]</sup>。如何结合 HVS 的频率感知特

性,对不同频率成分进行有效建模,已成为提升 NR-IQA 性能的重要研究方向。

近年来,基于卷积神经网络(Convolutional Neural Network, CNN)与 Transformer 的深度学习方法在 NR-IQA 任务中取得了显著进展,并在全局依赖建模、多尺度特征提取以及语义信息表达等方面展现出较强优势<sup>[11-12]</sup>。然而,现有研究仍存在两个亟待解决的问题:一方面,多数方法更加侧重空间域中的全局语义建模,而对边缘模糊、振铃伪影及纹理损失等与视觉感知密切相关的高频细节退化缺乏显式建模,导致模型对复杂真实失真的感知能力受限<sup>[13]</sup>;另一方面,在多尺度特征融合过程中,不同层级特征之间的关联关系尚未得到充分利用,局部细节与全局语义之间缺乏有效协同,从而影响模型对复杂失真特征的表达能力<sup>[14]</sup>。上述问题在一定程度上限制了模型在真实复杂失真场景以及跨数据库任务中的泛化性能。

针对上述问题,本文提出一种人眼视觉特征引导的频率感知无参考图像质量评价方法。首先,以 Swin Transformer-Tiny 作为骨干网络,构建层级化特征提取结构,实现从局部纹理到全局语义的信息建模;其次,设计受视觉 CSF 启发的频率感知门控模块,从通道、频率与空间三个维度对失真特征进行自适应调制,以增强模型对人眼敏感频率成分的感知能力;进一步地,提出基于离散余弦变换(DCT)的高频感知增强模块,通

过空间与通道交互机制强化边缘与纹理等高频结构信息,从而提升模型对局部细节退化的表征能力。最后,在 LIVE、CSIQ、TID2013、LIVEC 以及 KonIQ-10k 等公开数据集上的实验结果表明,本文方法在合成失真与真实失真场景下均取得了较好的预测性能,并表现出较强的跨数据集泛化能力。

## 2 相关工作

### 2.1 基于深度学习的无参考图像质量评价

无参考图像质量评价研究经历了由自然场景统计建模向深度特征学习的发展过程。早期方法主要依赖空间域统计规律,通过人工设计特征对图像退化程度进行量化,例如经典的盲图像空间质量评估模型<sup>[15]</sup>、自然图像质量评价模型<sup>[16]</sup>以及引入多尺度、多方向特征的改进模型<sup>[17]</sup>。随着深度学习技术的发展,基于卷积神经网络的方法显著提升了模型的特征表征能力。例如,引入自适应超网络机制后,模型能够根据输入图像内容动态生成质量预测参数,从而提升对复杂未知失真的评价能力。此外,在图像质量评价的特征提取层面,全参考方法也为无参考模型提供了重要借鉴。例如,基于纹理奇异值分解的评价框架<sup>[18]</sup>,通过引入图像梯度与纹理信息作为权重,在对比度失真与局部结构损伤的量化上展现了较优的性能。然而,由于实际应用中通常难以获得无失真参考图像,深入挖掘图像自身的语义与频域多尺度特征成为当前 NR-IQA 研究的核心。

近年来,视觉 Transformer 架构凭借其全局长程依赖建模能力,逐渐成为该领域的重要研究方向。针对复杂分辨率与多尺度失真难以兼顾的问题,多尺度图像质量 Transformer 模型<sup>[9]</sup>实现了对原始分辨率输入的支持;结合相对排序学习与自一致性约束的评价框架<sup>[13]</sup>提升了预测鲁棒性;多维注意力网络<sup>[10]</sup>则增强了全局与局部特征的交互能力。类似的多尺度结构与注意力融合机制在图像超分辨率重建等底层视觉任务中,也被证明能够有效强化网络对高频纹理与局部细节的捕捉能力<sup>[19]</sup>。同时,为提升模型在跨场景中的泛化性能,基于质量感知的预训练模型<sup>[11]</sup>以及结合提示学习策略的评价方法<sup>[12]</sup>被相继提出。此外,针对深层网络容易忽略局部退化信息的问

题,研究者进一步引入局部特征增强与语义引导机制。基于特征符号化的轻量化方法<sup>[20]</sup>、局部特征增强的 Transformer 微调策略<sup>[14]</sup>、多任务特征增强网络以及基于语义引导的多尺度特征提取网络<sup>[21]</sup>,均在一定程度上提升了模型对局部失真区域的感知能力。

尽管现有深度学习方法在空间域多尺度特征建模、预训练泛化以及语义引导等方面取得了较大进展,但多数模型仍主要依赖像素空间特征表示,对于与人类视觉系统感知敏感性密切相关的频率域信息缺乏显式利用。这在一定程度上限制了模型对模糊、压缩伪影等具有明显频率衰减特征失真的机理表征能力。

### 2.2 人眼视觉系统与频域特征建模

图像质量的主观评价最终取决于人类视觉系统。生理学与心理物理学研究表明,视觉感知在图像处理过程中具有明显的频率选择特性,即对不同空间频率成分及不同空间区域中的失真敏感度存在显著差异<sup>[22]</sup>。视觉对比敏感函数能够有效刻画这一响应规律,其整体呈现典型的带通特性,即对中等频率最为敏感,而对过高或过低频率的敏感性相对较弱<sup>[8]</sup>。随着现代显示技术的发展,针对新型显示设备构建的精确对比敏感模型<sup>[23]</sup>,进一步揭示了复杂视觉感知特性在质量评价中的重要作用。

然而,如何在深层网络架构中兼顾视觉感知特性与计算效率,仍是当前研究的重要问题。相关研究指出,显式引入多维视觉特性虽然能够提升质量评价精度,但往往会增加模型参数量并削弱泛化性能;而通过选择性量化亮度、局部对比度阈值及高频占比等关键感知因子,则能够在较低计算开销下实现预测精度与效率之间的平衡<sup>[24]</sup>。

在传统数字图像处理中,频域分析常用于分离图像中的结构、纹理与噪声信息。然而,当前无参考图像质量评价深度学习框架大多仍侧重于空间域特征建模,尚未将视觉系统的频率感知特性与深层网络进行有效融合。由于缺乏针对频率域特定失真成分的自适应调制机制,模型对模糊、压缩伪影等具有明显频率衰减或突变特征的失真机理刻画能力仍存在一定局限。相关研究进一步表明,传统空间域扩展指标与主观视觉感知之间存在一定偏差,而通过引入自适应分块

离散余弦变换对高频细节损伤进行显式频域量化后,其对最终评价性能贡献得到明显提升<sup>[25]</sup>。因此,探索符合人类视觉对比敏感特性的频率感知机制,并在频域中引导网络重点关注敏感退化成分,可为本文频率感知特征增强模块的构建提供理论依据。

### 2.3 图像高频细节增强与局部失真感知

在真实场景的图像退化过程中,边缘、纹理及局部细节等高频成分通常最先受到破坏,而这些高频结构的变化会直接影响人类对图像质量的主观感知结果<sup>[6]</sup>。然而,深层神经网络在连续下采样与特征压缩过程中,容易导致局部细节信息逐渐丢失。特别是对于视觉 Transformer 等主流架构,由于其在深层阶段更加关注全局长程依赖关系建模,局部纹理与结构特征容易被稀释甚至忽略,从而削弱模型对局部失真区域的感知与表征能力<sup>[26]</sup>。

