



# 尺度因子感知对比学习的无参考型超分辨率图像质量评价

樊丹丹<sup>1</sup>, 张凯兵<sup>1,2,3</sup>, 权星<sup>1</sup>, 孟雅蕾<sup>2,3</sup>

(1. 西安工程大学 电子信息学院, 陕西 西安 710048; 2. 西安工程大学 计算机科学学院, 陕西 西安 710048;  
3. 陕西省服装设计智能化重点实验室, 陕西 西安 710048)

**摘要** 超分辨率图像的质量不仅受重建算法的影响, 而且在不同的尺度因子下重建出的图像在质量退化等级方面存在一定差异。然而现有的无参考型超分辨率图像质量评价方法主要关注超分辨率图像的视觉特征, 忽略了可用的尺度因子信息。提出了一种尺度因子感知对比学习 (upscaling-factor aware contrastive learning, UFACL) 方法, 该网络结构分为尺度因子识别分支和质量分数分支。其中尺度因子识别分支从数据集本身出发, 将不同尺度因子的超分辨率图像作为彼此的正负样本, 在完成分类任务的同时引入对比学习, 提高有效特征的表达能力。在质量分数分支设计了一个频域注意模块 (frequency domain attention module, FDAM), 考虑了全局信息和通道信息, 同时, 该分支使用倒残差块 (inverted residuals blocks, IRB) 降低模型的计算量, 使得在训练过程中既保证了质量分数预测精度又提升了模型训练效率。实验结果表明, 提出的 UFACL 能够获得与主观感知质量更好的一致性。

**关键词** 超分辨率图像质量评价; 尺度因子; 对比学习; 频域注意模块; 倒残差块

**中图分类号:** TP391.41 **DOI:** 10.16152/j.cnki.xdxbr.2025-02-008

## No-reference super-resolution image quality assessment based on upscaling-factor aware contrastive learning

FAN Dandan<sup>1</sup>, ZHANG Kaibing<sup>1,2,3</sup>, QUAN Xing<sup>1</sup>, MENG Yalei<sup>2,3</sup>

(1. School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China;

2. School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China;

3. Shaanxi Key Laboratory of Clothing Intelligence, Xi'an 710048, China)

**Abstract** The quality of super-resolution images is not only affected by the reconstruction algorithm, but also there are some differences in the quality degradation levels of the reconstructed images under different upscaling-factors. However, the existing no-reference super-resolution image quality assessment (NR-SRIQA) methods mainly focus on the visual features of super-resolution images, ignoring the available upscaling-factor information. An upscaling-factor aware contrastive learning (UFACL) method is proposed. The network structure is divided into a upscaling-factor recognition branch and a quality score branch. The upscaling-factor recognition branch starts from the dataset, and takes the super-resolution images of different upscaling-factors as posi-

收稿日期: 2024-10-18

基金项目: 国家自然科学基金(61971339); 陕西省自然科学基金基础研究计划重点项目(2018JZ6002)。

第一作者: 樊丹丹, 女, 从事图像质量评价研究, 220411045@stu.xpu.edu.cn。

通信作者: 张凯兵, 男, 教授, 从事机器学习、图像超分辨重建等研究, zhangkaibing@xpu.edu.cn。

tive and negative samples of each other. Contrastive learning is introduced to complete the classification task, so as to improve the expression ability of effective features. In the quality score branch, a frequency domain attention module (FDAM) is designed, which considers both global information and channel information. At the same time, this branch uses inverted residuals blocks (IRB) to reduce the calculation amount of the model, which ensures the accuracy of quality score prediction and improves the training efficiency of the model in the training process. Experimental results show that the proposed UFACL can achieve better consistency with subjective perceived quality.

**Keywords** super-resolution image quality assessment; upscaling-factor; contrastive learning; frequency domain attention network; inverted residuals blocks

图像分辨率是评价图像最重要的指标,一般而言,图像分辨率越高,包含的信息越丰富,说明图像的视觉质量越高<sup>[1]</sup>。在资源受限的环境下,为了获取高质量的图像,超分辨重建技术<sup>[2]</sup>被提出,该技术主要利用低分辨率的图像重建出包含更多细节的高分辨率图像。然而,如何有效且公平地比较这些超分辨重建算法的优劣,评价不同算法生成的超分辨图像的质量,进一步优化不同超分辨算法的性能是一直存在的问题。因此,超分辨图像质量评价(super-resolution image quality assessment, SRIQA)是超分辨技术发展面临的关键挑战之一,对提升超分辨算法的性能具有重要意义<sup>[3-4]</sup>。

目前,现有的超分辨图像质量评价(SRIQA)方法可分为 2 大类,即主观质量评价和客观质量评价。主观质量评价通过人眼对图像的主观测试来实现<sup>[5]</sup>,并采用平均意见得分(mean opinion score, MOS)或差异平均意见得分(difference mean opinion score, DMOS)作为衡量图像质量的指标。尽管这种方法提供了一种直接且可靠的 SRIQA 方法,但是这是一件费时费力的工作,此外,主观质量评价方法还不能嵌入到 SR 系统中进行实际应用。相比之下,客观质量评价方法通过设计计算模型<sup>[6]</sup>,可以利用提取的感知特征自动计算给定超分辨图像的质量。根据评价时是否使用参考图像信息,客观质量评价又分为全参考型(full-reference, FR)、半参考型(reduced-reference, RR)和无参考型(no-reference, NR)3 类<sup>[7]</sup>。其中,FR-IQA 和 RR-IQA 模型主要通过量化参考图像和失真图像之间的差异进行质量评价,该过程均需要原始参考图像的参与,但是在实际应用中,很难获取到原始高分辨图像,使得这些方法的实用性受到限制。

