差异化建模与感知增强的轻微失真盲图像质量评价

程德强[1]，陈飞宇[1]，宋天舒[1]，寇旗旗[2]
（1. 中国矿业大学信息与控制工程学院，江苏 徐州 221116；2. 中国矿业大学计算机科学与技术学院，江苏 徐州 221116）

摘要：现有主流盲图像质量评价（BIQA）算法在评估轻微失真图像时仍然存在性能瓶颈。其根源在于全局训练策略所引发的特征建模冲突：人类视觉系统（HVS）感知显著失真侧重全局退化，而轻微失真高度依赖局部差异与语义内容。为此，提出基于差异化建模（DM）与感知增强的BIQA框架。首先，引入DM策略构建双分支架构，解耦轻微与显著失真特征表示，缓解跨失真等级建模冲突。其次，设计语义引导感知增强（SGPE）模块，自适应引导模型聚焦语义敏感区域，强化细微质量差异感知。最后，采用失真自适应加权融合（DAWF）策略动态聚合特征，在显著提升轻微失真评估精度的同时，维持全失真区间的预测鲁棒性。在公认基准数据集上的实验表明，所提方法有效突破轻微失真评估瓶颈，并具备强泛化能力。

关键词：盲图像质量评价；差异化建模；轻微失真；感知增强
中图分类号：TP391
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260211

Differential modeling and perception enhancement for slight-distortion blind image quality assessment

Cheng Deqiang[1], Chen Feiyu[1], Song Tianshu[1], Kou Qiqi[2]
1. College of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
2. College of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

Abstract: Despite significant progress in mainstream blind image quality assessment (BIQA) algorithms, performance bottlenecks persist when evaluating slightly distorted images. This stems from feature modeling conflicts induced by global training strategies: the human visual system (HVS) focuses on global degradation for severe distortions but relies heavily on local discrepancies and semantic content for slight ones. To address this, a BIQA framework was proposed based on differential modeling (DM) and perception enhancement. Firstly, a DM strategy constructed a dual-branch architecture to decouple slight and severe distortion representations, mitigating cross-level conflicts. Secondly, a semantic guidance perception enhancement (SGPE) module adaptively directed the model to focus on semantically sensitive regions, enhancing sensitivity to subtle quality variations. Finally, a distortion-adaptive weighted fusion (DAWF) strategy dynamically aggregated features, significantly improving slight-distortion accuracy while maintaining robust prediction across the full distortion spectrum. Experiments on standard benchmark datasets demonstrate that the proposed method effectively overcomes the slight-distortion bottleneck and exhibits strong generalization capability.

Key words: blind image quality assessment, differential modeling, slight-distortion, perception enhancement

收稿日期：2026–04–17；**修回日期：**2026–06–15
通信作者：宋天舒，tianshusong@cumt.edu.cn
基金项目：国家自然科学基金资助项目（No.62501597, No.52204177）；江苏省基础研究计划（No.BK20251651）；江苏省卓越博士后计划（No.2024ZB423）；中国博士后科学基金资助项目（No.2024M763543）
Foundation Items: The National Natural Science Foundation of China (No.62501597, No.52204177), The Basic Research Program of Jiangsu Province (No.BK20251651), The Jiangsu Funding Program for Excellent Postdoctoral Talent (No.2024ZB423), The China Postdoctoral Science Foundation (No.2024M763543)

0 引言

在数字化信息爆炸的今天，视觉数据已占据互联网流量的主导地位，成为人类获取信息最直观、最高效的载体。图像质量的高低直接影响用户的视觉体验，在图像的采集、传输以及处理过程中，常常会发生不同程度的质量退化。这种图像质量退化程度的精确评估，对图像压缩、去噪、增强和生成等领域的发展具有重要的指导意义[1-3]。

深度神经网络（deep neural network, DNN）拥有强大的特征表示能力，已广泛应用于盲图像质量评估。然而，目前主流的基于DNN的盲图像质量评价（blind image quality assessment, BIQA）方法大都采取全局训练的方式训练模型，不区分失真程度，对所有样本进行端到端优化。这类方法在粗粒度质量评价（包含不同失真程度图像）任务中表现优异。当评估范围聚焦于相同失真等级内的细粒度质量差异时，其性能仍存在明显瓶颈[4]。这一局限性源于人类视觉系统（human visual system, HVS）在评估轻微失真图像和显著失真图像时采用的两种不同感知策略[5]。具体而言，在评估轻微失真图像（存在一定程度的高频细节退化或微弱的噪声和伪影等，但图像的整体结构特征保持高度完整）时，由于图像本身内容非常明显，人类视觉系统会主动寻找并定位微小失真等高频视觉信息；对于显著失真图像（存在明显的结构破坏、严重的模糊、强烈的块效应或噪声等，此类失真已经主导了视觉感知体验），则更加依赖全局统计特征与结构完整性。更重要的是，轻微失真的感知质量高度依赖于图像内容语义（如人脸区域的微小模糊比背景区域更不可接受），而显著失真对内容语义则相对鲁棒[6]。因此，采用单一感知策略建模所有失真等级的现有方法，在进行细粒度质量评估任务时，不同失真图像质量感知规则构建时的相互影响导致对细粒度质量变化的敏感性下降。

受上述人类视觉系统差异化感知机制的启发，本文提出一种基于差异化建模与感知增强的盲图像质量评价网络框架，以改善对轻微失真的评价能力。本文的核心设计包含以下两个关键环节。①针对跨失真等级混合建模引发的特征冲突问题，引入差异化建模策略构建双分支并行架构。该策略在特征空间显式解耦轻微与显著失真样本，使网络能够独立学习适配各自感知机制的特征表示，从而缓解显著失真对轻微失真表征优化的压制与干扰。②针对轻微失真分支，设计语义引导感知增强模块。具体而言，首先，采用卷积神经网络[7]优异的局部感受野特性，精准捕获高频细微失真差异；然后，引入冻结预训练ViT[8]的高层语义先验，通过通道注意力机制实现自上而下的特征调制。该设计有效规避了失真对语义完整性的破坏，引导模型将感知资源动态聚焦于语义敏感区域的质量退化。为实现跨分支特征的协同利用，本文采用失真自适应加权融合策略，根据输入图像的实际失真程度自适应切换感知重心，融合双分支的质量感知特征增强模型，在显著强化轻微失真评估精度的同时，维持对全失真区间的泛化鲁棒性，从而胜任通用的图像质量评价任务。

1 相关工作

盲图像质量评估因其不依赖参考图像的特性，在实际工程应用中具有重要价值。传统图像质量评价方法主要依赖专家先验知识构建针对特定失真的手工特征，并结合浅层回归模型完成质量映射[9-11]，其在面对复杂多变的图像内容与混合失真类型时泛化能力受限[12-13]。相比之下，基于深度学习的方法凭借强大的特征表示能力，采用数据驱动范式，能够端到端地自动学习从底层像素到高层感知质量的多层次映射关系，成为当前该领域的主流研究方向[14-15]。其中，基于Transformer的架构因其强大的全局建模能力，逐渐取代传统卷积神经网络（convolutional neural network, CNN）成为BIQA研究的新范式。Golestaneh等[16]和Qin等[17]率先利用Transformer架构优势，通过引入相对排序策略和注意力面板解码器，显著增强了模型对图像结构信息长程依赖的捕捉能力，突破了局部感受野的限制。在此基础上，Shi等[18]借助Transformer架构预测失真图像与参考图像间的误差图，实现了对图像局部退化特征的有效提取与定位，进一步提升了模型对质量差异的评估能力。为进一步优化特征表示，研究者开始关注多维度的特征交互与针对性提取。Yang等[19]提出了在通道与空间双重维度应用注意力机制，通过强化全局上下文与局部细节的特征交互，提升了模型对多尺度失真模式的鲁棒性。Pan等[20]从失真分布建模的角度出发，设计了失真感知卷积模块，旨在提升特征提取对特定失真类型

