

DOI:10.11784/tdxbz202404015

基于显著性双流分层感知的 NR-IQA 方法

史再峰^{1,2}, 康泰¹, 王云峰¹, 肖云泽¹, 罗韬³

(1. 天津大学微电子学院, 天津 300072; 2. 天津市成像与感知微电子技术重点实验室, 天津 300072;

3. 天津大学智能与计算学部, 天津 300072)

摘要: 人类会对图像中自己感兴趣的区域给予更多关注, 所以发生在这些区域的失真更容易影响人类的主观质量分数, 而传统的图像质量评价 (IQA) 方法并未考虑到图像中不同区域受到关注度的差异, 导致预测分数与主观质量分数的拟合程度较低。针对上述问题, 本文利用显著图对图像中受关注区域的突出表达能力, 提出了一种基于显著性双流图像质量评价 (SDS-IQA) 方法。该方法采用显著图分支和原始图分支组成的双流分层结构, 从整体和重点两个方面实现图像的多尺度失真感知, 并通过双重注意力在所有维度上体现特征的重要性差异。SDS-IQA 在特征提取阶段通过在显著图分支使用空间注意力来体现尺度空间上的关注度差别, 并通过空间注意力权重来强化原始图分支的失真信息表达, 在特征融合阶段使用门控注意力强化通道间的交互, 使通道间的关注度差异在融合时得到体现, 最终实现对受关注区域中失真的重点表征。实验结果表明, 该方法在 3 个合成数据集 (LIVE, TID2013, CSIQ) 上的皮尔森线性相关系数分别达到 0.976、0.896 和 0.865, 在 2 个真实数据集 (LIVEC, KonIQ-10k) 上分别达到 0.869 和 0.877, 证明 SDS-IQA 的预测结果与人类主观评价有良好的拟合性。

关键词: 图像质量评价; 显著图; 注意力机制

中图分类号: TP391.4

文献标志码: A

文章编号: 0493-2137(2025)02-0189-10

No-Reference Image Quality Assessment Method Based on Saliency-Map and Dual-Stream Hierarchical Perception

Shi Zaifeng^{1,2}, Kang Tai¹, Wang Yunfeng¹, Xiao Yunze², Luo Tao³

(1. School of Microelectronics, Tianjin University, Tianjin 300072, China;

2. Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology, Tianjin 300072, China;

3. College of Intelligence and Computing, Tianjin University, Tianjin 300072, China)

Abstract: Humans pay more attention to the regions of interest in an image, so distortion in those areas is more likely to affect their subjective quality scores. However, traditional image quality assessment (IQA) methods do not take into account the difference in the attention received by different regions in the image, resulting in a lower degree of fitting between the predicted score and the subjective quality score. Aimed at the above problems, a saliency-map-based dual-stream image quality assessment (SDS-IQA) method to highlight the area of interest in the image was proposed. A dual-stream hierarchical structure composed of a saliency map branch and an original image branch was used to realize multi-scale distortion perception of images from both the whole and the focus, reflecting the importance difference of features in all dimensions through dual attention. In the feature extraction stage, SDS-IQA used spatial attention in the saliency-map branch to reflect the difference in attention in the scale space, strengthened the distortion information expression of the original image branch through spatial attention weight, and used gated attention to strengthen the interaction between channels in the feature fusion stage, so that the attention difference between channels could be

收稿日期: 2024-04-16; 修回日期: 2024-07-01.

作者简介: 史再峰 (1977—), 男, 博士, 副教授.

通信作者: 史再峰, shizaifeng@tju.edu.cn.

基金项目: 国家重点研发计划资助项目 (2022YFB2804401); 天津市科技计划资助项目 (22JCYBJC00140).

Supported by the National Key Research and Development Program of China (No. 2022YFB2804401), the Science and Technology Plan Program of Tianjin, China (No. 22JCYBJC00140).

reflected during fusion and the key characterization of distortion in the region of interest was eventually realized. Experimental results show that the Pearson linear correlation coefficient of this method reaches 0.976, 0.896 and 0.865 on three synthetic datasets(LIVE, TID2013, CSIQ), and 0.869 and 0.877 on two authentic datasets(LIVEC, KonIQ-10k), respectively, proving that the prediction results of SDS-IQA have good fit with human subjective assessment.

Keywords: image quality assessment(IQA); saliency map; attention mechanism

随着互联网和移动通信技术的发展,图像成为了人们感知世界和接收信息的重要媒介。从图像生成到被接收的过程中,对图像的处理和传输都有可能使原始图像出现失真,从而干扰人们从图像里获取信息^[1]。使用客观图像质量评价方法来对成像质量进行量化评测,能对图像质量优化和改善起到重要指导作用。图像质量评价(image quality assessment, IQA)方法根据是否有参考图像分为无参考和全参考,其中无参考难度更大,应用价值也更高。

早期的无参考方法通常基于手工设计特征提取,对自然场景失真的统计量从频率域^[2]、空间域^[3]和小波域^[4]进行特征提取和融合,对单一失真类型具有良好的评价效果。然而,当失真类型比较复杂时,基于手工模型提取到的特征无法充分表征图像的多尺度失真信息,导致预测结果与主观质量评分的一致性较差。

基于深度学习的方法使用卷积来提取图像质量的相关特征,其过程通过数据驱动来自动实现。Kang 等^[5]提出通过卷积神经网络(convolutional neural network, CNN)来建立端到端的图像质量评价模型。人类视觉系统(human visual system, HVS)处理信息的过程是分层的^[6],首先形成局部特征,然后再形成如语义信息之类的全局特征。CNN 的网络结构能实现图像特征的分层提取,其过程近似于 HVS 的感知机制,能对不同尺度的失真信息进行全面表达。受此启发, Gao 等^[7]使用深层 CNN 网络来提取分层特征,相较于只使用最后一层卷积的特征,其性能有显著提高。Wu 等^[8]通过实验系统分析了层次特征和退化的相关性,并提出了一种基于层次特征聚合的 IQA 方法。Sun 等^[9]将中间层特征分层整合到最终的特征表示中,并使用迭代混合数据库来学习更通用的特征表示方法。这些研究表明 CNN 的分层感知结构能同时提取不同抽象程度的失真特征信息。