针对高频特征在网络深层中的流失问题,频域子带信息的保留与复用逐渐成为一种有效思路。相关研究表明,通过在跳跃连接中嵌入高频特征增强模块以抑制低频干扰,或逐元素融合浅层网络生成的冗余高频小波子带<sup>[27]</sup>,能够有效缓解因网络加深导致的纹理细节退化问题。

在图像质量评价领域,研究者也围绕局部失真感知开展了大量研究。已有研究发现,过度依赖高层语义特征容易忽略局部细节退化信息。为此,结合元学习与多尺度失真感知的评价框架

被提出,以尽可能保留局部非均匀失真特征,并提升复杂真实场景下的泛化能力<sup>[28]</sup>。类似地,通过在大规模预训练视觉模型中显式引入局部失真特征,也进一步增强了模型对局部退化结构的感知敏感性<sup>[14]</sup>。

尽管上述方法在局部失真建模方面取得了一定进展,但现有研究大多仍依赖空间域中的多尺度卷积或池化操作,对高频细节的建模能力仍存在一定局限。对于压缩失真产生的块效应以及噪声引起的异常高频增强等典型高频退化,仅依赖空间域特征往往难以实现有效分离与精准表征。相比之下,基于离散余弦变换等频域分析方法能够更加直接地对不同频率成分进行分解与选择,从而更有效地提取边缘、纹理等高频退化信息。因此,结合频域变换与空间—通道交互机制,构建符合人类视觉感知规律的高频感知增强方法,是进一步提升无参考图像质量评价性能的重要方向。

## 3 提出的方法

为提升模型对复杂真实失真图像中频率信息和细节退化特征的表征能力,本文提出了一种人眼视觉特征引导的频率感知无参考图像质量评价方法。该方法主要包括骨干特征提取、频率感知特征增强、多尺度特征融合以及质量预测四个部分,整体结构如图 1 所示,并以端到端方式完成训练与推理。

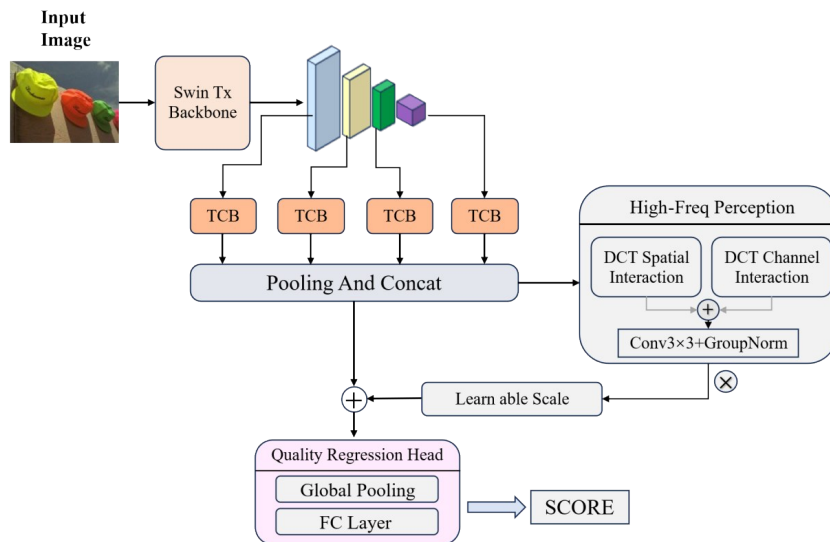


图 1 整体框架图

Fig. 1 Overall framework diagram

在训练阶段,输入RGB失真图像经预训练的SwinTransformer-Tiny,得到不同阶段输出的四组层级特征,对应不同感受野与语义程度,由浅至深覆盖从局部纹理到全局结构语义的多层次质量线索。四组层级特征分别送入对应的频率感知Transformer-CNN桥接模块进行特征增强,各模块在保持分支并行处理的同时,对每一路特征完成感知调制与更新,输出增强后的多尺度特征集合。多尺度特征融合模块将各分支输出对齐至统一空间分辨率后进行通道拼接,将多层次信息汇聚至同一质量表示空间,形成融合全局表征。将融合后的特征作为高频感知增强模块的输入,通过通道交互与空间交互提取高频注意信息,输出的高频增强特征经可学习缩放系数加权后与融合特征相加,得到最终增强表示。质量回归头对最终表示依次执行全局池化与全连

接映射,输出最终的质量分数。训练过程以数据集提供的主观评分为监督信号,通过回归损失优化网络参数。推理阶段保持相同前向流程,对任意输入图像直接输出质量预测分数。

### 3.1 语义特征提取网络

Swin-T采用基于局部窗口的层次化设计,在控制计算复杂度的同时,能够较好地兼顾局部细节信息与整体结构信息的表达,因而更适合用于图像质量评价任务。其整体结构如图2所示。该网络在分层特征提取过程中能够逐步获得不同尺度、不同语义层次的特征表示,形成具有金字塔结构的多尺度特征输出<sup>[29]</sup>。这种由浅入深的层次化表征方式,有利于刻画图像从纹理细节到结构语义的多层次质量线索,也能够与后续多尺度特征融合模块形成较为自然的衔接,为后续频率感知增强和质量预测提供可靠的特征基础。

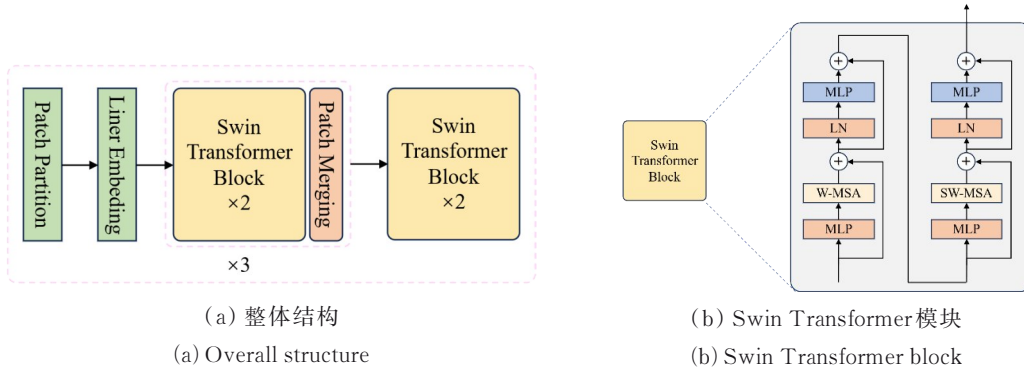


图2 Swin Transformer-Tiny的整体结构图

Fig. 2 Overall structure of Swin Transformer-Tiny

图2(a)展示了Swin Transformer-Tiny的整体结构。输入图像首先经过Patch Partition进行图像块划分,然后通过Linear Embedding将图像块映射到特征嵌入空间。随后网络进入由Swin Transformer Block $\times 2$ 与Patch Merging组成的基本结构单元,该结构在网络中重复三次。每经过一次Patch Merging操作,特征图的空间分辨率减半,通道数增加一倍。网络最后一个阶段只包含Swin Transformer Block $\times 2$ ,不再进行下采样操作。经过四个阶段后,网络输出多层特征 $\{l_1^m \| l_2^m \| l_3^m \| l_4^m\}$ ,对应的通道维度分别为96、192、384和768。这些特征从浅层到深层逐步描述图像信息,浅层特征更多反映局部纹理细节,中层特征包含一定的统计结构