相比之下, NR-IQA 方法不需要原始参考图像的任何信息,在实际场景下应用更加广泛。

Zhang 等人提出将结构统计特征、多尺度方向和频率统计特征以及颜色统计特征结合,计算图像的马氏距离,获取图像预测分数<sup>[8]</sup>; Ye 等人则通过使用多种 FR-IQA 模型计算图像的质量得分,将该分数作为图像的主观质量得分,以此构建 NR-IQA 模型<sup>[9]</sup>。然而这些方法针对单一失真类型表现良好,对于现实场景下的图像所表现出来的复杂失真类型并没有明显优势,因此 Gu 等人提出了一种通用的 IQA 模型,对于单一失真和多重失真都有不错的表现<sup>[10]</sup>。上述方法都是在传统机器学习上构建的模型,在性能方面与基于深度学习的模型相比不是很高。基于深度学习的模型主要通过建立图像特征与质量分数间的映射关系<sup>[11]</sup>,获取图像质量分数。Liu 等人提出一种基于排序学习的方法,通过图像对之间的排序信息获取图像的视觉质量<sup>[12]</sup>;艾娜等人采用自举算法进行超分辨率重构,训练具有双重稀疏性的字典对,进一步对图像进行深层特征提取<sup>[13]</sup>; Bourbia 等人提出了一种新的基于多任务深度学习的盲立体图像质量评估方法,用于立体图像质量评价<sup>[14]</sup>; Li 等人将显著性子任务与 IQA 任务联合优化,以提高显著性引导的 IQA 性能<sup>[15]</sup>; 李云红等人提出利用公共特征子空间提取图像的特征信息,实现精准特征识别<sup>[16]</sup>。然而,这些方法虽然在一定程度上提升了 SRIQA 的性能,但是这些方法主要关注于超分辨率图像的视觉特征,忽略了可用的尺度因子信息。

具体来说,超分辨图像的质量不仅受重建算法的影响,而且在不同的尺度因子下重建出的图像在质量退化等级方面存在一定的差异。一般情况下,尺度因子越大,说明重建出来的图像质量越差,因此尺度因子大小与超分辨图像质量之间存在一定程度的相关性。随着尺度因子的不断增大,超分辨图像质量不断降低,这是一个明显的质

量退化过程。相反地,尺度因子越小,超分辨率图像是一个明显的质量增强过程,这 2 个过程都符合人眼的视觉认知,证明尺度因子与图像质量之间存在一定的线性相关性,图像的尺度信息与其主观质量得分高度相关。受此启发,本文得出结论,尺度因子对超分辨率图像主观质量分数的影响具有统计学意义,可以提供一些关于图像视觉质量的特征。换句话说,尺度信息对超分辨率图像质量评价任务是有用的,可以用于指导 NR-SRIQA 的学习。从这个角度出发,文献[17]提出了一个尺度引导的超网络框架,该框架以尺度自适应的方式评价图像质量,经过内容感知后,根据超分辨率图像的尺度因子生成用于质量预测的评价规则,在一定程度上增强了模型的泛化能力。受上述研究启

发,本文结合轻量级的倒残差块(inverted residual blocks, IRB)<sup>[18]</sup>,提出了一种基于尺度因子感知对比学习(upscaling-factor aware contrastive learning, UFACL)的超分辨率图像质量评价方法,在训练过程中既保证了质量分数预测精度又提升了模型训练速度。

## 1 UFACL 方法

本文提出的 UFACL 方法通过双分支结构探索尺度因子与超分辨率图像质量之间的关系,从而在超分辨率图像质量评价过程中获得更丰富的特征,具体来说,该方法主要由尺度因子识别分支和质量分数分支组成,整体框架如图 1 所示。