在 HVS 中,注意力机制驱动着人们重点关注感兴趣的区域,而忽略其他不感兴趣的信息,这种感知机制被运用在 IQA 中。Ma 等^[10]利用语义感知注意力和质量感知注意力之间的差异来细化质量预测,通过共享的特征提取分支和独立的空间注意力分支联

合优化学习框架。You 等^[11]提出了一个轻量级注意力模块来模拟空间注意和对比度敏感性机制,通过空间和通道注意力机制来拟合 HVS 机制。Yu 等^[12]用分解的大核卷积来提取多尺度特征,使用轻量级注意力机制模块实现更低的计算复杂度,并通过一种新颖的特征增强模块来改进特征表示。以上方法通过注意力机制在通道维度和空间维度上依照重要性进行权重分配,使受关注区域的失真对质量分数预测的影响得到充分体现。

研究表明质量失真与视觉显著性(visual saliency, VS)之间有相关性^[13]。在观看图片时,人们重点关注区域的失真强度能通过视觉显著图来模拟。Zhang 等^[14]提出了一种基于 VS 的 IQA 方法,将 VS 用作失真图像的局部质量图的特征,并在计算质量分数时使用基于 VS 的加权函数来反映区域重要性。Zhu 等^[15]将显著性引入到 Transformer 中,引导模型在聚合全局信息时突出感兴趣区域,并结合本地嵌入来进行图像质量评估。Seo 等^[16]使用基于显著性的注意力机制,通过残差网络^[17]来提取感知上重要的特征。Yang 等^[18]在图像质量预测任务的基础上增加了视觉显著性预测的子任务,实现了感知一致的特征融合,但其显著性指导无法作用于原始图失真信息。Ma 等^[19]通过空间注意力模块强化了对原始图中受关注区域失真信息的表征能力,并在特征融合时使用全连接层进行通道间关系建模,但这种隐式方法得到的不同层间通道的关联性不够紧密。以上研究表明了显著图对受关注区域的失真映射能力可以提升评价结果与人类主观评价的一致性。

本文基于显著图和双流分层感知提出了一种无参考图像质量评价(no-reference image quality assessment, NR-IQA)方法。该方法使用原始图像分支和显著图分支组成双流结构,原始图像分支关注图像的整体结构,显著图分支用于引导模型对人眼更感兴趣的区域进行特征建模。利用 CNN 对两个分支进行分层提取来完成多尺度特征表达,并在显著图分支使用空间注意力^[20]对区域特征进行加权处理,利用权重指导原始图像分支的特征提取过程,实现对空间显著特征的优化。将提取的双流特征和空间注意力加

权特征进行融合,并将融合特征用作下一特征融合层的输入来增强特征信息的丰富性.在质量分数预测时通过门控通道注意力机制来优化跨通道交互,利用门控机制来抑制冗余通道的特征表达,以提高量化评测结果.

1 本文方法

1.1 方法整体结构

在 IQA 中,图像失真往往会导致不同程度的层次特征退化.由于失真类型的复杂性和失真分布位置的多变性,单一尺度和单一视觉复杂性的特征往往难以充分表达图像的失真信息.而图像的不同区域受到人们的关注度也不同,需要同时在空间维度和通道维度上体现不同区域的关注度差异性.因此,本文提出了一种基于显著性的双流图像质量评价方法 SDS-IQA,其结构如图 1 所示.

显著图由原始输入图像经过 LC^[21]显著性检测生成,并与原始图一起被输入至双流结构的不同分支.显著图分支和原始图分支均使用基于残差模块

的 CNN 网络进行特征提取.为了强化对不同像素在空间上的重要性差异的表达能力,在显著图分支里嵌入空间注意力模块,从而得到基于视觉显著性的空间注意力权重.由于显著图的空间注意力权重能充分体现受关注区域失真对预测质量分数的影响,将空间注意力权重与原始图像里提取的分层特征相乘,从而指导原始图分支提取出更具关注度差异性的空间特征.

为了实现图像信息的层次性表征,从第 2 层到第 5 层的特征都被用于图像质量预测.将原始图像特征、显著图像特征和具有关注度差异性的空间特征进行矩阵拼接,得到分层感知特征.由于 CNN 层数增加会导致局部信息丢失,在第 3 层及之后还将上一层的融合得到的特征进行降采样来作为特征融合的输入,在高层特征中降低了低层局部信息的损失.由于不同通道的特征图在 IQA 中的重要性也具有差异,本文使用基于门控机制^[22]的通道注意力模块,对分层感知特征进行融合,抑制对冗余通道特征的关注度.最后将加权处理后的特征输入至卷积层和池化层来实现降维,最终得到质量分数.