的针对性。为缓解现有方法在计算效率与特征完备性上的局限，Huo 等[21]利用频域变换缓解了 Transformer 的计算负担，同时通过四元数协同模块捕获精细纹理细节。Yu 等[22]构建了多尺度全局特征融合网络，通过多重注意力机制高效聚合了 CNN 的局部优势与 ViT 的全局语义。尽管上述研究从多个方面取得了显著进展，但是大多采用单一骨干网络统一处理所有失真程度的图像，忽略了不同失真等级下特征提取侧重点的内在差异，导致模型在细粒度质量评价任务中的表现仍显不足[4]。

此外，研究人员在局部特征提取与语义先验引导两个维度进行了广泛探索。在局部特征提取方面，为捕捉局部失真线索，Bosse 等[23]提出了基于图像块的评估框架，通过聚合局部质量图实现质量预测，但忽略了全局上下文依赖。Xu 等[24]设计了局部失真提取器与注入器模块，显式增强了局部失真特征与全局语义的交互。在语义先验引导方面，早期工作尝试利用视觉显著性图[25]加权空间特征，引导模型关注人眼固定点区域。随着预训练模型的发展，Su 等[26]利用 ImageNet 预训练权重提取语义特征，证明了语义破坏程度与质量感知的相关性。近期，基于基础模型的方法进一步提升了语义引导能力，如 Wang 等[27]利用语言–图像对齐先验评估图像“外观与感受”。尽管上述方法证实了语义引导的有效性，但其引导策略多为全局统一，忽略了不同失真程度对语义依赖的差异性。轻微失真图像质量高度敏感于图像内容，而严重失真图像的语义信息被破坏，对图像内容鲁棒。

人类视觉系统对图像质量的感知是一个涉及内容理解、失真识别及两者交互的复杂过程。为模拟这一机制，基于双分支或多分支结构的深度学习方法已成为盲图像质量评估的重要研究范式。早期工作主要基于失真来源或特征属性进行解耦，例如，Zhang 等[28]设计针对合成失真与真实失真的双分支结构以独立提取特征；Wang 等[29]基于视觉掩蔽效应，将内容感知与失真感知解耦，通过交互模块模拟人类视觉系统。近年来，研究者开始从更细粒度的感知认知机制出发模拟人类视觉系统。Li 等[30]借鉴 HVS 的“由粗到细”反馈审视机制，通过多阶段预测逐步细化对局部失真的感知，有效缓解了窄质量范围内的“范围效应”。Xia 等[31]侧重于模拟 HVS 的层级依赖学习机制，通过建模多阶段特征间的空间与通道依赖来聚合多阶段特征。然而，尽管上述方法在特征感知和交互上取得了一定进展，其主干设计或优化策略仍基于全局统一的建模范式，尚未触及失真程度这一关键维度。事实上，人眼对轻微失真的感知更依赖语义完整性，而对显著失真的感知更依赖结构破坏程度。Song 等[32]指出，现有模型在低质量图像上的评估性能显著低于中高质量图像，这主要归因于分布偏移与长尾分布，其通过引入中介增强图像与迭代训练策略，专门强化模型对低质量样本的特征提取能力。该工作有力地证明了不同质量等级的图像需要差异化的处理机制。

尽管上述研究显著推动了图像质量评价领域的发展，但针对轻微失真图像的评估特性仍关注不足。HVS 对质量的感知呈现出显著的失真程度依赖性：在评估轻微失真图像时，HVS 高度敏感于图像语义内容，倾向于捕捉局部细微失真；在评估显著失真图像时，HVS 对内容变化表现出鲁棒性，更关注于整体结构的退化。受此机制启发，本文提出了一种差异化建模与感知增强的盲图像质量评价网络框架。该框架旨在解耦不同失真等级下的质量感知规则，缓解混合训练对轻微失真感知的干扰，同时增强对细微失真差异的感知能力。

2　本文方法

本文提出了差异化建模与感知增强的盲图像质量评价模型。该方法旨在通过缓解跨失真等级的建模冲突和增强细微差异感知来提升对轻微失真图像质量的评估能力。如图 1 所示，为模拟 HVS 对轻微与显著失真差异化感知机制，本文框架采用双分支并行架构：轻微失真分支旨在模拟人眼评估轻微失真时对局部细微差异及图像内容的高度敏感性，增强对轻微失真图像的表征能力；显著失真分支则侧重于提取全局统计失真特性并对图像内容鲁棒的显著失真特征。最后，通过失真自适应聚合模块将两类特征动态融合，在改善轻微失真表征的同时，保持对显著失真的评估性能。具体而言，为缓解跨失真等级的建模冲突，本文引入差异化建模策略，依据失真程度将图像划分为轻微失真与显著失真两组，并基于 MANIQA[19]基线构建双分支架构以学习差异化表征。在轻微失真分支中，本文设计了语义引导感知增强模块注入局部失真特征，从而强化

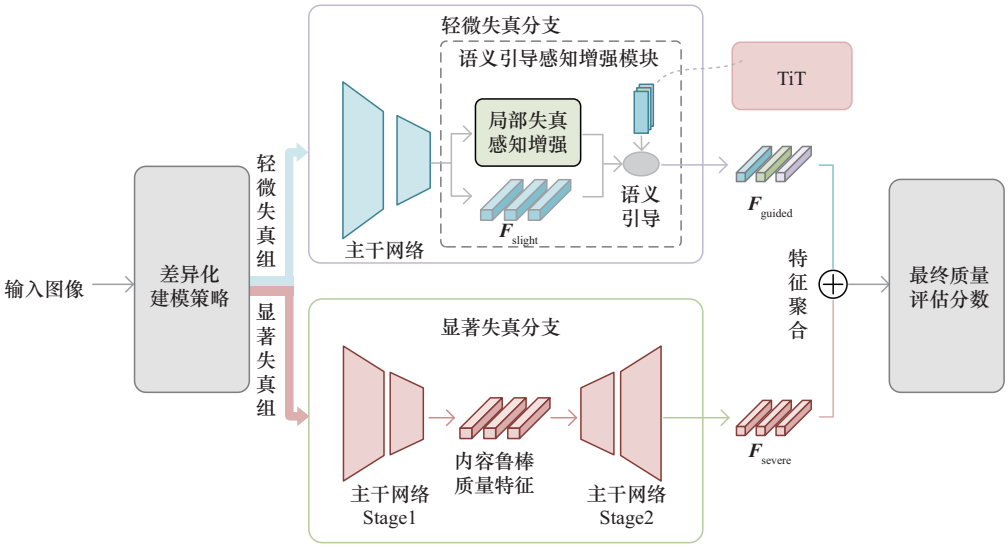

图1　本文方法框架

了模型对细微差异的感知能力。同时，利用预训练ViT的全局语义先验，引导模型关注局部感知增强特征中的语义敏感区域。在显著失真分支中，模型沿用基线网络提取质量表示。该路径擅长捕捉图像整体结构的退化，且经过分组后，显著失真图像中的语义干扰被弱化，使所得特征对图像内容更具鲁棒性。最后，采用失真自适应加权策略根据双分支的特征表示动态学习融合权重，旨在优化特征聚合过程，缓解因特征融合导致的轻微或显著失真的表征能力退化。