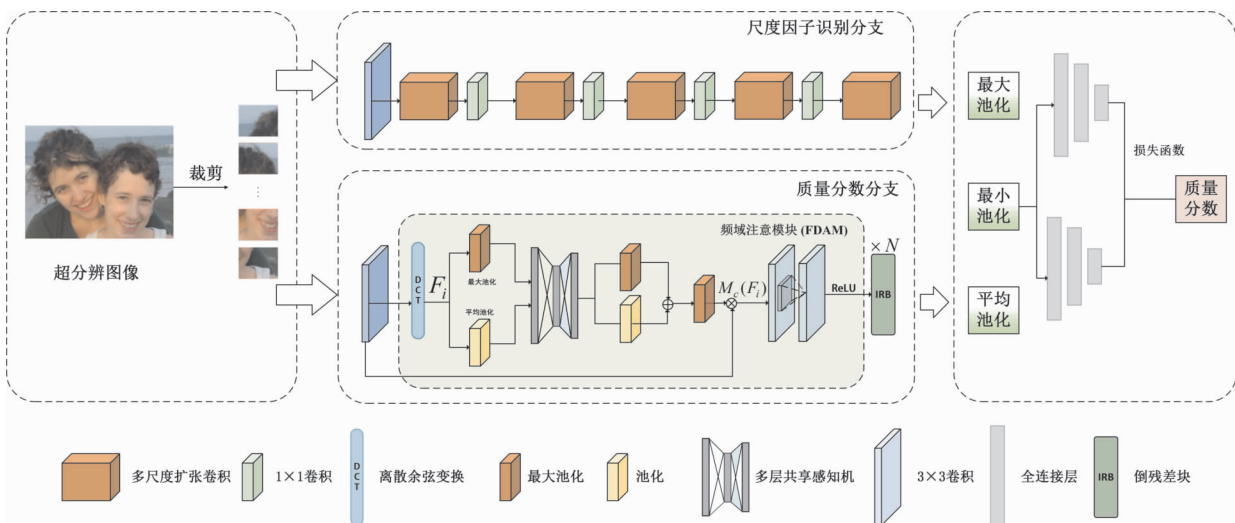


图 1 UFACL 整体框架

Fig. 1 The framework of UFACL

在图 1 中,首先,将超分辨率图像随机裁剪为  $32 \times 32$  大小的图像块,并为每个图像块分配对应的质量分数标签,之后,在尺度因子识别分支,输入图像块经过 1 个  $3 \times 3$  大小的卷积层,提取超分辨率图像块的浅层特征,然后,由 5 个相同的多尺度扩张卷积模块(multiscale extended convolution module, MECM)提取多尺度特征,输出特征经过池化操作获取特征张量,并通过全连接层完成尺度因子的分类。在质量评价分支,使用一个  $3 \times 3$  大小的卷积提取浅层特征,然后利用频域注意模块(frequency domain attention module, FDAM)从图像中提取频域信息的同时获得全局信息和通道信息,之后,使用一组堆叠的倒残差块(inverted residuals blocks, IRB)提取图像的深层语义特征并提升模型训练效率,再利用池化操作完成特征压

缩,并输入到全连接层进行非线性质量分数的映射,在训练过程中使用损失函数计算总损失,更新模型参数,对图像块进行质量分数预测。

### 1.1 多尺度扩张卷积模块

与单一尺度相比,多尺度学习可以将不同尺度的信息有效地融合在一起<sup>[19]</sup>,通过同时处理来自多个尺度的输入数据,模型可以更全面地提取数据的特征和结构。因此,在尺度因子识别中,首先通过一个  $3 \times 3$  大小的卷积提取超分辨率图像的浅层特征,之后使用 5 个相同的多尺度扩张卷积模块(MECM)获取多尺度信息,每一个 MECM 后都使用  $1 \times 1$  大小的卷积进行通道上的降维。该模块共设计了 4 种不同空洞率的卷积,空洞率越大,感受野越大。如图 2 所示,Conv $3 \times 3$ -1 意为卷积核尺寸-空洞率,使用不同空洞率大小的卷积提

取超分辨图像特征并进行拼接。

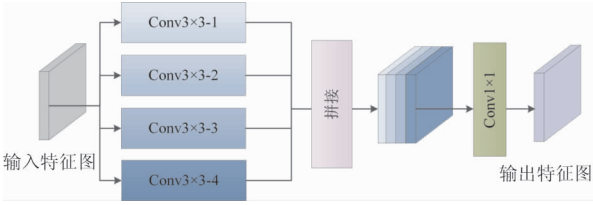


图 2 多尺度扩张卷积模块

Fig. 2 Multiscale extended convolution module

此外,具有不同感受野的特征图对尺度因子的识别任务具有不同程度的重要性,受文献[20]的启发,在特征拼接之后使用  $1 \times 1$  大小的卷积为每层特征图都分配相应的权重,调整特征图的依赖性。MECM 的具体操作可以表示为

$$MECM = \text{Conv}_1(C(\text{Conv}_3^r(X))) \quad (1)$$

式中:Conv 表示卷积操作,下标表示卷积核大小; $r$  表示空洞率( $r = 1, 2, 3, 4$ ); $X$  表示输入的浅层特征; $C$  表示在通道维度进行拼接操作。

### 1.2 对比学习构建

为了让尺度因子识别分支的精确度更高,本文在完成分类任务的同时引入对比学习。在构建对比学习的过程中,将来自同一种尺度因子下的超分辨图像块作为正样本,来自不同尺度因子下的超分辨图像块作为负样本,提高有效特征的表达能力。具体构建方式如图 3 所示,以锚点样本尺度因子  $s = 2$  举例,将尺度因子  $s = 3, 4, 5, 6, 8$  的图像块均作为该图像块的负样本,尺度因子为 2 的其他图像块为正样本。通过不断学习拉近相同尺度因子下样本之间的距离,推远不同尺度因子下样本之间的距离,让模型训练阶段学习到更具鲁棒性和判别能力的特征,获取更好的视觉表征,从而提高模型的泛化能力和表现性能。