信息,深层特征更侧重整体结构与语义表达,这些信息对于图像质量判断具有参考价值。

图2(b)给出了Swin Transformer Block的内部结构。每个模块由窗口多头自注意力(W-MSA)与移位窗口多头自注意力(SW-MSA)交替组成,并结合层归一化与多层感知机结构进行特征变换。其计算过程如公式(1)所示:

$$\begin{cases} \hat{Z}^l = W - MSA(LN(Z^{l-1})) + Z^{l-1} \\ Z^l = MLP(LN(\hat{Z}^l)) + \hat{Z}^l \\ Z^l = MLP(LN(\hat{Z}^l)) + \hat{Z}^l \\ \hat{Z}^{l+1} = SW - MSA(LN(Z^l)) + Z^l \end{cases} \quad (1)$$

### 3.2 频域失真感知的质量调制模块

人类视觉系统(HVS)在感知图像质量时,对不同空间频率成分以及不同空间区域中的失真并不具有相同的敏感性。已有研究表明,视觉对比敏感函数(CSF)能够刻画人眼对不同空间频率的响应特性,其总体上表现出带通特征,即人眼通常对中等空间频率更为敏感,而对过低和过高空间频率的敏感性相对较弱;其中,与纹理、边缘和细节相关的频率成分对质量感知具有重要影响<sup>[8]</sup>。此外,选择性视觉注意机制表明,不同空间区域中的失真对整体质量感知的影响并不相同<sup>[30]</sup>。因此,在图像质量评价任务中,从频率域角度刻画视觉感知机制,有助于更准确地描述人眼对图像失真的感知特点,并为面向感知的质量建模提供依据<sup>[22-23]</sup>。基于这一认识,本文在网络结构中引入了一种受视觉对比敏感函数(Contrast Sensitivity Function, CSF)启发的频率感知门控模块(CSF-WeightedFreqGate)。该模块通过在频率域对特征进行加权分析,使模型能够更加关注与人类视觉感知密切相关的频率失真信息,从而提升图像质量预测的准确性。具体而言,对于输入特征  $x \in R^{B \times C \times H \times W}$ ,首先通过二维快速傅里叶变换将其映射到频率域,并计算其幅值谱,用以表征不同频率分量的能量分布。在此基础上构建二维频率网格,根据视觉对比敏感函数建立频率权重模型,使网络能够对不同频率成分进行自适应加权。CSF权重函数形式如公式(2)所示:

$$CSF(f) = f^\alpha \exp(-\beta f), \quad (2)$$

其中  $f$  表示空间频率,  $\alpha$  与  $\beta$  为可学习参数,用于调节频率响应曲线的形状。该函数呈现典型的带通特性,能够强化人眼更敏感的中高频信息,同时抑制对视觉感知贡献较小的极低频或极高频成分。随后利用该权重对频域幅值进行加权,得到频率感知能量分布,并通过全局平均操作提取每个通道的频率能量描述,最终通过多层感知机生成通道级频率门控权重,实现对特征响应的自适应调制。

为进一步增强多维度的失真表征能力,本文构建了结合频率感知机制的特征建模模块,如图3所示,值得说明的是,为了实现多维度失真特征的渐进式精炼,该模块并未采用通道、频率与空间特征的并行调制,而是遵循“语义筛选—频率

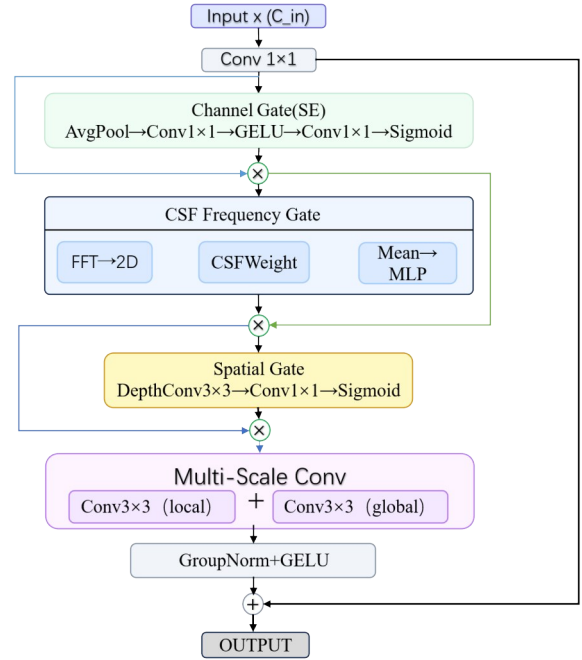


图3 频率感知 Transformer-CNN 混合

Fig. 3 Frequency-aware transformer-CNN hybrid module

建模—空间定位”的串行协同处理思路逐步完成失真特征增强。

具体而言:首先,利用通道注意机制对骨干网络输出的不同特征通道进行自适应建模,筛选并聚焦于与图像质量感知密切相关的关键特征维度,为后续的精细化失真分析剔除冗余干扰;随后,引入受 CSF 启发的频率感知门控机制,对筛选后的纯净特征进行频域加权,以显式增强模型对模糊、噪声及压缩伪影等失真所引起的全局频率能量变化的感知能力;最后,通过空间门控分支进一步关注边缘、纹理及结构退化等局部关键区域,实现失真位置的空间精准定位与增强表达。这种串行设计使得不同维度的特征调制能够形成有效的层级互补,最大程度地强化了模型对复杂复合失真的机理刻画能力。

在完成上述渐进式多维度特征调制之后,模块进一步引入由局部卷积与空洞卷积组成的多尺度卷积结构,以同时建模细粒度局部纹理与更大范围的结构上下文信息。最后采用质量感知残差连接方式进行特征融合,如公式(3)所示:

$$Y = x + \alpha \cdot F(x), \quad (3)$$

其中  $x$  表示输入特征,  $F(\cdot)$  表示经过多分支调制与多尺度建模后的特征变换,  $\alpha$  为可学习的残差

调节系数。通过该设计,模块能够在保持网络稳定性的同时自适应控制残差特征的贡献。

综上所述,该模块从通道、频率和空间三个维度对图像特征进行协同调制,并利用CSF函数显式建模人眼对不同频率失真的感知特性,使网络能够更加关注图像中对主观质量影响显著的结构与纹理信息。这种设计在增强失真感知能力的同时保持了较低的计算复杂度,从而为无参考图像质量评价任务提供更加有效的特征表达能力。

### 3.3 高频感知模块

为建模人眼视觉系统对边缘、纹理等高频结构的敏感性,本文设计高频感知模块,结构如图4所示。该模块基于离散余弦变换从空间与通道两个维度选择性增强高频成分,经卷积归一化后融合输出。