2.1　差异化建模

轻微失真与显著失真图像质量分布存在明显差异，前者高度依赖局部细节与语义内容，后者侧重全局统计特性。然而，现有主流方法大多采用混合训练范式，迫使模型在单一优化空间中同时学习这两种相互冲突的特征表示。本文对基线模型MANIQA在KADID-10k[33]上的性能进行了分析，如图2所示。其中，轻微与显著失真图像训练指使用参考图像对KADID-10k进行划分，80%的失真图像（包含5个失真等级）直接作为训练集；仅轻微失真图像训练指使用划分后的训练集中失真等级$k\in\{1,2,3\}$的失真图像作为训练集；仅显著失真图像训练指使用混合训练划分的训练集中失真等级$k\in\{4,5\}$的失真图像作为训练集。将它们分别对应在剩余20%的测试集图像中失真等级$k=1$或$k=5$的测试图像上进行测试，可以看到，基线模型MANIQA在混合训练范式下，无论是在失真等级$k=1$还是在$k=5$的测试集上，均低于仅使用轻微失真图像或显著失真图像训练的结果，在k=1上性能损失尤其明显。

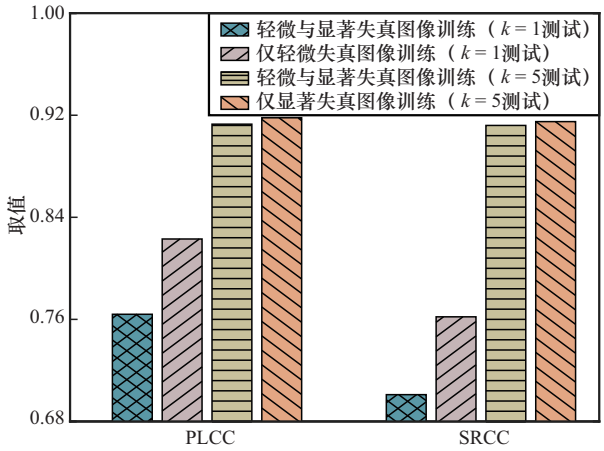

图2　不同训练策略在测试$k=1$和$k=5$子集上的性能对比

为进一步验证混合训练中跨失真等级建模冲突的存在性，本文对具有代表性的$k=1$轻微失真样本与$k=5$显著失真样本在混合训练过程中的梯度优化方向一致性进行了分析。具体而言，在KADID-10k[33]、TID2013[34]和CSIQ[5]这3个数据集上采用混合训练策略训练基线模型，并在训练收敛阶段分别计算$k=1$与$k=5$样本反向传播时的梯度向量。本文分别统计了网络整体参数与高层特征提取层的梯度余弦相似度，余弦相似度越接近1表示优化方向越一致，越接近−1表示优化冲突越强[35]，结果如图3所示。在3个数据集上，两类样本的整

体梯度余弦相似度均为负值，表明其在参数更新时产生了方向对立的梯度信号，存在一定的梯度冲突。同时，高层特征提取层的梯度优化方向同样呈负相关，这进一步验证了轻微失真与显著失真图像在高级语义和全局结构特征依赖上的内在差异，表明单一网络在统一特征空间中难以兼顾两者的感知规则。

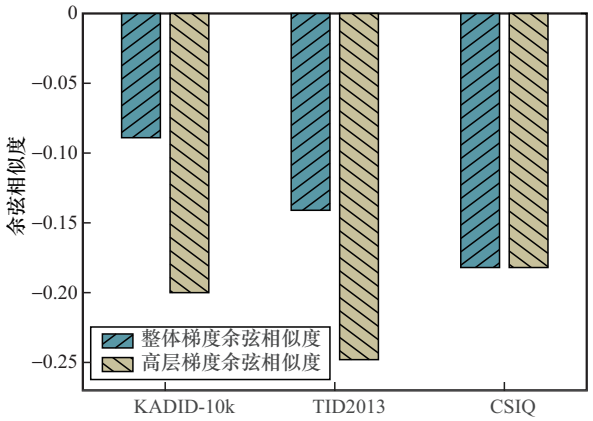

图3　轻微失真样本与显著失真样本的梯度余弦相似度统计结果

基于上述分析，为缓解不同失真等级间感知机制的相互干扰，本文引入差异化建模策略，结合双分支结构分别学习适配各自感知策略的特征表示，从而规避跨失真等级的建模冲突。该策略的优势在于显式地在特征空间解耦轻微与显著失真图像，专注于构建针对不同失真等级的特异性感知模型。

首先，对失真图像进行分组。遵循数据集构建时的失真程度划分，按照失真等级 $k \in \{1,2,3,4,5\}$，将失真图像 $\boldsymbol{I} \in \mathbb{R}^{H \times W \times 3}$ 划分为轻微失真组 $\boldsymbol{I}_{\text{slight}}$ 和显著失真组 $\boldsymbol{I}_{\text{severe}}$，相应失真等级分别为 $k \in \{1,2,3\}$ 和 $k \in \{4,5\}$。

然后，使用分组后的图像对轻微失真分支和显著失真分支进行微调以学习不同的质量规则。通过与双分支结构的协同作用，实现轻微与显著失真在特征学习与优化路径上的显式分离。

2.2　轻微失真分支

（1）轻微失真感知增强

相较于显著失真，轻微失真图像的失真信息较为微弱且难以感知，这对模型的失真判别能力提出了严峻挑战。此类失真通常表现为局部区域内的细微扰动，要求模型具备极高的局部细节敏感度。本文选用的基于 ViT 改进的基线模型 MANIQA 尽管利用大规模预训练的先验知识得到了鲁棒的特征表示，但是难以捕获轻微失真的细微差异（如图 2 所示），为此本文首先采用预训练初始化的卷积神经网络来增强对局部细微差异的感知。

具体而言，本文通过轻微失真分支第一阶段 $\text{Backbone}_{\text{Stage1}}$ 提取轻微失真质量特征 $\boldsymbol{F}_{\text{slight}}$：

$$\boldsymbol{F}_{\text{slight}} = \text{Backbone}_{\text{Stage1}}(\boldsymbol{I}_{\text{slight}}) \tag{1}$$

其中，$\boldsymbol{F}_{\text{slight}} \in \mathbb{R}^{B \times N \times C}$，$B$ 为批量大小，C 为通道数，$N = \dfrac{HW}{P^2}$ 为图像块数量，P 为图像块尺寸。为提取高效的局部失真特征，本文采用预训练 ResNet50 作为局部特征提取骨干。为避免网络层数过深导致局部细节信息退化，本文选取 ResNet50 的 Layer2 输出作为局部失真特征 $\boldsymbol{F}_{\text{CNN}}$。随后，通过 1×1 卷积层和重排列操作将两特征对齐，此时 $\boldsymbol{F}_{\text{CNN}} \in \mathbb{R}^{B \times N \times C}$。由于预训练 ResNet50 使用的预处理方式不同于 MANIQA，所提取特征存在分布差异，为此本文引入带有可学习缩放和偏移参数的 InstanceNorm2d 对 $\boldsymbol{F}_{\text{CNN}}$ 进行归一化，以消除域偏移。