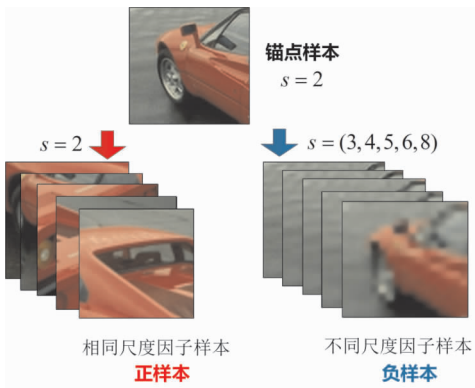


图 3 构建对比学习

Fig. 3 The construction of contrastive learning

### 1.3 频域注意模块

为了进一步提高所提取特征的代表能力,本文提出一种新的频域注意模块(frequency domain attention module, FDAM), FDAM 同时考虑了全局空间信息和通道信息。具体来说,将提取的低级特征首先经过二维离散余弦变换(two-dimensional discrete cosine transform, 2D-DCT)提取特征的频域信息  $F_i$ , 之后再经过池化和多层感知机的特征与输入图像块进行逐像素的相乘,得到特征  $M_c(F_i)$ ,最后经过  $3 \times 3$  的卷积操作提升特征维度。FDAM 的具体操作过程可以描述为

$$M_c(F_i) = \sigma(\text{MLP}(\text{maxPool}(F_i), W_{\text{mlp}}) + \text{MLP}(\text{AvgPool}(F_i), W_{\text{mlp}})) \quad (2)$$

式中: $\sigma$  为 sigmoid 激活函数;MLP 为多层感知机;maxPool 和 AvgPool 分别为最大池化和平均池化; $W_{\text{mlp}}$  为共享多层感知机的权重。

### 1.4 倒残差块

受 MobileNet-V2 倒残差块(inverted residuals block, IRB)模块的启发,本文利用堆叠的 IRB 和特殊的残差连接提取特征,如图 4 所示。与 MobileNet-V2 不同的是,残差连接是在输入特征  $C_{\text{in}}$  与输出特征  $C_{\text{out}}$  之间的特征映射通道不匹配,扩展系数  $E$  也不对称的情况下使用的。

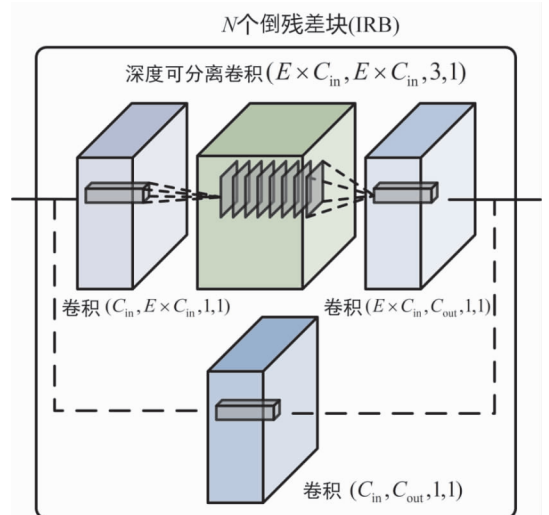


图 4 轻量化模块

Fig. 4 Lightweight module

### 1.5 损失函数设计

#### 1.5.1 联合回归损失

联合回归损失是图像质量评价任务中应用最广泛的损失函数,可表示为

$$L_1 = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

式中: $y_i$ 表示预测图像质量分数; $\hat{y}_i^l$ 表示图像真值; $N$ 表示切分的图像块个数。在测试阶段,给定的超分辨率图像的最终质量分数是所有图像块的预测质量得分平均值。

### 1.5.2 对比损失

对于重建算法识别任务,为了在模型训练阶段帮助获取一个更好的视觉表征,本文应用有监督的对比学习策略。具体来说,将同一类的样本作为正样本,不同类的样本作为负样本,构建正负样本的目的就是为了拉近锚点样本与正样本之间的距离,推远锚点样本与负样本之间的距离。由于超分辨率图像质量评价很大程度上与其生成算法有相关性,因此,我们根据数据集所提供的不同超分辨率图像生成算法构造正负样本,挖掘更多来源于样本间的监督信号,对比损失可表示为

$$l_i = -\log \frac{\sum_{j_i} \exp(\text{Sim}(r_i, r_j) \cdot \tau)}{J_i \sum_{k=1}^N I_{k \neq i} \exp(\text{Sim}(r_i, r_k) \cdot \tau)} \quad (4)$$

$$L_{\text{contrast}} = \frac{1}{N} \sum_{i=1}^N l_i \quad (5)$$

式中: $\text{Sim}(r_i, r_j)$ 表示 $r_i$ 和 $r_j$ 的余弦相似度; $I_{k \neq i}$ 是指示函数,如果 $k \neq i$ 则表示为1,否则为0; $\tau$ 表示用于控制分布密度水平的温度参数; $J_i$ 表示样本 $i$ 的正例个数; $N$ 表示图像块个数。