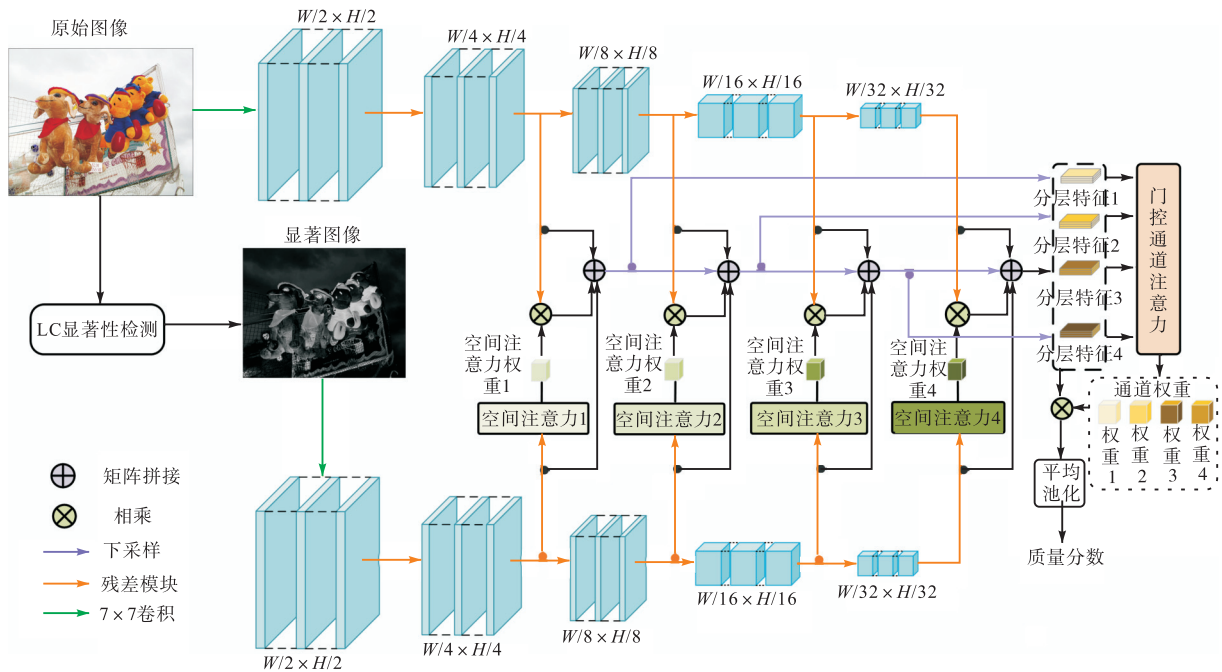


图 1 SDS-IQA 方法的结构
Fig.1 Structure of SDS-IQA method

1.2 空间注意力与分层感知

由于人的视觉处理机制对于空间上的不同区域存在关注度差异,在 IQA 里如果失真分布在人眼更关注的区域,则更容易影响人们对图像质量的整体评价.显著图能突出显示视觉系统首先关注的区域,并对该区域的失真进行强化表达,因此将空间注意力集

成到显著图像分支来提取特征,并使用显著图的注意力权重来指导原始图像对重点区域的特征进行提取,最后通过分层提取融合不同尺度的特征来实现多层信息联合.

1.2.1 空间注意力权重指导特征提取

双流分支的特征提取主要由残差模块完成,多个

残差模块堆叠形成特征提取骨干网络,其结构如图 2 所示. 残差模块包含两个分支,主分支由两个 3×3 卷积组成. 为了解决卷积层堆叠导致的网络退化问题,使用 1×1 卷积分支来跳跃连接,实现输入数据的恒等映射. 批量归一化对每个批数据在输入前进行标准化处理,在加快网络收敛的同时降低训练过程

中对于权重初始化的依赖. 残差模块的特征提取流程如下.

$$y = x + H(x, w_h) \quad (1)$$

式中: x 表示输入特征; $H(x, w_h)$ 为主分支两个卷积组成的非线性层; y 表示残差映射; w_h 为主分支中的神经网络参数.

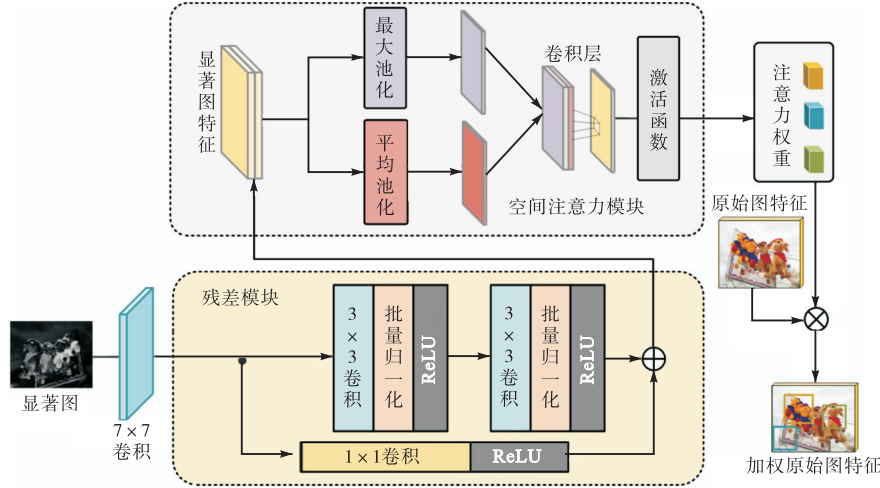


图 2 空间注意力权重指导特征提取流程

Fig.2 Flow chart of spatial attention weight guiding feature extraction

空间注意力模块依据受关注度对显著图分支中所提取的特征在二维空间上进行建模表征. 首先对输入特征进行通道压缩,在通道维度上分别通过平均值池化和最大池化进行降维,得到两个二维特征,并通过拼接得到一个双通道特征图. 之后两通道特征图通过单个卷积核的隐藏层实现降维,保证输入特征与输出特征的形状在空间层面上相同,并经过激活函数得到加权显著图特征. 同时注意力权重还被用于指导原始图分支的特征提取过程,将权重与原始图特征相乘,最终生成加权原始图特征. 该过程如下.

$$\begin{cases} W_s = \sigma(\text{conv}(\text{AP}(F_s), \text{MP}(F_s))) \\ F_p^w = W_s \times F_p \end{cases} \quad (2)$$

式中: σ 表示 Sigmoid 激活函数; AP 和 MP 分别表示平均池化和最大池化; conv 为卷积层; W_s 表示空间注意力权重; F_s 为显著图特征; F_p 和 F_p^w 分别为原始图特征和加权原始图特征.