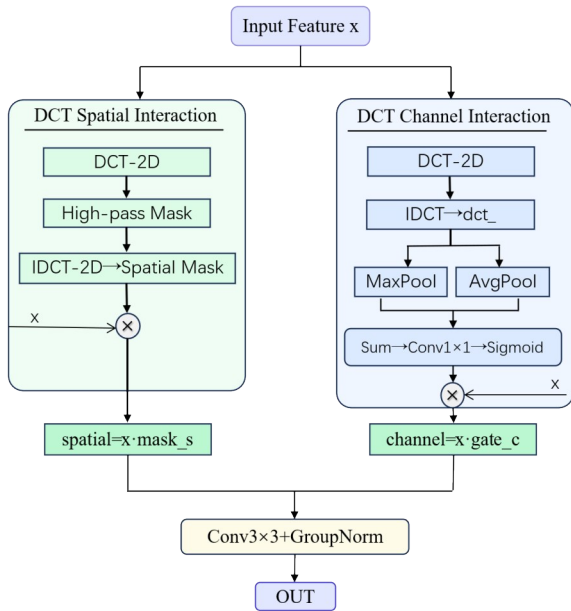


图4 高频感知模块

Fig. 4 High-frequency perception module

DCT空间交互分支(Spatial Interaction)以输入特征  $x \in \mathbb{R}^{B \times C \times H \times W}$  为对象进行二维DCT变换得到频域表示,如公式(4)所示:

$$X_f = DCT(x). \quad (4)$$

DCT系数在二维平面中通常呈现左上为低频、右下为高频的分布特征,因此根据比例参数  $r_h, r_w$  构建频域掩码  $M \in \mathbb{R}^{H \times W}$ ,将左上角低频区域置零并保留其余高频区域,从而得到高频保留后的频域特征,如公式(5)所示:

$$\tilde{X}_f = X_f \odot M. \quad (5)$$

对  $\tilde{X}_f$  执行逆DCT变换获得空间域高频响应掩码,如公式(6)所示:

$$\hat{m} = IDCT(\tilde{X}_f). \quad (6)$$

随后通过逐点乘法对原始特征进行重新标定,如公式(7)所示:

$$x_s = x \odot \hat{m}. \quad (7)$$

上述过程在特征层面对高频结构进行强化,使网络能够更容易感知模糊引起的高频衰减、压缩造成的振铃和块效应以及噪声带来的高频异常增强等失真现象。

DCT通道交互分支(Channel Interaction)同样先利用DCT与IDCT将特征映射到频域并保留高频信息得到增强特征  $\hat{x}$ ,随后分别进行自适应最大池化与自适应平均池化以提取统计特征,最大池化对局部尖峰响应与明显伪影更敏感,平均池化反映整体能量变化,两种统计方式能够从不同角度描述高频失真特征。两路统计结果通过分组  $1 \times 1$  卷积(groups=32)进行融合并生成通道注意力权重  $a_c$ ,随后经过Sigmoid函数得到通道权重并对输入特征进行通道重标定,如公式(8)所示:

$$x_c = x \odot \sigma(a_c). \quad (8)$$

该过程能够自动选择对高频失真更敏感的特征通道,同时抑制对质量评估贡献较小的通道响应。空间分支与通道分支的输出进行融合后,通过  $3 \times 3$  卷积完成局部特征混合,再利用GroupNorm进行归一化处理得到最终输出,如公式(9)所示:

$$y = GN(Conv_{3 \times 3}(x_s + x_c)). \quad (9)$$

空间分支反映高频失真出现的位置分布,通道分支刻画更适合表达该类失真的特征维度,两种信息在统一框架中进行联合建模,使网络在较低计算开销下增强对高频结构变化的感知能力,同时为后续质量预测提供更加集中于失真区域的特征表示。

## 4 实验结果与分析

### 4.1 数据集与评价指标

#### 4.1.1 实验数据集

本次实验在五个公开的IQA数据集(三个合

成失真和两个真实失真)上进行测试和评估。对于合成失真,使用LIVE<sup>[31]</sup>、CSIQ<sup>[32]</sup>、TID2013<sup>[33]</sup>,对于真实失真数据集使用CLIVE<sup>[34]</sup>和KonIQ-10k<sup>[3]</sup>。表1给出了这5个公开数据集的详细信息。

表1 数据集的主要参数信息

Tab. 1 Main parameter information of datasets

数据集	参考图像	失真图像	失真类型	失真等级	MOS/DMOS
LIVE	29	799	5	5或6	[0,100]
CSIQ	30	866	6	4或5	[0,1]
TID2013	25	3000	24	5	[0,9]
LIVEC	0	1162	真实失真	—	[0,100]
KonIQ-10k	0	10 073	真实失真	—	[0,100]

#### 4.1.2 评价指标

为客观地量化评估本文所提图像质量评价模型的性能,本文采用图像质量评价领域两种主流的标准统计评价指标:皮尔逊线性相关系数(Pearson Linear Correlation Coefficient, PLCC)和斯皮尔曼秩相关系数(Spearman Rank-order Correlation Coefficient, SRCC)。

##### (1) 皮尔逊线性相关系数(PLCC)

PLCC主要用于衡量模型预测分数与真实主观评分之间的线性相关性,直观反映了模型预测的准确度。如公式(10)所示:

$$PLCC = \frac{\sum_{i=1}^N (s_i - \bar{s})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^N (s_i - \bar{s})^2 \sum_{i=1}^N (q_i - \bar{q})^2}}, \quad (10)$$

式中: $N$ 为测试集中的图像总数; $s_i$ 和 $q_i$ 分别表示第 $i$ 幅图像的真实主观质量评分和模型预测的客观质量分数; $\bar{s}$ 和 $\bar{q}$ 分别代表所有测试图像真实主观评分的平均值和客观预测分数的平均值。

##### (2) 斯皮尔曼秩相关系数(SRCC)

SRCC仅依赖于数据样本的秩次进行计算,用于衡量预测分数与真实评分之间的单调一致性,重点反映了模型对图像质量的排序能力。如公式(11)所示:

$$SRCC = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}, \quad (11)$$

式中: $N$ 为测试集中的图像总数; $d_i$ 表示第 $i$ 幅图像在主观评分序列和客观预测分数序列中的秩

次(Rank)之差。

上述两种评价指标的取值范围均为 $[-1, 1]$ 。在实际评估中,探讨其绝对值即可,即 $[0, 1]$ 之间,评价指标的绝对值越接近于1,表明客观质量评价模型的预测结果与人眼主观视觉感知越吻合,即模型的综合预测性能越优异。

## 4.2 实验环境与实现细节

### 4.2.1 实验环境

本文实验算法基于Python3.9及PyTorch 2.7.1深度学习框架实现。硬件配置采用NVIDIA GeForce RTX 4090显卡(24GB显存),底层加速库使用CUDA 11.8,操作系统为Ubuntu 20.04 LTS。

### 4.2.2 实验细节

在训练阶段,从每张失真图像随机截取若干 $224 \times 224$ 像素图像块,并进行随机水平翻转等数据增强操作,各图像块的质量分数与源图像保持一致。具体地,LIVE、CSIQ和LIVEC数据集每张图像随机采样25个图像块,TID2013采样20个,KonIQ-10k采样4个。采用Adam优化器进行训练,权重衰减系数为 $5 \times 10^{-4}$ ,最大训练轮数(epoch)为10,小批量大小(batch size)在LIVE、CSIQ、TID2013和LIVEC上设为32,在KonIQ-10k上设为64。为避免对预训练骨干网络的过度扰动,骨干网络与预测头采用差异化学习率策略:骨干网络初始学习率设为 $1 \times 10^{-5}$ ,预测头初始学习率设为 $1 \times 10^{-4}$ ,并采用余弦退火重启调度策略(Cosine LR with Restart)进行学习率衰减,最小学习率衰减至 $1 \times 10^{-6}$ ,前2个epoch执行线性预热。在测试阶段,LIVE、CSIQ和LIVEC每张测试图像随机抽取25个图像块,TID2013抽取20个,KonIQ-10k抽取8个,取各图像块预测分数的中位数作为最终图像质量评分。实验采用相同设置,以固定随机种子(2025)生成10组不同的数据划分方案,将数据集随机划分为80%训练集与20%测试集,测试数据不参与训练,最终取10组实验的SRCC与PLCC中值作为汇报结果。在网络关键超参数设置方面,本文对所提模块中的核心参数进行了明确限定与动态优化。具体而言,3.2节所述CSF频率感知门控模块中的形状调节参数 $\alpha$ 与 $\beta$ 均设定为动态可学习参数,其初始值分别赋予1.0和2.0,以使网络在训