因此,本文通过 $L_l$ 联合回归损失计算质量分数的回归损失,通过交叉熵损失 $L_{\text{cls}}$ 计算分类误差,联合对比损失 $L_{\text{contrast}}$ 共同组成最终的损失函数,并以此训练模型。具体可表示为

$$L = \alpha L_{\text{cls}} + \beta L_{\text{contrast}} + (1 - \alpha - \beta) L_l \quad (6)$$

式中: $\alpha$ 和 $\beta$ 表示权值系数,按照经验设置为0.33。

## 2 实验结果与分析

### 2.1 实现细节

本文在MA和QADS两个基准超分辨率图像质量评价数据集上进行实验。为了扩充数据集,将所有超分辨率图像进行局部对比度归一化后裁剪成不重叠的大小为 $32 \times 32$ 的图像块,每个图像块的标签对应原始图像标签值。具体的训练参数设置:batch\_size为128;初始学习率为0.001;模型的迭代次数为100;优化器为Adam,使用联合回归损失、分类损失、对比损失对模型进行优化。模型在验证和测试阶段,将同一幅图像中图像块的平均预测质量分数作为图像最终的预测分数。本文实验均在配备GTXT2080Ti GPU的服务器中完成,并在Pytorch 1.6.0编程环境中实现。

### 2.2 超分辨率图像数据库

为了更好地验证UFACL方法的性能,本文使用的超分辨率图像质量评价数据集为MA和QADS数据集,它们的数量规模、图像来源、下采样程度都不同,可以用于评价不同方法的性能。其中,MA数据集将30张原始图像在尺度因子为 $s$ ( $s=2,3,4,5,6,8$ )、高斯核宽度为sigma(sigma=0.8,1.0,1.2,1.6,1.8,2.0)的情况下进行下采样,获取到低分辨率图像,再由9种超分辨率算法生成1620张高分辨率图像,MOS标签范围从0到10。QADS数据集则来源于20张原始图像,在尺度因子为 $s$ ( $s=2,3,4$ )的情况下,由21种超分辨率算法重建获得980张超分辨率图像,其中,包括4种基于插值的方法、11种基于字典的方法,以及6种基于DNN的方法,MOS范围从0到1。2个数据集的具体细节如表1所示,后续所有实验均在这2个数据集上完成。

表1 超分辨率质量评价数据库

Tab.1 The super-resolution quality assessment database

数据集	参考图像数	尺度因子	重建算法类别	超分辨率图像数	标签类型	标签范围
MA	30	6	9	1 620	MOS	[0, 10]
QADS	20	3	21	980	MOS	[0, 1]

### 2.3 评价指标

对于客观质量评价算法模型,一般需要具备以下2个方面的要求:一方面是模型预测应具备准确性,即客观质量评价模型的预测结果应该与主观质量分数相接近;另一方面是预测的相关性

与单调性,即客观质量评价模型的预测结果应该与人眼视觉感官具有高度的一致性。在图像质量评价领域,通常采用斯皮尔曼相关系数(Spearman rank-order correlation coefficient, SROCC)、皮尔逊相关系数(Pearson linear correlation coefficient,

PLCC)、肯德尔秩相关系数(Kendall rank ordered correlation coefficient, KROCC)和均方根误差(root mean squared error, RMSE)4 个定量指标来客观评价超分辨图像质量评价方法的性能。

1) RMSE 用于比较主客观评价值之间的绝对误差值, RMSE 值越小, 说明图像质量越好, RMSE 计算公式为

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - f_i)^2} \quad (7)$$

式中:  $s_i$  表示第  $i$  幅待测图像的 MOS 值;  $f_i$  表示待测图像的预测质量分数;  $N$  表示待测图像数。

2) KROCC 及 SROCC 用于衡量模型预测的单调性。其中, SROCC 比较主客观评分值在各自数据序列中的序号数之差, KROCC 比较主客观评分值中序号对一致与不一致的个数之差。SROCC 和 KROCC 值越大, 说明性能越好。SROCC 和 KROCC 计算过程分别为式(8)和式(9)。

$$SROCC = 1 - \frac{6 \sum_{i=1}^N (r_{s_i} - r_{x_i})}{N(N^2 - 1)} \quad (8)$$

式中:  $r_{s_i}$  表示第  $i$  个待测图像的 MOS 排名值;  $r_{x_i}$  表示第  $i$  个待测图像的预测分数排名值;  $N$  表示待测图像数。