1.2.2 双流分层感知

HVS 对于视觉信号的处理流程具有层次性,对图片的理解会从低级简单的特征逐渐扩展到高级抽象的特征. CNN 通过分层特征提取和感受野尺寸的变化来模拟大脑皮层的分级处理,特征的抽象程度也会随着网络深度和感受野大小的增加逐渐上升. IQA 任务里,对图像质量的感知是多尺度特征共同影响的

过程,如果只通过深层的特征信息来进行质量评估,会导致低层次的图像底层特征被忽略,无法充分表征图像的退化. 因此本文通过特征的多阶统计量来全面评价图像的整体质量.

从双流分支的特征提取网络第 2 层开始,为了增加图像失真信息的丰富度,把第 l 层原始图特征、加权原始图特征和显著图特征在通道维度上进行融合,将不同尺度和不同关注度的特征图都用于质量信息的表征. 第 l 层的融合特征经过下采样后,被用做第 $(l+1)$ 层的输入,使低层次特征信息得到保留. 每层的融合特征可以表示为

$$\begin{cases} C_i = F_{p,i} \oplus F_{s,i} \oplus F_{p,i}^w \\ F_{p,i}^w = F_{p,i} \otimes W_{s,i} \end{cases} \quad i = 2,3,4,5 \quad (3)$$

$$C_k = C_i \oplus D(C_{k-1}) \quad k = 3,4,5 \quad (4)$$

式中: \oplus 为矩阵拼接; $F_{p,i}$ 和 $F_{s,i}$ 分别表示第 i 层原始图像特征和显著图特征; $W_{s,i}$ 为第 i 层空间注意力; $F_{p,i}^w$ 表示第 i 层加权原始图特征; \otimes 表示相乘; D 为下采样; C_k 表示第 k 层融合后的特征.

1.3 门控通道注意力模块

通过 CNN 提取的特征由多通道组成,每个通道都可以对 IQA 的最终结果做出贡献. 实际上,不同通道的特征信息对质量分数预测的贡献是不相同的,受到 HVS 的注意力机制启发,每个通道应该按照重要

性被赋予不同的权重, 以实现冗余通道信息的抑制. 本文通过门控通道注意力模块对通道间交互进

行强化, 实现通道权重的分配, 其结构如图3所示.

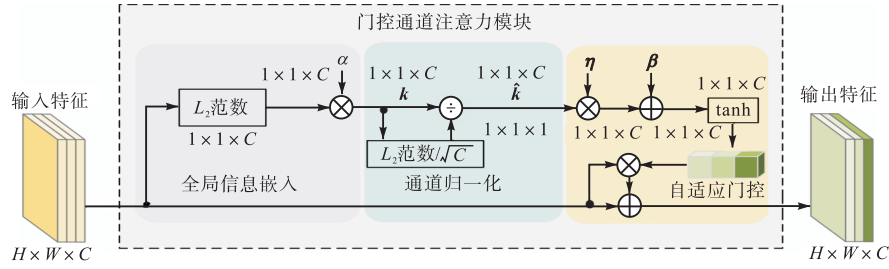


图3 门控通道注意力模块的结构

Fig.3 Structure of gated channel attention module

门控通道注意力模块分为全局信息嵌入、通道归一化、自适应门控3个部分. 全局信息嵌入模块对输入特征的各个通道进行压缩, 实现全局信息聚合. 为了避免通道特征产生固化, 选择使用 L_2 范数来代替全局平均池化, 其过程如下所示.

$$\begin{cases} \mathbf{k}_c = \alpha \|\mathbf{x}_c\|_2 = \alpha \left\{ \left[\sum_{i=1}^H \sum_{j=1}^W (\mathbf{x}_c^{i,j})^2 \right] + \phi \right\}^{\frac{1}{2}} \\ \alpha = [\alpha_1, \alpha_2, \dots, \alpha_c] \end{cases} \quad (5)$$

式中: \mathbf{k}_c 表示压缩通道特征; α 表示每个通道的嵌入权重, 这是一个可训练参数; \mathbf{x}_c 为输入特征; ϕ 为极小常数, 避免在零点处求导的问题. 使用归一化模块代替全连接层来进行通道间关系建模, 实现对分层特征跨通道交互的优化并完成通道间竞争和协同关系的构建, 其过程如下.

$$\begin{cases} \hat{\mathbf{k}}_c = \frac{\sqrt{C} \mathbf{k}_c}{\|\mathbf{k}\|_2} = \frac{\sqrt{C} \mathbf{k}_c}{\left(\sum_{c=1}^C \mathbf{k}_c^2 + \phi \right)^{\frac{1}{2}}} \\ \mathbf{k} = [k_1, k_2, \dots, k_c] \end{cases} \quad (6)$$

式中: $\hat{\mathbf{k}}_c$ 为归一化后的输出特征; \sqrt{C} 为尺度因子, 用于对 $\hat{\mathbf{k}}_c$ 的尺度进行归一化, 避免 C 较大时特征尺度过小. 通过引入自适应门控机制来促使训练过程中冗余通道的特征表达被抑制. 当通道的门控权重为正时, 通道之间为竞争关系; 当通道的门控权重为负时, 通道之间为协同关系. 其过程如下所示.

$$\hat{\mathbf{x}}_c = \mathbf{x}_c \left[1 + \tanh(\eta \hat{\mathbf{k}}_c + \beta) \right] \quad (7)$$

式中: η 和 β 分别为门控权重和偏置, 用于控制门限的开关; $\hat{\mathbf{x}}_c$ 为输出特征.

2 实验设置与结果分析

2.1 实验环境与数据集

实验的训练和测试环境如下: 操作系统为

Windows10, CPU 为 i9-9900KF@3.60 GHz, GPU 为 NVIDIA GeForce RTX 3090, Pytorch 版本为 1.6.0.