随后，$\boldsymbol{F}_{\text{CNN}}$ 通过多头交叉注意力模块与 $\boldsymbol{F}_{\text{slight}}$ 交互，将局部失真感知信息注入 $\boldsymbol{F}_{\text{slight}}$ 以增强模型对局部细微差异的感知。多头交叉注意力模块包含 h 头，每个头的尺寸为 $d = \dfrac{C}{h}$。使用 3 个不同的线性投影层将 $\boldsymbol{F}_{\text{CNN}}$ 转换为查询矩阵 $\boldsymbol{Q}$，$\boldsymbol{F}_{\text{slight}}$ 转换为键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$，其中 $\boldsymbol{Q} = \{\boldsymbol{Q}_1,\cdots,\boldsymbol{Q}_h\} \in \mathbb{R}^{B \times N \times C}$，$\boldsymbol{K} = \{\boldsymbol{K}_1,\cdots,\boldsymbol{K}_h\} \in \mathbb{R}^{B \times N \times C}$，$\boldsymbol{V} = \{\boldsymbol{V}_1,\cdots,\boldsymbol{V}_h\} \in \mathbb{R}^{B \times N \times C}$，并且 $\boldsymbol{Q}_h,\boldsymbol{K}_h,\boldsymbol{V}_h \in \mathbb{R}^{B \times N \times d}$。最终局部感知增强特征 $\boldsymbol{F}_{\text{enchanced}}$ 可表示为：

$$\begin{cases} \boldsymbol{F}_{\text{c}} = \text{Cat}(\text{Attention}(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\boldsymbol{Q}_3),\cdots, \\ \qquad \text{Attention}(\boldsymbol{Q}_h,\boldsymbol{K}_h,\boldsymbol{V}_h))\boldsymbol{W}_{\text{L}} \\ \boldsymbol{F}_{\text{enhanced}} = \text{Norm}(\boldsymbol{F}_{\text{c}} + \boldsymbol{F}_{\text{slight}}) \end{cases} \tag{2}$$

其中，$\boldsymbol{F}_{\text{c}}$ 是多头交叉注意力模块的输出特征，$\text{Attention}(\boldsymbol{Q}_h,\boldsymbol{K}_h,\boldsymbol{V}_h))\boldsymbol{W}_{\text{L}} = \text{softmax}\left(\dfrac{\boldsymbol{Q}_h\boldsymbol{K}_h^{\text{T}}}{\sqrt{d}}\right)\boldsymbol{V}_h$，$\text{Cat}(\cdot)$ 表示拼接操作，$\boldsymbol{W}_{\text{L}}$ 是线性映射层的权重，Norm 是层归一化。通过上述过程，最终可以得到轻微失真信息感知增强的质量特征。

（2）语义引导

轻微失真图像质量对图像内容高度敏感，然而在图像质量评价任务微调过程中，模型往往倾向于优化质量回归目标，导致通用的语义理解能力不可

避免地发生退化。为解决这一问题，本文利用冻结的预训练 ViT 提取鲁棒的全局语义先验，提取其 CLS token 作为全局语义描述符，并结合通道注意力对经过局部增强的质量特征进行 $\boldsymbol{F}_{\text{enhanced}}$ 语义引导，增强对语义关键区域的失真感知能力，确保质量评估策略与轻微失真场景下 HVS 的内容敏感性保持一致。

具体来讲，将输入图像 $\boldsymbol{I}_{\text{slight}}$ 输入预训练 ViT 中，将最后一层 Transformer Block 输出的 CLS token 记为 $\boldsymbol{V}_{\text{s}} \in \mathbb{R}^{C}$，并将其作为全局语义描述符。通过一个轻量级映射网络，语义引导的自适应权重 $\boldsymbol{g} \in [0,1]^{C}$ 可表示为：

$$\boldsymbol{g} = \text{Sigmod}\left(\boldsymbol{W}_2\,\text{ReLU}\left(\boldsymbol{W}_1\boldsymbol{V}_{\text{s}}\right)\right) \tag{3}$$

其中，$\boldsymbol{W}_1 \in \mathbb{R}^{\frac{C}{4}\times C}$、$\boldsymbol{W}_2 \in \mathbb{R}^{C\times\frac{C}{4}}$ 是两个不同线性映射层的权重。首先，利用 $\boldsymbol{W}_1$ 将输入语义特征从 C 维压缩至 $\frac{C}{4}$ 维，通过低维嵌入学习不同通道之间的相关性。随后，利用 $\boldsymbol{W}_2$ 将特征恢复至原始维度，并结合 Sigmoid 函数生成通道注意力权重 $\boldsymbol{g}$。将自适应权重 $\boldsymbol{g}$ 点乘局部增强后的质量特征 $\boldsymbol{F}_{\text{enhanced}}$，重校准其的通道响应。最后，将语义引导模块的输出记为 $\boldsymbol{F}_{\text{guided}}$，其计算式可表示为：

$$\boldsymbol{F}_{\text{guided}} = \boldsymbol{g}\cdot\boldsymbol{F}_{\text{enhanced}} + \boldsymbol{F}_{\text{enhanced}} \tag{4}$$

在完成局部特征的语义引导增强后，该表征被输入基线网络的后续回归模块以预测质量分数，由此构建出端到端的轻微失真感知分支。该流程有效模拟了 HVS 针对轻微失真的核心感知机制：对局部高频残差的精细捕获，以及对图像内容的强依赖性。

2.3 显著失真分支

人类视觉系统在评估显著失真图像时，主要聚焦于全局统计特征的破坏程度，而对图像内容变化表现出较强的鲁棒性。受此启发，本文采用基于 ViT 改进的基线模型[1]构建显著失真感知分支。得益于 Transformer 架构的全局感受野，该分支能够有效建模图像整体结构的退化模式。同时，在差异化建模策略下，显著失真样本的特征空间由失真信号主导，自然削弱了网络对高层语义内容的依赖，从而实现了兼顾内容鲁棒性与表征效率的质量特征提取。

具体而言，本文首先将分组后的显著失真样本输入主干网络 Stage1，获取内容鲁棒的多尺度质量特征表示。其次，在主干网络 Stage2，利用 Swin Transformer Block 进一步捕获严重失真带来的整体结构破坏与区域级质量退化。本文采用均方误差（MSE）损失函数进行优化，学习显著失真图像的质量退化规律。损失函数表示为：

$$L = \frac{1}{M}\sum_{i=1}^{M}(y_i - s_i)^2 \tag{5}$$

其中，M 为显著失真样本的数量，s_i 是第 i 张图像的预测分数，y_i 为对应的真实标签。

本文取出模型质量回归前的特征表示作为显著失真质量特征 $\boldsymbol{F}_{\text{severe}}$，该特征对图像内容变化具有鲁棒性，并有效捕捉了全局结构的破坏程度，为后续的特征融合提供了可靠的失真表示。

2.4 失真自适应加权融合

为增强模型在复杂失真场景下的适应性，本文引入失真自适应加权融合策略，旨在模拟 HVS 针对不同失真程度动态调整感知策略的机制。通过动态学习融合权重，实现对轻微与显著失真分支贡献度的自适应调节。

具体而言，设失真自适应加权融合的权重为 α，其表达式为：

$$\alpha = \text{Linear}\left(\text{pool}\left(\text{cat}\left(\boldsymbol{B}_{\text{slight}}(\boldsymbol{I}_i),\boldsymbol{B}_{\text{severe}}(\boldsymbol{I}_i)\right)\right)\right) \tag{6}$$

其中，Linear 为输入维度 C、输出维度 1 的全连接层，pool 为自适应平均池化操作，$\boldsymbol{B}_{\text{slight}}$ 为冻结的轻微失真分支，$\boldsymbol{B}_{\text{severe}}$ 为冻结的显著失真分支。最终预测的质量分数 s_i 可表示为：

$$s_i = \text{MLP}\left(\alpha\cdot\boldsymbol{B}_{\text{slight}}(\boldsymbol{I}_i) + (1-\alpha)\cdot\boldsymbol{B}_{\text{severe}}(\boldsymbol{I}_i)\right) \tag{7}$$