$$KROCC = 1 - \frac{2N_c - N_d}{N(N - 1)} \quad (9)$$

表 2 在基准数据库 QADS 和 MA 上的质量评价算法的性能比较结果

Tab. 2 Comparison results of different IQA methods on the QADS and MA databases

评价方法	QADS				MA			
	SROCC ↑	PLCC ↑	KROCC ↑	RMSE ↓	SROCC ↑	PLCC ↑	KROCC ↑	RMSE ↓
SSEQ <sup>[21]</sup>	0.867 9	0.864 3	0.688 7	0.150 2	0.885 4	0.883 2	0.701 3	1.108 7
BLINDS-II <sup>[22]</sup>	0.888 9	0.883 8	0.710 0	0.143 7	0.898 3	0.892 1	0.725 2	1.262 1
Ma <sup>[15]</sup>	0.895 4	0.896 4	0.728 0	0.117 4	0.913 9	0.925 8	0.753 1	0.883 3
CNN <sup>[23]</sup>	0.953 3	0.950 2	0.811 4	0.094 3	0.922 6	0.936 4	0.759 9	0.949 4
CNN ++ <sup>[24]</sup>	0.952 5	0.950 9	0.816 0	0.093 2	0.931 2	0.930 7	0.771 0	0.987 4
Two-stream <sup>[25]</sup>	0.954 1	0.958 4	0.819 2	0.104 7	0.942 4	0.942 3	0.793 1	0.964 5
DBCNN <sup>[26]</sup>	0.957 5	0.944 9	0.820 0	0.092 1	0.946 5	0.950 9	0.802 1	0.838 9
HyperIQA <sup>[27]</sup>	0.954 1	0.956 8	0.815 1	0.098 6	0.932 7	0.928 4	0.771 7	1.016 6
C <sup>2</sup> MT <sup>[28]</sup>	0.969 0	<u>0.971 9</u>	0.853 5	0.091 8	<u>0.947 1</u>	0.943 6	0.796 0	1.180 4
TADSR <sup>[29]</sup>	<u>0.969 4</u>	0.971 1	<b>0.857 9</b>	<u>0.069 5</u>	0.945 3	<u>0.947 3</u>	<u>0.802 4</u>	<u>0.812 1</u>
本文方法	<b>0.969 7</b>	<b>0.972 7</b>	<u>0.854 9</u>	<b>0.066 9</b>	<b>0.947 7</b>	<b>0.957 5</b>	<b>0.808 2</b>	<b>0.749 4</b>

式中:  $N_c$  表示序号对一致的图像数;  $N_d$  表示序号对不一致的图像数,  $N$  表示待测图像数。

3) PLCC 用于衡量模型的相关性, PLCC 值越大, 说明模型性能越好, PLCC 指标用公式表示为

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

式中:  $s_i$  和  $f_i$  分别表示第  $i$  幅待测图像的 MOS 值和预测质量分数;  $\bar{s}$  和  $\bar{f}$  分别表示待测图像的 MOS 值和预测质量分数的均值;  $N$  表示待测图像数。

### 2.4 不同方法的对比实验

为了验证本文方法的有效性, 在 QADS 和 MA 数据集上与 3 种基于传统机器学习的主流图像质量评价算法 SSEQ<sup>[21]</sup>、BLINDS-II<sup>[22]</sup> 和 Ma<sup>[15]</sup> 以及 7 种基于深度学习的先进的图像质量评价算法 CNN<sup>[23]</sup>、CNN ++<sup>[24]</sup>、Two-stream<sup>[25]</sup>、DBCNN<sup>[26]</sup>、HyperIQA<sup>[27]</sup> 和最新提出的方法 C<sup>2</sup>MT<sup>[28]</sup>、TADSR<sup>[29]</sup> 进行比较, 所有的对比方法均由作者提供的公开代码在推荐参数配置下实现。在 2 个基准数据集上的实验结果如表 2 所示, 其中, 最优指标用加粗表示, 次优用下划线表示。

根据表 2 可以看到,与其他所有的对比方法相比,本文方法 UFACL 在 MA 数据集上的所有指标上均取得了最好的效果,在 QADS 数据集上的 SROCC、PLCC 和 RMSE 指标值达到最优,KROCC 指标也具有一定的可比性。其中,最优的 SROCC 和 KROCC 值表示所提出的 UFACL 方法预测的质量分数与超分辨图像的主观分数之间存在很好的单调性;最优的 PLCC 表示所提的 UFACL 方法获取的预测质量分数与原图像主观分数之间存在最好的线性相关性;此外,最低的 RMSE 值则表示,本文方法获取到了最好的预测精度。值得注意的是,相较于 TADSR,本文方法在 2 个数据集上的大部分指标都有了不同程度的提升,主要表现在 QADS 数据集上的 SROCC、PLCC 指标分别提升了 0.03% 和 0.16%, RMSE 值下降了 0.26%。在 MA 数据集上 SROCC、PLCC 和 KROCC 分别提升了 0.24%、1.02% 和 0.58%。这主要是因为本文方法 UFACL 考虑到了尺度因

子与超分辨图像质量之间的相关性,有利于质量评价任务学习到更加全面的感知特征,从而对超分辨图像质量分数进行更加精准的预测。

为定性分析不同质量评价模型的预测分数与主观分数之间的一致性,图 5 分别展示了其他图像质量评价算法与本文方法在 QADS 和 MA 数据集上绘制的散点图,蓝色和绿色的点分别表示 MA 和 QADS 数据集的主客观散点图。其中,水平轴表示超分辨图像的主观质量分数,纵坐标表示不同方法的预测质量分数。每个点对应一张测试图像,红线表示拟合结果。如果预测分数与主观分数的散点与拟合线越接近,则表明该方法对超分辨图像质量分数的预测得分与人类主观感知得分具有很强的一致性。从图 5 可以清楚看出,基于深度学习的方法相较于传统机器学习方式有明显优势,并且本文提出的 UFACL 方法所预测的分数与主观 MOS 分数要具有更好的一致性。