测试和训练过程中使用的合成数据集包括 TID2013^[23]、LIVE^[24]、CSIQ^[25], 真实数据集有 LIVE 挑战赛 (LIVEC)^[26] 和 KonIQ-10k^[27]. TID2013 包含 25 张参考图像和 3 000 张失真图像, 有 24 种失真类型, 每种失真对应 5 个不同级别. LIVE 数据集包含从 29 张原始图像生成的 779 张失真图像, 每张失真对应 5 种失真类型. CSIQ 包含由 30 张原始图像生成的总共 866 张扭曲图像, 共有 6 种失真类型. LIVEC 数据库由 1 162 张不同相机设备拍摄的真实失真图像组成. KonIQ-10k 包含 10 073 张失真图像. 每个数据集按照 8 : 2 的比例划分为训练集和测试集. 数据集的标签为平均主观得分 (MOS) 或平均主观得分差 (DMOS). MOS 越高或 DMOS 越低代表图像的感知质量更好. 本文选择斯皮尔曼秩相关系数 (Spearman rank-order correlation coefficient, SROCC) 和皮尔森线性相关系数 (Pearson linear correlation coefficient, PLCC) 作为评价 NR-IQA 方法的性能指标. SROCC 和 PLCC 的值的范围为 [0, 1], 值越接近 1, 模型性能越好, 与人类的主观评价越拟合.

2.2 合成数据集上的性能对比

为了验证 SDS-IQA 在面对合成失真图像时的表现, 在 3 个合成数据集 (LIVE, TID2013, CSIQ) 上将 SDS-IQA 与一些经典的全参考 IQA 方法和无参考 IQA 方法进行了对比, 结果如表 1 所示.

从表 1 中可知, SDS-IQA 在 3 个数据集上总共 6 项对比指标里有 3 项结果排名前二, 在 TID2013 数据集上的 PLCC 值达到了最高的 0.896. 作为一种无参考 IQA 方法, 与早期的全参考 IQA 方法 (SSIM^[28], PSNR) 相比也具有一定的性能优势, 与一些效果较好的全参考 IQA 方法 (DeepSim^[29], GMSD^[30]) 相比也在 3 个指标上更好, 在没有参考图的情况下实现了更精确的质量评估. 相较于仅使用

CNN 方法最后输出的单层特征的方法 (DIQA^[31], BIECON^[32], TSIG^[33]), SDS-IQA 方法的总体表现优于 DIQA 和 BIECON 方法, 与 TSIG 方法相比在 TID2013 数据集的 PLCC 上提升了 0.041, 这证明了提取不同尺度的特征可以更加全面地评价图像的整体质量. 相较于 AIGQA 方法^[34], SDS-IQA 方法在 3 个指标上都更优秀, 这是由于 SDS-IQA 方法在特征提取时通过显著图分支来丰富特征中的关注度差异信息, 弥补了原始图像分支特征丰富度不足的缺点, 但在 CSIQ 数据集上的 2 个指标和 TID2013 数据集上的 SROCC 值 AIGQA 方法都更高, 这可能是由于 AIGQA 方法通过生成伪参考图来实现对图中内容的预测, 以类似于全参考 IQA 方法进行质量评价. 相较于 DB-CNN 方法^[35], SDS-IQA 方法在 LIVE 和 TID2013 数据集上性能更优秀, 但在 CSIQ 数据集的

两个指标上 DB-CNN 方法表现更好. 这是由于 DB-CNN 方法使用了经过大规模图像分类任务预训练的 VGG-16 网络, 还合并了两个大规模数据集进行预训练, 而 SDS-IQA 方法只使用单一数据集进行训练和测试, 用于训练的数据量存在一定差距. 相较于端到端多级特征提取的 CaHDC 方法^[36], SDS-IQA 方法在 LIVE 数据集上的 PLCC 值提升了 0.012, 原因在于 SDS-IQA 方法将注意力机制嵌入网络中来突出受关注区域, 更符合 HVS 的感知特点. 与同样利用视觉显著性进行特征提取的方法 (SETNet^[18], SGDNet^[19]) 相比, SDS-IQA 方法在 LIVE 和 TID2013 数据集上表现也更好, 这是因为在通道注意力的建模过程中 SDS-IQA 方法使用通道归一化加强了通道间信息的关联性, 使特征信息在通道维度的重要性差异得到体现.

表 1 合成数据集上不同 IQA 方法的性能比较

Tab.1 Performance comparison of different IQA methods on synthetic datasets

方法		LIVE 数据集		TID2013 数据集		CSIQ 数据集	
		PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
全参考 IQA 方法	DeepSim ^[29]	0.968	0.974	0.872	0.846	0.919	0.919
	SSIM ^[28]	0.945	0.948	0.691	0.637	0.861	0.876
	GMSD ^[30]	0.960	0.960	0.859	0.801	0.954	0.957
	PSNR	0.872	0.886	0.706	0.636	0.800	0.806
无参考 IQA 方法	DIQA ^[31]	0.971	0.969	0.850	0.825	0.915	0.884
	DB-CNN ^[35]	0.971	0.968	0.865	0.816	0.959	0.946
	BIECON ^[32]	0.962	0.961	0.761	0.721	0.823	0.815
	AIGQA ^[34]	0.957	0.960	0.893	0.871	0.952	0.927
	CaHDC ^[36]	0.964	0.965	0.878	0.862	0.914	0.903
	TSIG ^[33]	0.978	0.969	0.855	0.878	0.929	0.923
	SETNet ^[18]	0.972	0.964	0.891	0.864	0.934	0.918
	SGDNet ^[19]	0.965	0.969	0.861	0.843	0.872	0.851
SDS-IQA (本文方法)	0.976	0.972	0.896	0.868	0.865	0.858	

本文使用散点图来直观展示测试图像的图像标签值和预测分数之间的关系, 实验结果如图 4 所示. 图中使用蓝色点来表示测试图像, 蓝色斜线表示测试图的线性拟合曲线, 纵轴为失真图像的标签值, 横轴表示 SDS-IQA 对测试图像进行质量评价后得到的预测分数. 从图中可看出在合成数据集中预测值与标签值的线性相关度较好, 拟合效果良好.