练初期具备标准的带通感知特性;同时,3.3节高频感知增强模块中的频域掩码比例参数ratio统一设置为(0.25, 0.25),即严格抑制频谱左上角25%的低频区域能量,从而最大程度保留并聚焦于核心的高频退化结构线索。上述超参数的显

式说明有效提高了实验过程的完整性与可复现性。

训练过程采用SmoothL1Loss作为损失函数,以减少异常值影响并提升训练稳定性,其定义如公式(12)所示:

$$SmoothLoss = \begin{cases} 0.5(x-y)^2, & |x-y| < \beta \\ \beta(|x-y| - 0.5\beta), & otherwise \end{cases}, \quad (12)$$

式中: $x$ 为模型预测的质量分数, $y$ 为失真图像对应的真实主观质量标签值。

### 4.3 整体性能对比与分析

#### 4.3.1 主流方法性能对比

为验证本文所提方法的有效性,在LIVE、CSIQ、TID2013、LIVEC和KonIQ-10k五个基准数据集上与多种主流IQA方法进行了对比实验,如表2和表3所示。对比方法包括传统手工特征方法以及基于深度学习的方法。相较于对比方法,本文方法在评价指标上均取得了提升,验证了所提特征建模机制的有效性。在合成失真数据集上,模型在LIVE、CSIQ和TID2013上均取得当前最优结果,在LIVE上PLCC、SRCC达到

表2 各评价方法在合成失真数据集上的性能对比

Tab.2 Performance comparison of different evaluation methods on synthetic distortion datasets

方法	LIVE		CSIQ		TID2013	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
DIIVINE <sup>[35]</sup>	0.906	0.891	0.776	0.804	0.567	0.643
BRISQUE <sup>[15]</sup>	0.944	0.929	0.747	0.812	0.571	0.626
CORNIA <sup>[36]</sup>	0.938	0.937	0.768	0.695	0.733	0.657
HOSA <sup>[37]</sup>	0.943	0.942	0.805	0.721	0.835	0.749
MEON <sup>[38]</sup>	0.955	0.951	0.864	0.852	0.824	0.808
WaDIQaM <sup>[39]</sup>	0.955	0.960	0.844	0.852	0.855	0.835
TS-CNN <sup>[40]</sup>	0.965	0.969	0.904	0.892	0.824	0.783
RAN4IQA <sup>[41]</sup>	0.962	0.961	0.931	0.914	0.859	0.820
TIQA <sup>[42]</sup>	0.965	0.949	0.838	0.825	0.858	0.846
MetaIQA <sup>[43]</sup>	0.959	0.960	0.908	0.899	0.868	0.856
HyperIQA <sup>[44]</sup>	0.966	0.962	0.942	0.923	0.858	0.840
CaHDC <sup>[44]</sup>	0.964	0.965	0.914	0.903	0.885	0.862
TReS <sup>[13]</sup>	0.968	0.969	0.942	0.922	0.883	0.863
MMNet <sup>[45]</sup>	0.970	0.970	0.937	0.924	0.853	0.832
VCRNet <sup>[46]</sup>	0.973	0.975	0.932	0.914	0.886	0.865
本文方法	<b>0.981</b>	<b>0.980</b>	<b>0.955</b>	<b>0.947</b>	<b>0.926</b>	<b>0.911</b>

表3 各评价方法在真实失真数据集上的性能对比

Tab.3 Performance comparison of different evaluation methods on authentic distortion

方法	LIVEC		KonIQ	
	PLCC	SRCC	PLCC	SRCC
DIIVINE	0.591	0.588	0.558	0.546
BRISQUE	0.629	0.629	0.685	0.681
CORNIA	0.636	0.617	0.569	0.545
HOSA	0.656	0.621	0.703	0.684
MEON	0.710	0.697	0.628	0.611
WaDIQaM	0.671	0.681	0.807	0.804
TS-CNN	0.687	0.654	0.724	0.713
RAN4IQA	0.612	0.586	0.763	0.752
TIQA	0.861	0.846	0.903	0.892
MetaIQA	0.802	0.835	0.856	0.887
HyperIQA	0.881	0.859	0.917	0.906
CaHDC	0.744	0.738	/	/
TReS	0.877	0.846	0.928	0.915
MMNet	0.876	0.852	/	/
VCRNet	0.878	0.864	0.919	0.906
本文方法	<b>0.911</b>	<b>0.888</b>	<b>0.933</b>	<b>0.922</b>

0.981、0.980,在CSIQ上达到0.955、0.947,在失真类型更加复杂的TID2013上达到0.926、0.911,相比现有代表性方法如VCRNet、TReS和HyperIQA均取得不同程度提升,特别是在TID2013数据集上的提升最为明显,表明模型在多失真场景下的性能更强。在真实失真数据集上,本文方法同样展现出稳定的性能优势。其中,在LIVEC数据集上分别取得0.911和0.888的评价结果,在KonIQ-10k数据集上达到0.933和0.922。相较于现有主流模型,本文方法均表现出较为优异的性能。整体而言,实验结果表明所提出方法在合成失真与真实失真场景中,其预

测精度和泛化能力都表现得十分出色,说明本文所提出的频率感知特征建模机制,在提升图像质量评估性能上起到了关键的作用。

#### 4.3.2 预测一致性分析

客观质量评价模型的准确性,主要通过衡量模型预测分数与人眼真实主观评分(MOS/DMOS)之间的一致性来严格评判。具体而言,本文从定量验证与定性可视化两个维度对预测结果的准确性进行综合评估:在定量维度上,依据皮尔逊线性相关系数(PLCC)来衡量预测分数与主观评分的线性准确度,同时利用斯皮尔曼秩相关系数(SRCC)来检验模型在图像质量排序上的相对准确性;在定性维度上,则通过散点分布与回归拟合直线的贴合程度,直观检验预测分数分布的整体偏差。

基于上述评判准则,为直观评估模型预测效果,图5绘制五个数据集上预测分数和真实MOS之间的散点分布,红色直线表示线性回归拟合结果,KONIQ-10K、LIVE、LIVEC三类数据集的散点,大多聚集在回归直线周边,预测分数和MOS呈现正相关关系,说明模型放在同域和真实场景数据中,可输出稳定预测能力。TID2013的散点分布相对分散,整体依旧保持线性趋势,反映模型可以适配多种合成失真类型,CSIQ数据集的分布呈现负相关,该数据集采用DMOS作为标注,评分方向和模型输出方向相反,散点仍集中在回归直线周边,说明模型依旧保有稳定的质量排序能力,散点分布和回归线整体契合度较高,说明模型在不同数据集上,都可输出可靠的质量预测结果。