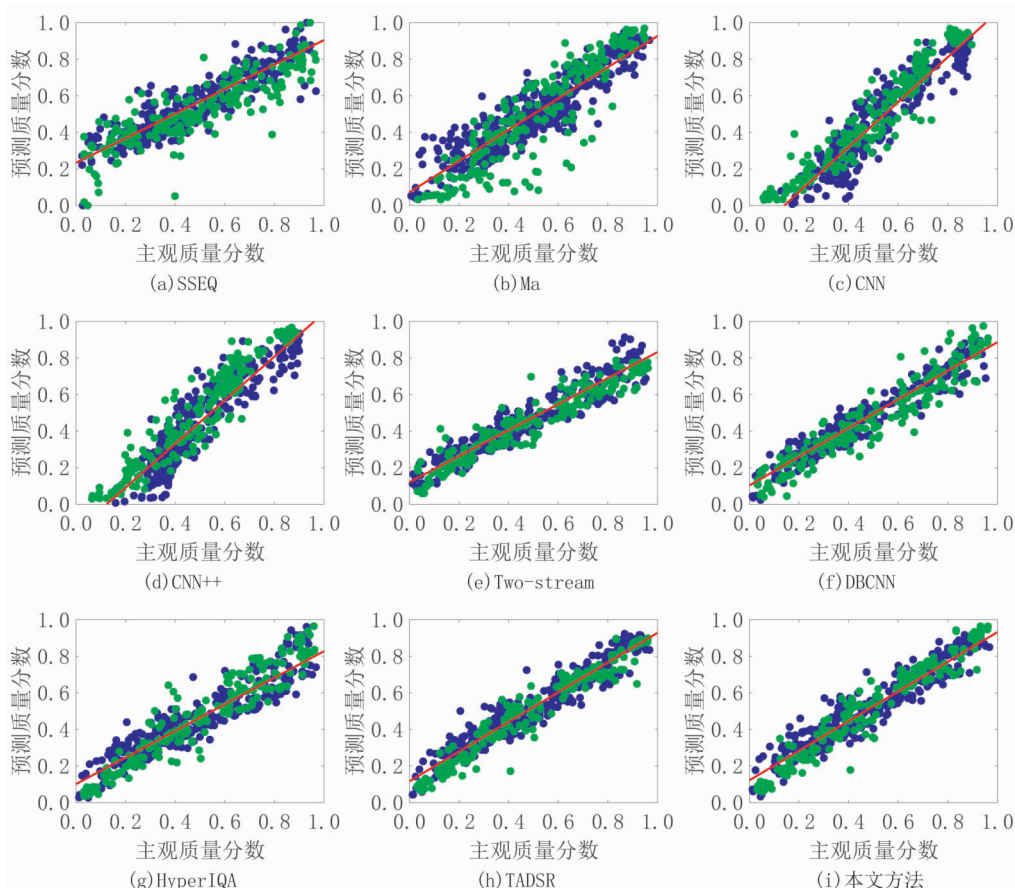


图 5 超分辨图像数据集上质量评价算法预测图像质量的散点图

Fig. 5 Scatter plot of image quality predicted by quality evaluation algorithm on the SR image databases

### 2.5 消融实验

#### 2.5.1 轻量级主干设计

在提出的 UFACL 中,关键的任务之一是构建一个有效的轻量级主干网络,提取与感知质量分数相关的内在特征。本文选择 MobileNetV2 作为 UFACL 的骨干网络,因为它的轻量级架构以及具有倒残差和线性瓶颈的性能。为了证实选择的合理性,使用另外 2 个轻量级模型 ShuffleNet<sup>[30]</sup> 和 MobileNet-V1<sup>[31]</sup> 进行性能比较。如图 6 所示,符号“√”表示使用该主干网,“×”表示未使用。蓝色行和红色行分别表示模型使用 ShuffleNet 和 MobileNet-V1 作为主干网络,灰色行表示使用本文提出的 MobileNet-V2 作为主干网络。从比较的结果可以发现,MobileNet-V2 在 MA 和 QADS 数据集上比 ShuffleNet 和 MobileNet-V1 获得更好的性能。在 MA 数据集上,相比较于 MobileNet-V1,4 个指标都是最优的,在 QADS 数据集上,SROCC、PLCC 和 RMSE 与另外 2 个轻量级骨干网络也具有一定的可比性。