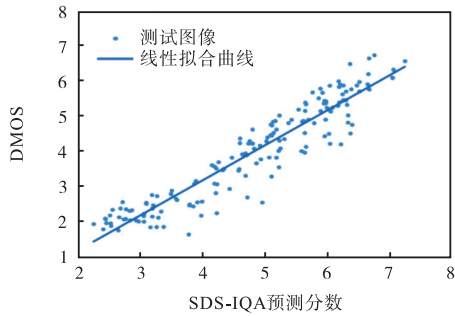
对于不同类型的失真, IQA 方法都应表现出鲁棒性. 为了验证 SDS-IQA 方法对特定失真类型的预测效果, 本文在 LIVE 数据集上针对 5 种不同类型的失真进行 SROCC 比较实验, 结果如表 2 所示. 从表中可以发现, 与一些全参考方法和经典的无参考方法相比, SDS-IQA 方法在 LIVE 数据集上的 5 种失真类型中面对 JPEG2000 压缩 (JP2K)、白噪声 (WN) 和高斯模糊 (GB) 表现最好, 分别达到 0.971、0.983 和

0.958, 在 JPEG 压缩 (JPEG) 和快速衰落 (FF) 里也排名前二, 这表明 SDS-IQA 的预测性能不受失真的分布位置和类型所影响, 充分体现了该方法优秀的泛化性.

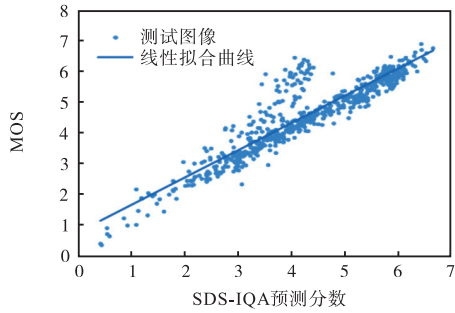
2.3 真实数据集上的性能对比

本文在 LIVEC 和 KonIQ-10k 两个真实数据集上进行对比实验来对 SDS-IQA 方法进行测试, 结果如表 3 所示. 从实验结果中可以发现, SDS-IQA 方法在 LIVEC 数据集上 SROCC 达到 0.853, 在 KonIQ-10k 数据集上 PLCC 达到 0.877, 在其余两个指标上也排名前二. 与其他基于 CNN 的 IQA 方法 (DIQA, BIECON, AIGQA) 相比, 除了在 KonIQ-10k 数据集上的 SROCC 比经过大规模预训练的 DB-CNN 方法低之外, SDS-IQA 方法在性能上都更加优秀, 这表明 SDS-IQA 方法通过空间注意力权重引导, 增强了原

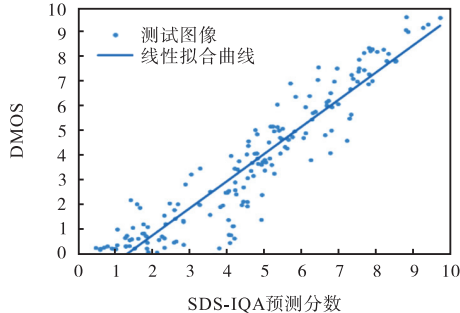
始图像的 CNN 方法分支对空间显著特征的提取能力,使融合后的特征能包含更多失真信息. 与使用空间注意力和通道注意力的方法(DQM-IQA^[37], SETNet^[18], SGDNet^[19])相比, SDS-IQA 方法在性能指标上都更高,笔者认为这是由于基于门控机制的通道注意力模块在通道间关系的建模上更加优秀,增强了层间通道间的信息关联性. 实验结果证明了 SDS-



(a) 在 LIVE 数据集上的散点图



(b) 在 TID2013 数据集上的散点图



(c) 在 CSIQ 数据集上的散点图

图 4 SDS-IQA 方法在合成失真测试集的散点图

Fig.4 SDS-IQA scatter plots on synthetic distortion test sets

表 2 评估不同失真类型图片的结果

Tab.2 Assessment results for different distortion types of images

方法	失真类型				
	JP2K	JPEG	WN	GB	FF
SSIM ^[29]	0.939	0.946	0.964	0.904	0.941
PSNR	0.870	0.885	0.942	0.763	0.874
BIECON ^[32]	0.952	0.974	0.970	0.945	0.923
DB-CNN ^[35]	0.955	0.972	0.980	0.940	0.930
DQM-IQA ^[37]	0.952	0.950	0.951	0.956	0.956
TSIG ^[33]	0.966	0.950	0.979	0.963	0.911
SDS-IQA (本文方法)	0.971	0.968	0.983	0.958	0.954

表 3 真实数据集测试的 SROCC 和 PLCC 结果对比

Tab.3 Comparison of SROCC and PLCC results on authentic datasets

方法	LIVEC 数据集		KonIQ-10k 数据集	
	PLCC	SROCC	PLCC	SROCC
DIQA ^[31]	0.680	0.671	0.761	0.739
DB-CNN ^[35]	0.869	0.851	0.869	0.856
BIECON ^[32]	0.596	0.582	0.598	0.591
AIGQA ^[34]	0.761	0.751	—	—
SETNet ^[18]	0.776	0.745	—	—
DQM-IQA ^[37]	0.676	0.652	0.673	0.655
SGDNet ^[19]	0.872	0.851	0.868	0.811
SDS-IQA (本文方法)	0.869	0.853	0.877	0.854

IQA 方法在面对复杂且未知的真实失真时也具有良好的预测性能,具有很高的实际应用价值.