其中，MLP 表示用于质量分数回归的多层感知机。

这种自适应融合策略避免了单一固定权重的局限性，使模型能够根据图像的失真类型灵活切换感知侧重点，从而在保持显著失真评估性能的同时，大幅提升对轻微失真场景的细粒度评价能力。

3 实验结果与分析

3.1 实验细节和评估指标

本文在 4 个合成失真图像数据集上进行实验，包括 KADID-10k[33]、TID2013[34]、CSIQ[5] 和 LIVE[36]。在训练阶段，本文根据数据集的失真程度进行分类。对于失真等级 $k \in \{1,2,3,4,5\}$，选择 $k \in \{1,2,3\}$ 的失真图像构成轻微失真组，$k \in \{4,5\}$ 构成显著失真组。由于 LIVE 数据集无

统一的失真程度分类，因此本文选择每种失真类型对应平均主观得分差（differential mean opinion score，DMOS）最低（质量最好）的 3 张失真图像构成轻微失真组，其余构成显著失真组。在对轻微失真图像测试时，为了更加直观地展现模型对轻微失真图像的细粒度评估能力，本文缩小了测试区间，仅在 $k = 1$ 或 LIVE 数据集 DMOS 排序最低的失真图像上进行测试。遵循先前工作的设定，本文使用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’s rank correlation coefficient, SRCC）作为评估模型性能的指标。

本文方法基于 PyTorch 框架，使用 NVIDIA RTX 3090 显卡进行实验，图像尺寸首先被调整为 256 像素×256 像素。本文随机裁剪一个大小为 224 像素×224 像素的图像块作为训练时输入，选取中心同样尺寸的图像块用于测试。对于轻微失真分支和显著失真分支的训练，本文使用随机梯度下降优化器，初始学习率为 1×10^{-5}，学习率更新使用重量衰减为 1×10^{-5} 的 Adam 优化器、参数 $T_{\max}$ 和 $\text{eta}_{\min}$ 设置为 50 和 0 的余弦退火学习率更新策略。对于失真自适应加权融合部分和最终的质量回归器，初始学习率为 1×10^{-3}，当 10 个时期的损失不减少时，本文通过乘以 0.3 的权重降低学习率，直到小于 1×10^{-5}。为缓解双分支特征提取过程中可能出现的过拟合问题，在 TID2013、CSIQ 和 LIVE 数据集上，本文对整体网络参数进行了微调，以提升模型在不同数据分布下的泛化能力。

遵循先前 BIQA 工作的实验准则，4 个合成失真图像数据集上的所有实验均使用相同的设置。本文根据参考图像随机划分训练集与测试集，其中 80% 用于训练，20% 用于测试。为降低数据划分随机性带来的影响，所有实验均执行 10 次随机划分，并最终将 PLCC 和 SRCC 的平均值作为模型性能的评估结果。

3.2　对比实验

本文方法旨在改善对轻微失真的细粒度评估能力，因此本文首先针对轻微失真图像进行了实验。由于很少有研究关注这方面，遵照原文实验设定，本文重新训练了 14 种流行的方法，包括传统方法 BLINDS-II[9]、BRISQUE[10]、HOSA[11]，以及基于深度学习的方法 TReS[16]、DEIQT[17]、TempQT[18]、DACNN[20]、WAQNIQA[21]、MGFFNeT[22]、LoDa[24]、HyperIQA[26]、DBCNN[28]、REQA[30]、HDLaQA[31]。所有对比算法的性能均为使用 10 个不同随机数种子重复 10 次实验后的平均值。其中，对于 WAQNIQA[21]、MGFFNeT[22]、LoDa[24]、REQA[30] 和 HDLaQA[31]，为保证实验设置的一致性，本文对所有对比方法采用统一的数据预处理和测试流程，遵循 3.1 节中的实验设定。

表 1 展示了本文方法的轻微失真分支以及最终模型在 4 个数据集的轻微失真图像上的性能对比。可以看到，除 LIVE 数据集的 PLCC 指标外，所提出的轻微失真分支在其余评价指标上均优于对比方法。另外，可以注意到，MGFFNeT 同样在此任务上展现出较为优异的成绩。这是因为其采用的多尺度 CNN 和 ViT 混合增强多重注意力聚合机制，一定程度上符合轻微失真图像需要的对局部细微差异和内容敏感的特性。然而，其忽略了显著失真对轻微失真差异感知的影响，因此本文采用的差异化建模策略以及内容引导感知增强模块以更少的计算资源实现了更优的效果，有效捕获了评估轻微失真质量所需的信息。另外，表 1 同样展现了聚合双分支特征的最终模型在 4 个主流数据集上的性能，可见所采用的聚合策略有效保持了其在轻微失真图像评估任务中的优势。

理想的图像质量评价指标不仅应该适用于轻微失真图像，还应具备良好的整体质量评估能力。因此本文在表 2 进一步总结了本文最终模型在完整数据集的性能结果。可以看到，所提方法在 LIVE 以及更有挑战性的 TID2013 和 KADID-10k 数据集上均取得了最佳的性能，并在 CSIQ 上展现出较强的竞争力。这表明本文采用的差异化建模策略有效缓解了轻微失真与显著失真同时建模的冲突，增强了对不同程度失真特性的表征能力，从而在通用 IQA 任务中展现出优异的鲁棒性和泛化能力。

另外，通过对比表 1 和表 2 中各算法的表现可以发现，在损失轻微失真分支部分性能的情况下，本文方法不仅在整体性能优异的数据集上展现出卓越的轻微失真评估能力，并且在 CSIQ 数据集上，虽整体性能略逊于 MGFFNeT，但在轻微失真子集上，仍然展现出与其相当的性能。这一结果进一步凸显了本文方法在轻微失真评估中的有效性。

表 1 在 4 个数据集的轻微失真图像上的性能对比

方法	LIVE		CSIQ		TID2013		KADID-10k	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
BLINDS-II[9]	0.692	0.496	0.053	0.007	0.313	0.299	0.148	0.151
BRISQUE[10]	0.611	0.414	0.054	0.017	0.274	0.245	0.022	0.001
HOSA[11]	0.737	0.533	0.050	0.112	0.295	0.334	0.183	0.134
TempQT[18]	0.866	0.733	0.288	0.381	0.678	0.475	—	—
DBCNN[28]	**0.877**	0.708	0.341	0.363	0.586	0.458	0.601	0.525
HyperIQA[26]	0.757	0.624	0.272	0.333	0.590	0.472	0.512	0.443
DACNN[20]	0.833	0.695	0.315	0.346	0.678	0.541	0.560	0.522
TReS[16]	0.762	0.645	0.242	0.293	0.661	0.483	0.518	0.489
DEIQT[17]	0.850	0.724	0.273	0.306	0.690	0.520	0.532	0.474
LoDa[24]	<u>0.865</u>	0.747	0.267	0.335	0.765	0.603	0.719	0.662
REQA[30]	0.655	0.545	0.054	0.148	0.731	0.618	0.538	0.472
HLaQA[31]	0.534	0.386	0.161	0.042	0.694	0.533	0.441	0.394
MGFFNeT[22]	0.857	0.774	<u>0.408</u>	0.451	<u>0.809</u>	0.641	0.779	0.703
WAQNIQA[21]	0.784	0.616	0.389	0.430	0.762	0.591	0.706	0.672
本文(slight）	0.845	**0.818**	**0.451**	**0.502**	**0.877**	**0.744**	**0.827**	**0.768**
本文（whole）	0.830	<u>0.778</u>	0.404	<u>0.453</u>	0.804	<u>0.669</u>	<u>0.790</u>	<u>0.737</u>