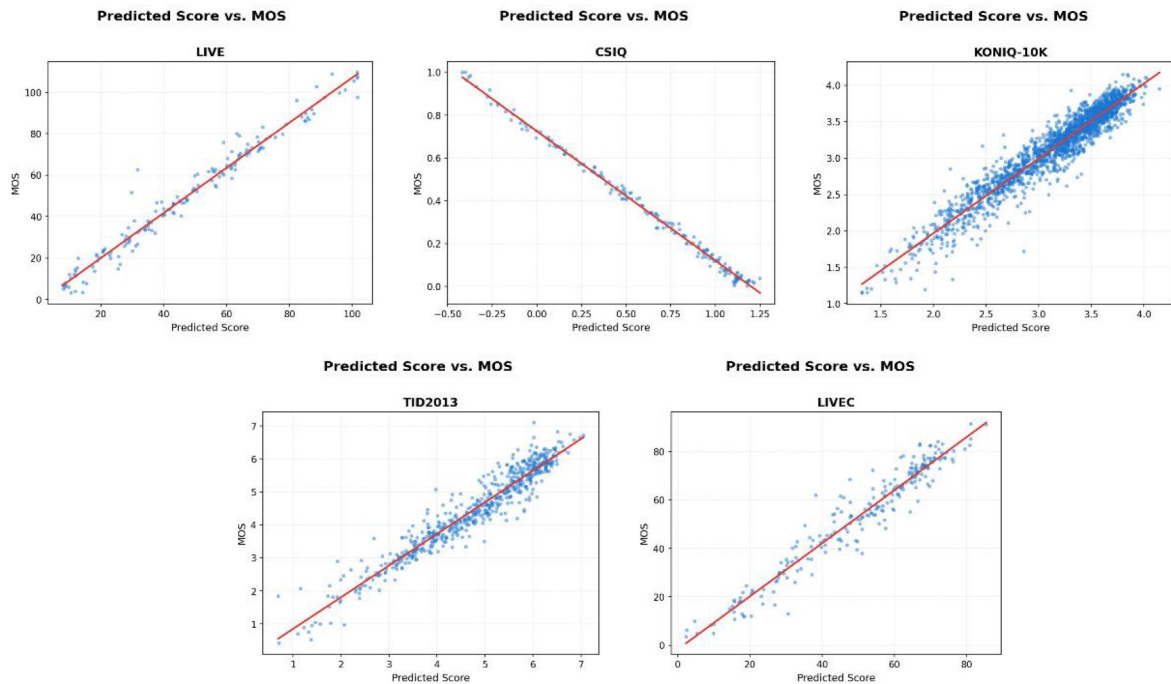


图5 五个数据集测试结果散点图

Fig. 5 Scatter plots of test results on five datasets

#### 4.3.3 单一失真类型评估

为验证模型对各类特定失真的表征能力,本文在LIVE数据集上开展了单一失真类型评估。为深入剖析对比算法的性能差异,本文将表4与表5中的对比方法归纳为三大类主流方法:一是以DIIVINE、BRISQUE、CORNIA和HOSA为代表的传统自然场景统计方法,其依赖空间或变

换域的浅层先验,面对复杂非线性退化时判别力受限;二是以CNN、MEON、WaDIQaM和TS-CNN为首的标准深度卷积网络,虽能实现失真特征的自适应提取,但往往缺乏对频域失真机理的显式刻画;三是CaHDC、HyperIQA、MMMNet和VCRNet为代表的引入高级感知机制的前沿网络,此类方法具备强大的空间语义建

模能力,但在深层网络的连续特征压缩过程中,极易导致边缘、纹理等局部高频细节流失。

表4与表5全面对比了各方法在JPEG2000、JPEG压缩、高斯白噪声、高斯模糊及瑞利衰落等单一失真下的性能。实验结果表明,本文方法在绝大多数场景中均取得较优评价。尤其在JPEG

**表5 LIVE数据集中单一失真类型上的PLCC值比较**  
Tab.5 PLCC comparison on single distortion types in the LIVE dataset

方法	PLCC				
	JP2K	JPEG	WN	GBLUR	FF
DIIVINE	0.901	0.887	0.987	0.787	0.879
BRISQUE	0.923	0.973	0.985	0.951	0.903
CORNIA	0.924	0.906	0.945	0.934	0.913
HOSA	0.923	0.924	0.959	0.965	0.923
CNN	0.942	0.973	0.976	0.953	0.907
MEON	0.965	0.955	0.978	0.942	0.868
WaDIQaM	0.965	0.960	0.973	0.957	0.936
TS-CNN	0.962	0.970	0.990	0.973	0.943
CaHDC	0.953	0.973	0.982	0.955	0.913
HyperIQA	0.973	0.962	0.975	0.971	0.935
MMMNet	/	/	/	/	/
VCRNet	0.972	<b>0.984</b>	0.981	0.968	0.951
本文方法	<b>0.979</b>	<b>0.984</b>	<b>0.989</b>	<b>0.975</b>	<b>0.970</b>

**表4 LIVE数据集中单一失真类型上的SRCC值比较**  
Tab.4 SRCC comparison on single distortion types in the LIVE dataset

方法	SRCC				
	JP2K	JPEG	WN	GBLUR	FF
DIIVINE	0.925	0.913	0.985	0.789	0.873
BRISQUE	0.914	0.965	0.979	0.951	0.877
CORNIA	0.936	0.934	0.962	0.926	0.912
HOSA	0.928	0.936	0.964	0.964	0.934
CNN	0.936	0.965	0.974	0.952	0.906
MEON	0.965	0.954	0.974	0.941	0.885
WaDIQaM	0.962	0.956	0.969	0.946	0.927
TS-CNN	0.959	0.965	0.981	0.970	0.930
CaHDC	0.948	0.970	0.978	0.951	0.898
HyperIQA	0.965	0.956	0.974	0.963	0.923
MMMNet	<b>0.968</b>	0.974	<b>0.985</b>	0.935	0.936
VCRNet	0.964	0.970	0.972	0.955	0.935
本文方法	0.966	<b>0.977</b>	0.982	<b>0.965</b>	<b>0.951</b>

压缩与瑞利衰落失真中表现尤为突出,SRCC和PLCC分别高达0.977、0.984与0.951、0.970。这一结果充分证明,本文所设计的高频感知模块能够有效打破单一空间域的建模局限,精准捕捉并强化特定失真中的高频退化线索,显著提升模型对局部失真特征的综合感知与质量预测性能。

#### 4.3.4 跨数据集泛化能力

为进一步评估模型的跨数据集泛化能力,本文在不同数据集之间开展了域迁移实验。具体而言,分别以LIVE、CSIQ和LIVEC数据集作为训练集,并在未参与训练的CSIQ、LIVE以及TID2013数据集上进行测试,实验结果如表6所示,其中采用SRCC作为评价指标。

当以LIVE数据集作为训练集、CSIQ数据集作为测试集时,本文方法的SRCC达到0.732,略低于VCRNet的0.768和DB-CNN的0.762。该结果表明,在合成失真数据库之间的跨库迁移任务中,本文方法与当前性能较优的方法相比仍存在一定差距。这可能与两个数据库在失真类型分布、失真强度以及主观评分范围等方面存在差异有关,从而对模型的跨域泛化能力带来一定影响。

当以CSIQ数据集作为训练集、LIVE数据集作为测试集时,本文方法取得了0.907的SRCC,优于所有对比方法,较次优方法VCRNet的0.886提升了0.021。该结果说明,本文模型在合成失真场景下具有较好的跨数据集迁移能力,能够较有效地学习不同数据集之间共享的失真特征表示。