2.4 跨数据集测试

优秀的 NR-IQA 方法不仅在单个数据集中表现良好,而且还可以很好地扩展到其他 IQA 数据集. 为了评估 SDS-IQA 方法面对未知数据集的泛化能力,本文选择在 TID2013 数据集上对 SDS-IQA 方法进行训练,并在其他 4 个数据集上进行测试. 在数据预处理时将其他数据集上的分数标签线性缩放到与 TID2013 数据集中的 MOS 值相同的范围[0, 9], 以便于直接进行结果对比. 跨数据集测试结果如表 4 所示. 从表中可知, SDS-IQA 方法在其中 3 个数据集里的 SROCC 值最高,在 CSIQ 数据集上的表现也具有一定竞争力,证明了 SDS-IQA 方法的性能不依赖于数据集的种类,面对其他数据集上的真实失真和合成失真都具有很好的适应能力.

表 4 跨数据集测试的 SROCC 结果

Tab.4 Cross dataset test results of SROCC

方法	SROCC			
	LIVE 数据集	CSIQ 数据集	LIVEC 数据集	KonIQ-10k 数据集
BLIINDS-II ^[38]	0.836	0.568	0.142	0.304
RankIQA ^[39]	0.599	0.667	0.289	0.460
DIQA ^[31]	0.834	0.679	0.048	0.182
SETNet ^[18]	0.854	0.710	0.462	0.407
TSIG ^[33]	0.853	0.735	0.455	0.324
SDS-IQA (本文方法)	0.858	0.686	0.542	0.586

2.5 消融实验

本文设计了消融实验来评估 SDS-IQA 中不同模块的有效性. 在 KonIQ-10k 数据集上进行训练,每次实验都使用相同的训练参数和设置. 将双重注意力模块去除,仅保留双流分层网络作为基线,以此为基础逐步添加改进模块,实验结果如图 5 所示. 在只保留基线网络的情况下预测效果不理想, SROCC 和 PLCC 均小于 0.700. 在结合空间注意力对原始图像

分支进行特征提取指导后,效果得到明显提升, SROCC 和 PLCC 分别提升 0.141 和 0.135, 这表明了模拟 HVS 关注重点区域能显著增加预测和主观结果之间的拟合性. 在使用门控注意力为不同通道给予差异性关注度后, SROCC 和 PLCC 相较于基线网络分别提升 0.168 和 0.152, 证明了通道注意力的有效性. 在将所有模块结合后效果达到最佳, SROCC 和 PLCC 达到 0.854 和 0.877. 实验结果表明每个模块都对性能提升起正向作用, 不同模块进行组合可以产生更好的结果, 文中所提出的模块都使得预测结果更加符合人类主观观感.

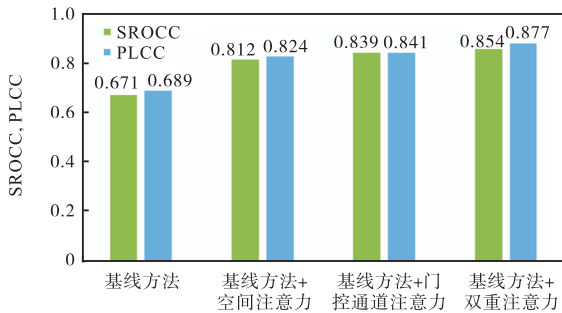
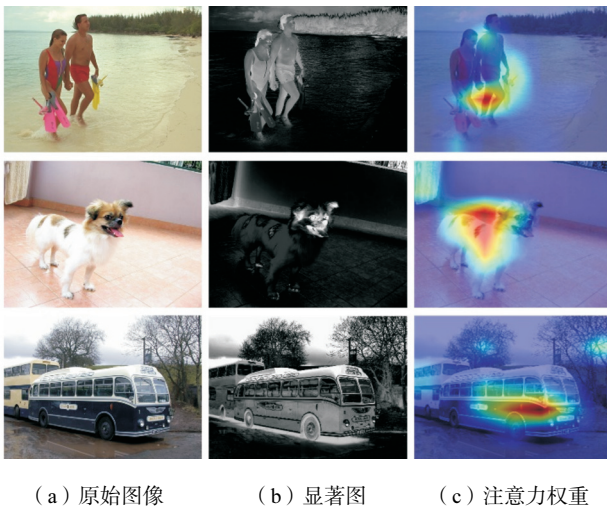


图 5 消融实验结果

Fig.5 Ablation experiment results

2.6 可视化分析

为了更直观地展现显著图和空间注意力所获得的信息, 本文选取了 3 张具有显著目标的失真图像, 将显著区域和空间注意力权重进行了可视化, 结果如图 6 所示.



(a) 原始图像 (b) 显著图 (c) 注意力权重

图 6 可视化结果

Fig.6 Visualization results

图 6(a) ~ (c) 分别为原始图像、显著图和注意力权重. 从图 6(b) 中可以看出, 显著图对图中的人、汽车和狗都判定为显著区域. 图 6(c) 通过热力图来表

示权重分配情况, 可以看出发生在显著区域中的失真具有更高的数据密度, 这表明 SDS-IQA 方法在训练时学习到了显著图中受关注区域的失真信息, 并根据不同区域的关注度差异分配了相对应的权重. 这些可视化结果直观地证明了基于显著图的空间注意力机制在处理失真图像时与 HVS 具有相似性, 可以更充分地反映受关注区域的失真信息.