注：粗体表示最优值，下划线表示次优值。

表 2 在 4 个数据集的整体图像上的性能对比

方法	LIVE		CSIQ		TID2013		KADID-10k	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
BLINDS-II[9]	0.923	0.921	0.778	0.741	0.601	0.533	0.533	0.516
BRISQUE[10]	0.915	0.914	0.796	0.734	0.613	0.551	0.522	0.497
HOSA[11]	0.941	0.943	0.829	0.777	0.696	0.648	0.639	0.629
TempQT[18]	0.973	0.970	0.937	0.930	0.901	0.885	—	—
DBCNN[28]	0.968	0.969	0.943	0.936	0.834	0.808	0.872	0.873
HyperIQA[26]	0.968	0.966	0.935	0.919	0.856	0.833	0.866	0.868
DACNN[20]	0.973	0.972	0.945	0.931	0.885	0.867	0.880	0.877
TReS[16]	0.972	0.97	0.933	0.920	0.881	0.854	0.858	0.859
DEIQT[17]	0.973	0.973	0.953	0.941	0.890	0.869	0.867	0.867
LoDa[24]	0.974	0.972	0.954	0.941	0.911	0.895	0.920	0.916
REQA[30]	0.942	0.935	0.889	0.862	0.873	0.855	0.825	0.821
HDLaQA[31]	0.947	0.942	0.929	0.913	0.873	0.848	0.820	0.834
MGFFNeT[22]	0.967	0.966	**0.960**	**0.954**	0.931	0.919	0.934	**0.934**
WAQNIQA[21]	0.961	0.959	0.957	0.948	0.916	0.900	**0.937**	0.932
本文（whole）	**0.979**	**0.977**	0.954	0.947	**0.933**	**0.923**	0.935	**0.934**

注：粗体表示最优值。

3.3　跨数据集性能对比

为全面评估所提方法在未知数据分布下的泛化能力，本文严格遵循3.1节的实验设定开展跨数据集评测。首先，针对模型核心关注的轻微失真场景，本文将训练完成的轻微失真分支独立部署于目标测试集的轻微失真子集上。表3展示了跨数据集轻微失真图像的SRCC性能对比。可以看出，本文方法在3个目标数据集上均取得最优性能，尤其在LIVE与TID2013数据集上展现出显著优势（SRCC分别领先次优方法约39.3%与26.2%）。该结果表明，得益于差异化建模策略对跨等级特征冲突的显式解耦，以及语义引导感知增强模块对局部细微退化的敏感性，模型能够有效迁移至未见轻微失真分布中。

表3　跨数据集轻微失真图像的SRCC性能对比

方法	TID2013		KADID-10k	CSIQ
	LIVE	KADID	TID2013	LIVE
HyperIQA[26]	0.309	0.217	0.263	0.517
DACNN[20]	0.304	0.241	0.290	0.513
TReS[16]]	0.313	0.230	0.308	0.567
DEIQT[17]]	0.276	0.212	0.311	0.500
DBCNN[28]	0.299	0.208	0.257	0.540
MGFFNeT[22]	0.338	0.452	0.351	**0.620**
本文（slight）	**0.471**	**0.479**	**0.443**	0.616

进一步地，为验证双分支协同架构在通用图像质量评估任务中的整体泛化性，本文在完整测试集上进行了跨库性能评估，结果如表4所示。可以看到，本文方法依然保持了极强的竞争力，在多个目标数据集上均取得了最优泛化性能。这一现象可归因于以下协同机制：其一，显著失真分支通过全局统计特性提取与内容鲁棒表征学习，为模型提供了跨数据集稳定的结构退化先验，有效抑制了内容分布偏移带来的性能波动；其二，失真自适应加权融合策略能够根据输入图像的实际失真程度动态校准双分支贡献度，使模型在未见场景中自动切换至最匹配的感知通路。这一结果有力凸显了本文方法的鲁棒性和强大的域适应能力，特别是在处理数据集之间的分布变化方面。轻微失真和整体数据集设置中的一致增益验证了本文方法在不同IQA场景中的普遍适用性。

表4　跨数据集整体失真图像的SRCC性能对比

方法	TID2013		KADID-10k	CSIQ
	LIVE	KADID	TID2013	LIVE
HyperIQA[26]	0.851	0.560	0.690	0.927
DACNN[20]	0.868	0.560	0.670	0.915
TReS[16]]	0.847	0.508	**0.715**	0.931
DEIQT[17]]	0.892	0.577	0.694	0.932
DBCNN[28]	0.853	0.541	0.633	0.917
MGFFNeT[22]	0.901	0.705	0.712	0.934
本文（whole）	**0.907**	**0.718**	0.697	**0.939**

3.4　消融实验

为系统验证差异化建模（DM）策略与语义引导感知增强模块（SGPE）在轻微失真评估中的有效性，本文在4个合成失真数据集上开展了消融实验。为了清晰地展现差异化建模的影响，实验分别从单一轻微失真分支与完整双分支架构两个维度展开评估。此外，为验证将轻微失真与显著失真间的过渡失真等级 $k = 3$ 划分为轻微失真组的合理性，本文对不同分组方式进行了消融实验。

（1）轻微失真分支消融

如图4所示，首先，初始化在ImageNet上预训练的基线模型。其次，在不做任何修改的情况下在目标数据集上对其进行微调，记为“Baseline”。再次，在基线架构上引入差异化建模策略，仅对轻微失真图像特征空间进行建模，结果对应图4中“w/DM”。最后，在轻微分支中进一步集成SGPE模块，对应“w/DM+SGPE”。

可以观察到，引入差异化建模策略后，模型在所有数据集上的性能均取得稳定提升（SRCC平均增幅约10.8%）。该增益主要源于DM对跨失真等级优化冲突的有效抑制：通过构建独立于显著失真的特征优化域，轻微分支得以摆脱混合训练中强梯度信号的压制，专注于高频细微残差与局部纹理变化的建模。进一步集成SGPE模块后，在4个数据集上均实现了一致的性能增益趋势。这是因为SGPE利用CNN的密集局部感受野和ViT高层语义特征构建的自上而下注意力调制通路，有效模拟了HVS评估轻微失真图像时关注细微失真差异并对图像内容敏感的特性，从而提升了对轻微失真图像的质量评估能力。