在更具挑战性的跨域实验中,以包含真实失真的LIVEC数据集作为训练集,并在包含合成失真的TID2013数据集上进行测试时,本文方法

**表6 跨数据集性能比较**

训练数据集	LIVE	CSIQ	LIVEC
测试数据集	CSIQ	LIVE	TID2013
DIIVIN	0.602	0.817	0.327
BRISQUE	0.573	0.823	0.275
WaDIQaM	0.601	0.813	0.141
DB-CNN	<b>0.758</b>	0.877	0.424
Hall-IQA	0.668	0.833	/
VCRNet	0.693	0.886	0.416
本文方法	0.732	<b>0.907</b>	<b>0.468</b>

取得了0.468的SRCC,较次优方法VCRNet的0.416提升了0.052,表现出相对明显的性能优势。这表明,本文提出的频率感知特征增强机制在一定程度上有助于学习更具域不变性的失真表征,从而提升模型在真实失真与合成失真之间的跨域泛化能力。

综合三组跨数据集实验结果可以看出,本文方法在其中两组实验中取得了最优性能,整体泛化表现优于大多数对比方法,说明所提出的框架在域外场景下具有较好的鲁棒性与泛化能力。

#### 4.4 消融实验

为了验证本文所提出各模块的有效性,本文在LIVE-C数据集上进行了消融实验,结果如表7所示。Baseline模型的SRCC/PLCC为0.874/0.900;加入Transformer-CNN Bridge(TCB)模块后,性能提升至0.880/0.908,分别提高了0.006和0.008,说明TCB能够增强模型的全局质量表征能力。加入High-Freq Perception(HFP)模块后,SRCC提升至0.885,增幅达到0.011,表明在真实场景多源复合失真条件下,高频细节信息对图像质量评估具有重要作用,HFP能够有效增强模型对细节与纹理失真的感知能力。

表7 各模块性能比较

Tab. 7 Performance comparison of different modules

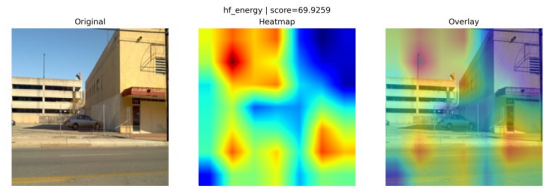
Baseline	TCB	HFP	LIVE-C	
			SRCC	PLCC
✓			0.874	0.900
✓	✓		0.880	0.908
✓		✓	0.885	0.900
✓	✓	✓	0.888	0.911

当同时引入TCB与HFP模块时,模型取得了0.888/0.911的最优结果,相较Baseline分别提升0.014和0.011。该结果表明,两种模块在功能上具有较好的互补性:其中TCB更侧重于提升全局语义层面的质量表征一致性,而HFP更侧重于增强模型对局部纹理与细节失真的敏感性,二者结合能够进一步提升模型在真实失真场景下的鲁棒性与泛化能力。

#### 4.5 特征响应可视化

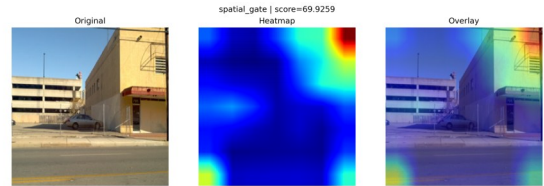
为更直观地理解模型在图像质量判断中的关注区域,图6展示了同一图像在三种不同注意

力下的Grad-CAM可视化结果,包括融合梯度激活(gradcam\_fused)、高频能量图(hf\_energy)以及空间门控权重(spatial\_gate)。通过图6(c)热图可以看出,模型在评估图像质量时会同时关注多处结构信息,建筑、道路及远处边缘等多个结构位置产生响应。图6(a)反映出该分支对细节与清晰度变化较为敏感,对建筑顶部和纹理细节较丰富的位置更加敏感,这些区域与图像中的高频边缘分布基本一致。图6激活区域集中,集中范围覆盖图像右上方结构区域,背景区域响应偏弱,空间门控可突出与质量判断相关的区域,同时压制无关信息输出。整体来看,三种热图从梯度响应、频率特征和空间选择三个角度感知图像质量,共同形成较稳定的质量感知表示。



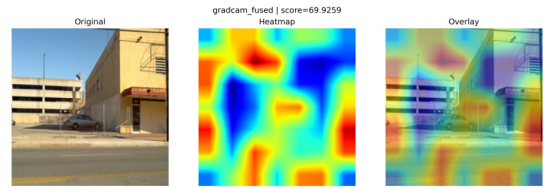
(a) 高频能量图

(a) High-frequency energy map



(b) 空间门控权重

(b) Spatial gate weight



(c) 融合梯度激活

(c) Fused gradient activation

图6 特征响应可视化图

Fig. 6 Visualization of feature responses

## 5 结论与展望

针对无参考图像质量评价中频率信息利用

不足以及高频细节表征能力受限的问题,本文提出了一种人眼视觉特性引导的频率感知评价方法。该方法基于层次化视觉 Transformer 实现多尺度特征建模,并通过受视觉对比敏感函数启发的频率门控机制对失真特征进行自适应调制;同时结合离散余弦变换强化模型对边缘、纹理等局部高频退化信息的响应能力。

实验结果表明,本文方法在 LIVE 数据集上的 PLCC 与 SRCC 分别达到 0.981 和 0.980;在真实失真数据集 LIVEC 上分别达到 0.911 和 0.888;在以 CSIQ 为训练集、LIVE 为测试集的跨数据集实验中,模型的 SRCC 达到 0.907。上述结果表明,所提出的频域特征建模机制能够有效

提升模型对复杂失真场景的质量预测能力,并在合成失真与真实失真场景下表现出较好的预测精度、鲁棒性与跨库泛化能力。

尽管本文方法在多个公开数据集上取得了较好的实验结果,但在开放环境与未知复杂失真场景中的适应能力仍有进一步提升空间。未来研究将主要围绕以下几个方向展开:一是结合动态频率选择策略,构建更加符合人类视觉动态感知规律的细粒度频率表征方法;二是引入领域泛化与对比学习机制,提升模型对未知复杂失真场景的自适应能力;三是探索将多模态大模型的视觉语义与场景理解能力引入图像质量评价框架,推动图像质量评价向高层语义感知与主观体验建模方向发展。

#### 参 考 文 献:

- [1] PENG Y, WANG Z P, JIANG L, *et al.* Reconfigurable holographic patterns for optical security multiplexing fabricated by 3D spatially modulated femtosecond pulses [J]. *Light: Advanced Manufacturing*, 2025, 6: 46.
- [2] GHADIYARAM D, BOVIK A C. Perceptual quality prediction on authentically distorted images using a bag of features approach [J]. *Journal of Vision*, 2017, 17(1): 32.
- [3] HOSU V, LIN H, SZIRANYI T, *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.
- [4] SU S, YAN Q, ZHU Y, *et al.* Blindly assess image quality in the wild guided by a self-adaptive hyper network [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3667-3676.
- [5] GAO M, SONG C, ZHANG Q, *et al.* Research progress on color image quality assessment [J]. *Journal of Imaging*, 2025, 11(9): 307.
- [6] CHEN C, MO J, HOU J, *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.
- [7] YU L. Multi-task feature enhancement network for no-reference image quality assessment [R]. arXiv preprint arXiv: 2411.07556, 2024.
- [8] CAMPBELL F W, ROBSON J G. Application of Fourier analysis to the visibility of gratings [J]. *The Journal of Physiology*, 1968, 197(3): 551.
- [9] KE J, WANG Q, WANG Y, *et al.* Musiq: Multi-scale image quality transformer [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5148-5157.
- [10] YANG S, WU T, SHI S, *et al.* Maniq: Multi-dimension attention network for no-reference image quality assessment [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 1191-1200.
- [11] ZHAO K, YUAN K, SUN M, *et al.* Quality-aware pre-trained models for blind image quality assessment [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 22302-22313.
- [12] CHEN Z, QIN H, WANG J, *et al.* Promptiq: Boosting the performance and generalization for no-reference image quality assessment via prompts [C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 247-264.
- [13] 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. 2022: 1220-1230.