3 结 语

本文针对图像中不同区域的关注度差异问题, 提出了一种基于显著图的双流分层感知的无参考图像质量评价方法. 该方法通过模拟人类视觉感知过程的层次性流程, 通过分层感知在不同尺度上对图像失真信息进行表征. 使用显著图和原始图的双流分支结构进行特征提取, 同时强化图像整体和重点区域的特征表达. 利用空间注意力模块获取显著区域的权重分布, 进而指导原始图像分支对不同区域实行差异性关注. 采用门控通道注意力模块进行通道间关系建模, 体现不同通道的重要性差别. 消融实验证明了两个注意力模块均能提升 IQA 方法的性能. 在合成数据集和真实数据集的实验结果均证明了 SDS-IQA 方法的预测结果更加符合人类主观评价.

参考文献:

[1] Cao Y Q, Min X K, Sun W, et al. Attention-guided neural networks for full-reference and no-reference audio-visual quality assessment[J]. IEEE Transactions on Image Processing, 2023, 32: 1882-1896.

[2] 张 涛, 潘希豪, 王 昊, 等. 基于时空域方向自适应结合的去隔行方法[J]. 天津大学学报(自然科学与工程技术版), 2019, 52(11): 1211-1218.
Zhang Tao, Pan Xihao, Wang Hao, et al. De-interlacing method based on adaptive combination of time-domain and space-domain directions[J]. Journal of Tianjin University(Science and Technology), 2019, 52(11): 1211-1218(in Chinese).

[3] 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.

[4] 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.
- [5] Kang L, Ye P, Li Y, et al. Convolutional neural networks for no-reference image quality assessment[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1733-1740.
- [6] Hochstein S, Ahissar M. View from the top: Hierarchies and reverse hierarchies in the visual system[J]. *Neuron*, 2002, 36(5): 791-804.
- [7] Gao F, Yu J, Zhu S G, et al. Blind image quality prediction by exploiting multi-level deep representations[J]. *Pattern Recognition*, 2018, 81: 432-442.
- [8] Wu J J, Yang W, Li L D, et al. Blind image quality prediction with hierarchical feature aggregation[J]. *Information Sciences*, 2021, 552: 167-182.
- [9] Sun W, Min X K, Tu D Y, et al. Blind quality assessment for in-the-wild images via hierarchical feature fusion and iterative mixed database training[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2023, 17(6): 1178-1192.
- [10] Ma X Y, Wang Y Q, Liu C, et al. ADGNet: Attention discrepancy guided deep neural network for blind image quality assessment[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York, USA, 2022: 1309-1318.
- [11] You J Y, Yan J. Explore spatial and channel attention in image quality assessment[C]//IEEE International Conference on Image Processing. Bordeaux, France, 2022: 26-30.
- [12] Yu L, Li J Y, Pakdaman F, et al. MAMIQA: No-reference image quality assessment based on multiscale attention mechanism with natural scene statistics[J]. *IEEE Signal Processing Letters*, 2023, 30: 588-592.
- [13] Engelke U, Kaprykowsky H, Zepernick H J, et al. Visual attention in quality assessment[J]. *IEEE Signal Processing Magazine*, 2011, 28(6): 50-59.
- [14] Zhang L, Shen Y, Li H Y. VSI: A visual saliency-induced index for perceptual image quality assessment[J]. *IEEE Transactions on Image Processing*, 2014, 23(10): 4270-4281.
- [15] Zhu M M, Hou G Q, Chen X J, et al. Saliency-guided transformer network combined with local embedding for no-reference image quality assessment[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021: 1953-1962.
- [16] 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*, 2020, 31(7): 2602-2616.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778.
- [18] Ma H H, Cui Z G, Gan Z L, et al. Saliency-enhanced two-stream convolutional network for no-reference image quality assessment[J]. *Journal of Electronic Imaging*, 2022, 31(2): 023010.
- [19] Yang S, Jiang Q P, Lin W S, 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, USA, 2019: 1383-1391.
- [20] Woo S H, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Munich, Germany, 2018: 3-19.
- [21] Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues[C]//Proceedings of the 14th ACM International Conference on Multimedia. New York, USA, 2006: 815-824.
- [22] Yang Z X, Zhu L C, Wu Y, et al. Gated channel transformation for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 11794-11803.
- [23] Ponomarenko N, Jin L, Ieremeiev O, et al. Image database TID2013: Peculiarities, results and perspectives[J]. *Signal Processing: Image Communication*, 2015, 30: 57-77.
- [24] 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.
- [25] 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.
- [26] Ghadiyaram D, Bovik A C. Massive online crowd-sourced study of subjective and objective picture quality [J]. IEEE Transactions on Image Processing, 2015, 25(1): 372-387.
- [27] Hosu V, Lin H 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.
- [28] Wang Z. Image quality assessment: From error measurement to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13: 600-613.
- [29] Gao F, Wang Y, Li P P, et al. DeepSim: Deep similarity for image quality assessment[J]. Neurocomputing, 2017, 257: 104-114.
- [30] Xue W F, Zhang L, Mou X Q, et al. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index[J]. IEEE Transactions on Image Processing, 2013, 23(2): 684-695.
- [31] 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.
- [32] Kim J, Lee S. Fully deep blind image quality predictor[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(1): 206-220.
- [33] Yan Q S, Gong D, Zhang Y N. Two-stream convolutional networks for blind image quality assessment [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2200-2211.
- [34] Ma J P, Wu J J, Li L D, et al. Blind image quality assessment with active inference[J]. IEEE Transactions on Image Processing, 2021, 30: 3650-3663.
- [35] 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, 2018, 30(1): 36-47.
- [36] Wu J J, Ma J P, Liang F H, et al. End-to-end blind image quality prediction with cascaded deep neural network[J]. IEEE Transactions on Image Processing, 2020, 29: 7414-7426.
- [37] Shi Z F, Wang Z, Kong F N, et al. Dual-quality map based no reference image quality assessment using deformable convolution[J]. Digital Signal Processing, 2022, 123: 103398.
- [38] 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.
- [39] Liu X L, van de Weijer J, Bagdanov A D. RankIQA: Learning from rankings for no-reference image quality assessment[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 1040-1049.

(责任编辑:孙立华)