（2）双分支架构消融实验

为验证上述组件在通用IQA任务中的协同价

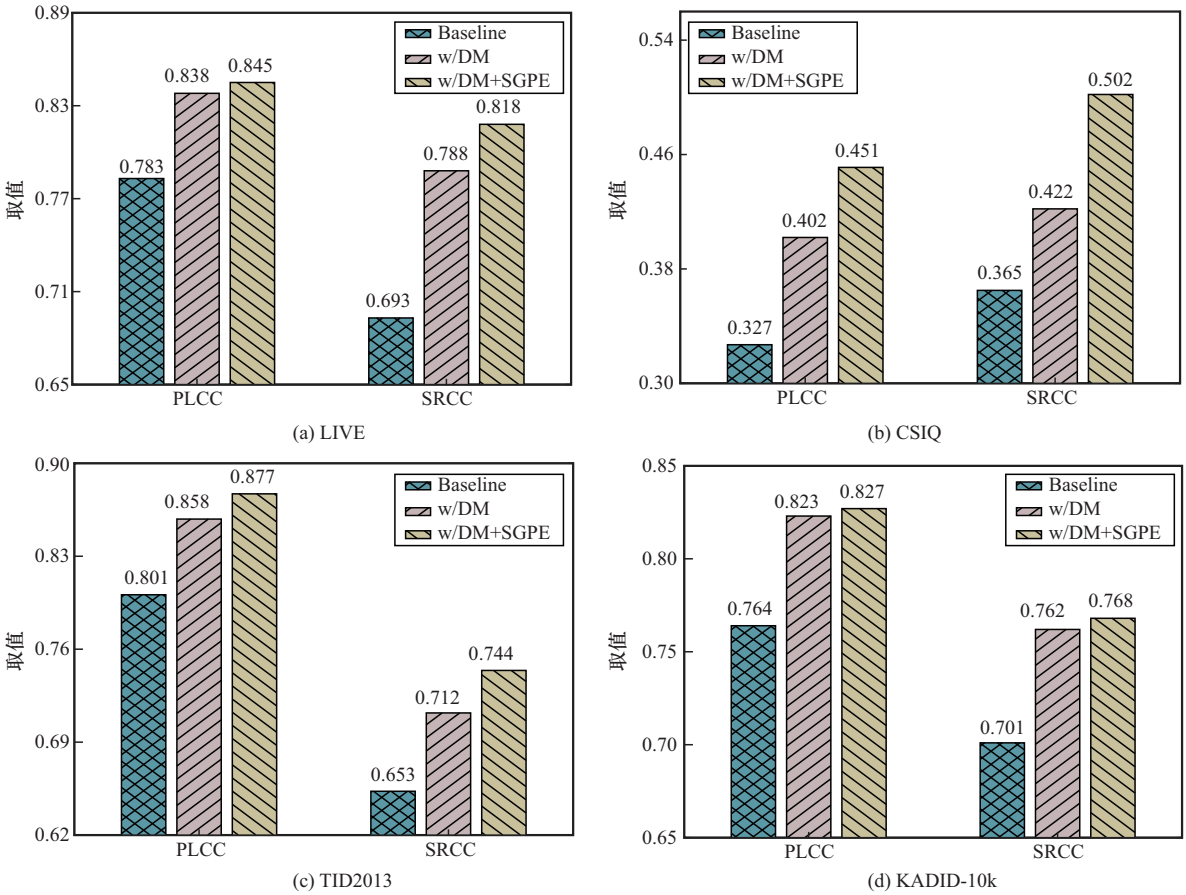

图4　本文方法在4个数据集的轻微失真图像上的消融实验结果

值，实验分别在轻微失真子集与整体测试集上评估了 Baseline、引入 DM 与失真自适应加权融合（DAWF）的中间模型，以及最终完整模型的性能。如表5所示，在引入差异化建模与失真自适应加权融合策略融合双分支特征后，在所有数据集上均实现稳定的增益。这表明通过轻微与显著失真分支分别对轻微与显著失真图像建模，在特征空间解耦两者的特征表示，有效缓解了跨失真等级建模冲突，增强了对轻微与显著失真图像的表征能力，使其在通用IQA任务中受益。

可以观察到，在引入语义引导感知增强模块后，模型整体性能得到进一步提升。这是因为SGPE强化后的轻微分支特征为自适应融合头提供了更高置信度的参考信号，使失真自适应加权融合策略能够根据输入图像的实际失真程度自适应切换感知重心，动态融合双分支的质量感知特征，保持轻微失真分支的增益，从而提升其在通用IQA任务中的评估性能。双分支结构组件在轻微失真图像子

表5　　**双分支结构在整体数据集和轻微失真子集上的消融实验结果**

数据集	Baseline	DM	SGPE	DAWF	LIVE		CSIQ		KADID-10k	
					PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
整体数据集	√				0.974	0.973	0.953	0.943	0.930	0.930
	√	√		√	0.977	0.976	0.955	0.946	0.934	0.932
	√	√	√	√	0.979	0.977	0.956	0.949	0.935	0.934
轻微失真子集	√				0.783	0.693	0.327	0.365	0.764	0.701
	√	√		√	0.810	0.769	0.336	0.381	0.784	0.732
	√	√	√	√	0.830	0.778	0.404	0.453	0.790	0.737

集中测试性能的一致性提升，充分验证了本文提出的差异化建模策略、语义引导感知增强模块和失真自适应加权融合的范式的有效性。这表明，所提架构不仅能在细粒度的轻微失真子集上突破评估瓶颈，还能在通用质量评价场景中保持领先的预测精度与鲁棒性。

（3）$k=3$ 失真等级不同分组方式消融实验

由于失真等级 $k=3$ 为轻微失真到显著失真之间的过渡等级，为进一步验证失真等级划分的合理性，本文额外对 $k=3$ 图像的归属问题进行了分析。除本文采用的划分方式外，进一步构建了另一种划分策略，将 $k=3$ 图像划入显著失真组（即轻微失真组：$k\in\{1,2\}$；显著失真组：$k\in\{3,4,5\}$），并在相同实验设置下于 LIVE 数据集上进行训练与测试，实验结果如表 6 所示。

表 6　$k=3$ 失真等级不同归属方式下的性能比较

归属方式	轻微失真子集		整体数据集	
	PLCC	SRCC	PLCC	SRCC
属于显著失真组	0.814	0.765	0.978	0.977
属于轻微失真组	0.830	0.778	0.979	0.977

可以观察到，当 $k=3$ 被划入显著失真组后，虽然对整体测试集性能影响较小，但在轻微失真测试子集上的性能出现一定程度的下降。这表明，尽管 $k=3$ 相较于 $k=1,2$ 包含更明显的失真信息，但其质量感知特征仍更接近轻微失真图像。将其纳入轻微失真组能够为模型提供更加丰富的细粒度质量变化样本，从而有助于学习轻微失真场景下的质量感知规律。因此，本文最终采用 $k\in\{1,2,3\}$ 作为轻微失真组，$k\in\{4,5\}$ 作为显著失真组。

3.5　计算复杂度与推理时间分析

为评估本文方法的实际可用性，本节从模型参数量及单张图像推理时间两个维度，对所提方法与主流对比方法进行了系统性对比分析，如表 7 所示。所有模型均在相同的硬件环境（NVIDIA RTX 3090）下进行，测试时输入图像随机裁剪为 224 像素×224 像素，批处理大小设为 1，以模拟单张图像的实时推理场景。推理时间的测量基于端到端测量原则，即记录从输入原始图像到输出最终单一质量分数所需的总时间。对于需多次随机裁剪取均值的方法，本文遵循原文中的测试设定进行实验，以真实反映实际应用中的计算成本。从表 7 可以看出，本文采用的差异化建模策略和 SGPE 模块虽使模型参数量与推理时间略有上升，但得益于单次前向传播的架构优势，本文方法在整体推理速度上仍显著优于依赖多轮采样的对比算法，展现出优异的运行效率与工程应用潜力。

表 7　计算复杂度与推理时间对比

方法	模型参数量	推理时间/ms
DEIQT*[17]	24.0×10^6	378.8
TempQT*[18]	87.6×10^6	363.3
TReS*[16]	152.5×10^6	1 824.6
WAQNIQA*[21]	133.8×10^6	694.9
HyperIQA*[26]	27.3×10^6	243.6
DBCNN[28]	15.3×10^6	18.2
本文（slight）	249.0×10^6	65.3
本文（whole）	385.1×10^6	86.6

注：*表示需采用多轮随机裁剪取均值的方法。

4　结束语

本文提出了一种基于差异化建模与语义引导感知增强的盲图像质量评价方法。该方法通过差异化建模策略构建双分支架构，在特征空间显式解耦轻微与显著失真表征，从根本上缓解了跨等级混合优化引发的梯度竞争与感知规则冲突；同时，引入语义引导感知增强模块协同局部细节与全局语义先验，精准模拟人类视觉系统对轻微失真图像的感知特性。大量的实验结果表明，所提方法显著提升了对轻微失真图像的评估能力，在数据集内和跨数据集测试中均实现了最先进的性能。

参考文献：

[1] 寇旗旗, 刘规, 江鹤, 等. 基于多域信息增强的轻量级图像超分辨率网络[J]. 通信学报, 2025, 46(4): 144-159.
Kou Q Q, Liu G, Jiang H, et al. Lightweight image super-resolution network based on Multi-domain information enhancement[J]. Journal on Communications, 2025, 46(4): 144-159.