- [14] XU K, LIAO L, XIAO J, *et al.* Boosting image quality assessment through efficient transformer adaptation with local feature enhancement [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 2662-2672.
- [15] 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.
- [16] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer [J]. *IEEE Signal Processing Letters*, 2012, 20(3): 209-212.
- [17] ZHANG L, ZHANG L, BOVIK A C. A feature-enriched completely blind image quality evaluator [J]. *IEEE Transactions on Image Processing*, 2015, 24(8): 2579-2591.
- [18] 李佳欣, 段发阶, 傅骁, 等. 基于纹理奇异值分解的全参考图像质量评价 [J]. *光学精密工程*, 2025, 33(1): 107-122.  
LI J X, DUAN F J, FU X, *et al.* Full-reference image quality assessment based on texture singular value decomposition [J]. *Optics and Precision Engineering*, 2025, 33(1): 107-122. (in Chinese)
- [19] 陈纯毅, 吴欣怡, 胡小娟, 等. 多尺度注意力融合的图像超分辨率重建 [J]. *中国光学(中英文)*, 2023, 16(5): 1034-1044.  
CHEN C Y, WU X Y, HU X J, *et al.* Image super-resolution reconstruction with multi-scale attention fusion [J]. *Chinese Optics*, 2023, 16(5): 1034-1044. (in Chinese)
- [20] 宋巍, 李嘉瑾, 刘晓晨, 等. 基于特征符号化和 Transformer 的无参考图像质量评价方法 [J]. *液晶与显示*, 2023, 38(3): 356-367.  
SONG W, LI J J, LIU X C, *et al.* No-reference image quality assessment method based on feature symbolization and Transformer [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(3): 356-367. (in Chinese)
- [21] JI P, WANG W, LV Z, *et al.* No-reference image quality assessment via semantic-guided multi-scale feature extraction [J]. *IET Image Processing*, 2025, 19(1): e70221.
- [22] MOORTHY A K, WANG Z, BOVIK A C. Visual perception and quality assessment [M]. New Jersey: John Wiley & Sons, 2011: 419-439.
- [23] MOZHAJEVA A, CREE M J, DURRANT R J, *et al.* A contrast sensitivity model of the human visual system in modern conditions for presenting video content [J]. *PLOS One*, 2024, 19(5): e0303987.
- [24] 姚军财, 申静. 基于图像画面信息及其视觉感知的无参考图像质量评价 [J]. *光学精密工程*, 2025, 33(18): 2944-2961.  
YAO J C, SHEN J. No-reference image quality assessment based on image picture information and its visual perception [J]. *Optics and Precision Engineering*, 2025, 33(18): 2944-2961. (in Chinese)
- [25] 万照麟, 刁奕宁, 徐靖奇, 等. 基于视觉感知机制的全景图像质量评价 [J]. *计算机辅助设计与图形学学报*, 2026: 1-12.  
WAN Z L, DIAO Y N, XU J Q, *et al.* Panoramic image quality assessment based on visual perception mechanism [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2026: 1-12. (in Chinese)
- [26] 周荣荣. 基于细节增强与全局结构感知的图像修复方法研究及应用 [D]. 南宁: 广西大学, 2025.  
ZHOU R R. Research and application of image inpainting method based on detail enhancement and global structure perception [D]. Nanning: Guangxi University, 2025. (in Chinese)
- [27] 郑博伟, 李宗辉, 陈锐彬, 等. 基于离散小波变换的细节增强的图像修复优化方法 [J]. *信息技术与信息化*, 2024, (6): 46-51.  
ZHENG B W, LI Z H, CHEN R B, *et al.* Detail enhancement image inpainting optimization method based on discrete wavelet transform [J]. *Information Technology and Informatization*, 2024, (6): 46-51. (in Chinese)
- [28] 万丙辰, 张选德. 基于元学习和失真感知的图像质量评价 [J]. *液晶与显示*, 2024, 39(11): 1519-1531.  
WAN B C, ZHANG X D. Image quality assessment based on meta-learning and distortion perception [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(11): 1519-1531. (in Chinese)
- [29] LIU Z, LIN Y, CAO Y, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012-10022.

- [30] ENGELKE U. Selective attention mechanisms for visual quality assessment [J]. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2015, 98(8): 1681-1688.
- [31] 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.
- [32] 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.
- [33] PONOMARENKO N, JIN L, IEREMEIEV O, *et al.* Image database TID2013: Peculiarities, results and perspectives [J]. *Signal Processing: Image Communication*, 2015, 30: 57-77.
- [34] 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.
- [35] MOORTHY A K, BOVIK A C. Blind image quality assessment: From natural scene statistics to perceptual quality [J]. *IEEE Transactions on Image Processing*, 2011, 20(12): 3350-3364.
- [36] YE P, KUMAR J, KANG L, *et al.* Unsupervised feature learning framework for no-reference image quality assessment [C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012: 1098-1105.
- [37] XU J, YE P, LI Q, *et al.* Blind image quality assessment based on high order statistics aggregation [J]. *IEEE Transactions on Image Processing*, 2016, 25(9): 4444-4457.
- [38] MA K, LIU W, ZHANG K, *et al.* End-to-end blind image quality assessment using deep neural networks [J]. *IEEE Transactions on Image Processing*, 2017, 27(3): 1202-1213.
- [39] 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.
- [40] YAN Q, GONG D, ZHANG Y. Two-stream convolutional networks for blind image quality assessment [J]. *IEEE Transactions on Image Processing*, 2018, 28(5): 2200-2211.
- [41] REN H, CHEN D, WANG Y. RAN4IQA: Restorative adversarial nets for no-reference image quality assessment [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018, 32(1).
- [42] YOU J, KORHONEN J. Transformer for image quality assessment [C]// 2021 IEEE International Conference on Image Processing (ICIP). 2021: 1389-1393.
- [43] ZHU H, LI L, WU J, *et al.* MetaIQA: Deep meta-learning for no-reference image quality assessment [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 14143-14152.
- [44] WU J, MA J, LIANG F, *et al.* End-to-end blind image quality prediction with cascaded deep neural network [J]. *IEEE Transactions on Image Processing*, 2020, 29: 7414-7426.
- [45] LI F, ZHANG Y, COSMAN P C. MMMNet: An end-to-end multi-task deep convolution neural network with multi-scale and multi-hierarchy fusion for blind image quality assessment [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(12): 4798-4811.
- [46] PAN Z, YUAN F, LEI J, *et al.* VCRNet: Visual compensation restoration network for no-reference image quality assessment [J]. *IEEE Transactions on Image Processing*, 2022, 31: 1613-1627.

#### 作者简介:



高敏娟,女,博士,讲师,硕士生导师,2021年于陕西科技大学获得博士学位,主要从事图像处理、图像质量评价方面的研究。E-mail:gaominjuan1984@tynu.edu.cn