[2] Zheng P, Jiang H, Sun F Q, et al. HMSR: hyper complex-guided mamba for fine-texture coupling in single image super-resolution[J]. Pattern Recognition, 2026, 178: 113403.

[3] 程德强, 姬广凯, 张皓翔, 等. 基于多粒度融合和跨尺度感知的跨模态行人重识别[J]. 通信学报, 2025, 46(1): 108-123.
Cheng D Q, Ji G K, Zhang H X, et al. Cross-modality person re-

identification based on multi-granularity fusion and cross-scale perception[J]. Journal on Communications, 2025, 46(1): 108-123.

[4] Zhang X F, Lin W S, Huang Q M. Fine-grained image quality assessment: a revisit and further thinking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5): 2746-2759.

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

[6] Zhang Y J, Yang Q, Xu Y L, et al. Perception-guided quality metric of 3D point clouds using hybrid strategy[J]. IEEE Transactions on Image Processing, 2024, 33: 5755-5770.

[7] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.

[8] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//Proceedings of the 8th International Conference on Learning Representations. Washington DC: ICLR, 2021: 1-21.

[9] 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.

[10] 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.

[11] Xu J T, Ye P, Li Q H, et al. Blind image quality assessment based on high order statistics aggregation[J]. IEEE Transactions on Image Processing, 2016, 25(9): 4444-4457.

[12] 万国挺, 王俊平, 李锦, 等. 图像拼接质量评价方法[J]. 通信学报, 2013, 34(8): 76-81.
Wan G T, Wang J P, Li J, et al. Method for quality assessment of image mosaic[J]. Journal on Communications, 2013, 34(8): 76-81.

[13] 蒋铭, 马兆丰, 辛宇, 等. 基于DWT和视觉加权的图像质量评价方法研究[J]. 通信学报, 2011, 32(9): 129-136.
Jiang M, Ma Z F, Xin Y, et al. Image quality evaluation method base on digital wavelet transform and vision weighted[J]. Journal on Communications, 2011, 32(9): 129-136.

[14] 张玉龙, 潘庆一, 杨培. 手工LBP特征增强的无参考图像质量评价[J]. 计算机辅助设计与图形学学报, 2026, doi: link.cnki.net/urlid/11.2925.TP.20250625.1651.004.
Zhang Y L, Pan Q Y, Yang P. Boosting no-reference image quality assessment using handcrafted LBP features[J]. Journal of Computer-Aided Design & Computer Graphics, 2026, doi: link.cnki.net/urlid/11.2925.TP.20250625.1651.004.

[15] 邹良涛, 蒋刚毅, 郁梅, 等. 基于张量域感知特征的无参考高动态范围图像质量评价[J]. 计算机辅助设计与图形学学报, 2018, 30(10): 1850-1858.
Zou L T, Jiang G Y, Yu M, et al. No-reference image quality assessment of high dynamic range image based on tensor domain perceptual features[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(10): 1850-1858.

[16] Golestaneh S A, Dadsetan S, Kitani K M. No-reference image quality assessment via transformers, relative ranking, and self-consistency[C]//Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2022: 3989-3999.

[17] Qin G Y, Hu R Z, Liu Y T, et al. Data-efficient image quality assessment with attention-panel decoder[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(2): 2091-2100.

[18] Shi J S, Gao P, Smolic A. Blind image quality assessment via transformer predicted error map and perceptual quality token[J]. IEEE Transactions on Multimedia, 2024, 26: 4641-4651.

[19] Yang S D, Wu T H, Shi S W, et al. MANIQA: multi-dimension attention network for no-reference image quality assessment[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2022: 1190-1199.

[20] Pan Z Q, Zhang H, Lei J J, et al. DACNN: blind image quality assessment via a distortion-aware convolutional neural network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32 (11): 7518-7531.

[21] Huo Y J, Huang G H, Yu Z W, et al. WAQNIQA: wavelet-augmented quaternion network for no-reference image quality assessment[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2026, PP(99): 1-14.

[22] Yu Q Q, Chen Y H, Ma J L, et al. Multi-scale and global feature fusion network with multiple attentions for no-reference image quality assessment[J]. Expert Systems with Applications, 2026, 307: 131011.

[23] 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.

[24] Xu K M, Liao L, Xiao J, et al. Boosting image quality assessment through efficient transformer adaptation with local feature enhancement[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 2662-2672.

[25] Yang S, Jiang Q, Lin W, et al. SGDNet: an end-to-end saliency-guided deep neural network for no-reference image quality assessment[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 1383-1391.

[26] Su S L, Yan Q S, Zhu Y, et al. Blindly assess image quality in the wild guided by a self-adaptive hyper network[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 3664-3673.

[27] Wang J, Chan K C K, Loy C C. Exploring clip for assessing the look and feel of images[J]. Proceedings of the AAAI Conference on Artifi-

cial Intelligence, 2023, 37(2): 2555-2563.

[28] Zhang W X, Ma K D, 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, 2020, 30(1): 36-47.

[29] Wang X Q, Xiong J, Lin W S. Visual interaction perceptual network for blind image quality assessment[J]. IEEE Transactions on Multimedia, 2023, 25: 8958-8971.

[30] Li B H, Huo F S. REQA: Coarse-to-fine assessment of image quality to alleviate the range effect[J]. Journal of Visual Communication and Image Representation, 2024, 98: 104043.

[31] Xia J L, He L H, Gao X B, et al. Blind image quality assessment based on hierarchical dependency learning and quality aggregation[J]. Neurocomputing, 2024, 585: 127621.

[32] Song T S, Li L D, Chen P F, et al. Blind image quality assessment for authentic distortions by intermediary enhancement and iterative training[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(11): 7592-7604.

[33] Lin H H, Hosu V, Saupe D. KADID-10k: a large-scale artificially distorted IQA database[C]//Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). Piscataway: IEEE Press, 2019: 1-3.

[34] Ponomarenko N, Ieremeiev O, Lukin V, et al. Color image database TID2013: peculiarities and preliminary results[C]//Proceedings of European Workshop on Visual Information Processing. Piscataway: IEEE Press, 2013: 106-111.

[35] Liu B, Liu X, Jin X, et al. Conflict-averse gradient descent for multi-task learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 18878-18890.

[36] 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.

[作者简介]



程德强（1979– ），男，博士，中国矿业大学教授、博士生导师，主要研究方向为机器视觉与模式识别、图像智能检测与信息处理。



陈飞宇（2000– ），男，中国矿业大学硕士生，主要研究方向为图像质量评价。



宋天舒（1994– ），男，博士，中国矿业大学副教授，主要研究方向为图像质量评价、人工智能+物理、表面生长动力学。



寇旗旗（1988– ），男，博士，中国矿业大学副教授，主要研究方向为图像增强与复原、智能检测与模式识别。