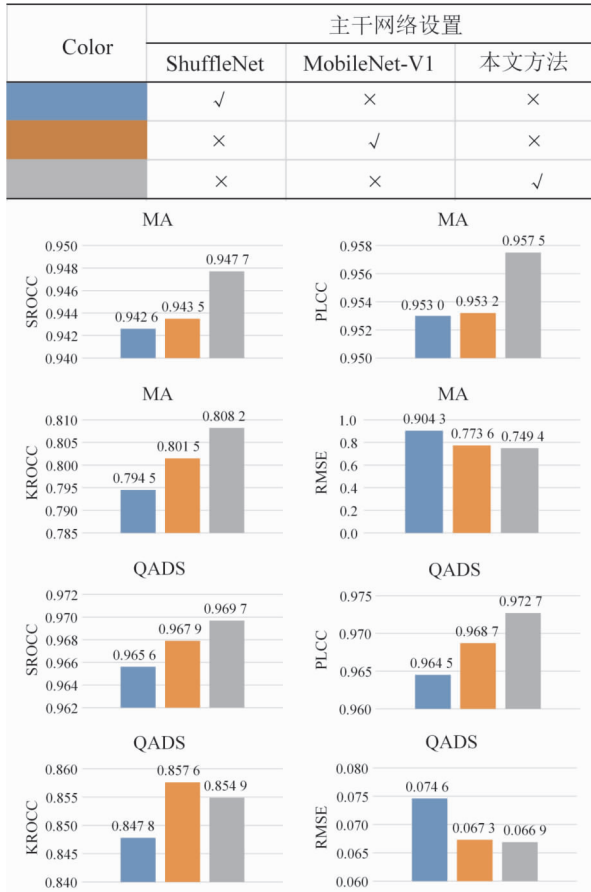


图 6 轻量级主干设计消融实验

Fig. 6 The ablation study of backbone network

#### 2.5.2 损失函数消融实验

为了验证模型训练过程中使用多损失对模型性能的影响,设计了 3 种情况进行对比。如图 7 所示, $L_1$  表示回归损失, $L_d$  表示分类损失, $L_c$  表示对比损失,符号“√”表示使用该损失,“×”表示未使用。蓝色行表示在训练过程中只使用  $L_1$  损失,黄色行表示使用  $L_1$  损失和  $L_d$  损失相结合,红色行表示本文方法 3 个损失同时使用。可以看出,在只使用  $L_1$  损失时,除在 QADS 数据集上 RMSE 指标最优外,其他指标性能都较差;当  $L_1$  损失和  $L_d$  损失相结合时,在 2 个数据集上的所有指标值均有提升;当 3 个损失共同使用时,除在 QADS 数据集上的 RMSE 外,所有指标均达到最优,这说明对比损失和分类损失在一定程度上提高了分类的精确性,尺度因子分支学习到的特征对质量评价任务产生了一定的积极影响。

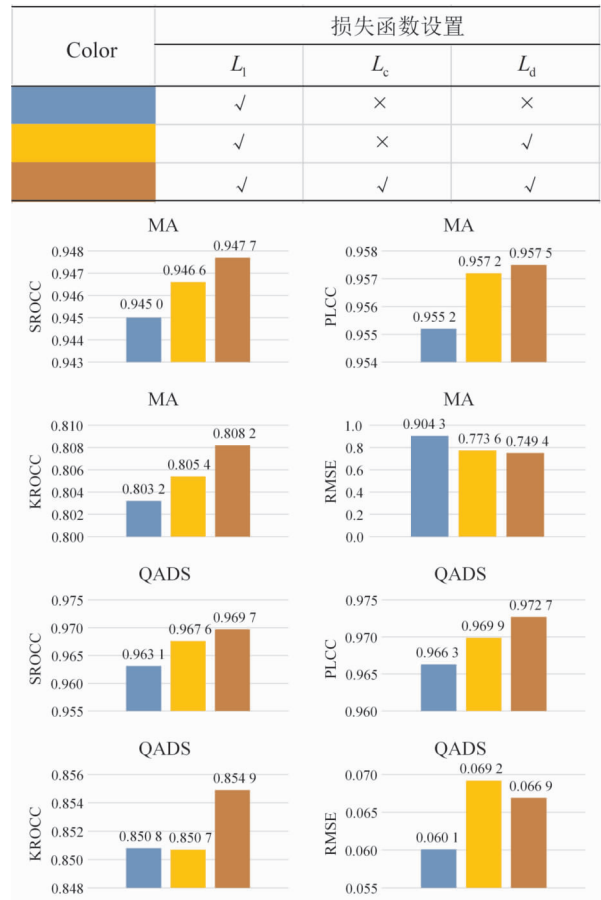


图 7 损失函数消融实验

Fig. 7 The ablation study of loss function

### 2.6 泛化实验

为了进一步验证模型的泛化能力,我们分别对其他几种经典的基于深度学习的方法与本文方法 UFACL 进行跨数据集实验,数据集采用

QADS、MA 以及 SISRSet<sup>[32]</sup>。其中, QADS 和 MA 分别选取数据集的 20% 作为测试集图像, SISRSet 数据集共 360 张超分辨率图像, 但图像大小不一, 因此我们去掉部分尺度较大的图像, 最终选取 312 张图像进行测试。首先, 在 QADS 数据集上的交叉验证结果如表 3 所示, 当在 SISRSet 数据集上测试时, 本文方法 UFACL 表现较好。SROCC、KROCC、PLCC 分别达到 0.903 9、0.820 1

和 0.816 6 的指标值, 相较于其他方法有明显提升, 这主要是因为 UFACL 考虑到尺度信息对于超分辨率图像质量的相关性。当在 MA 数据集上测试时, UFACL 也表现出中等的性能, HyperIQA 表现最优, TADSR 次之, 这主要是由于 HyperIQA 考虑到图像内容对质量评价的影响, 更接近于人眼视觉感知。

表 3 验证 QADS 数据集上模型的泛化性能实验

Tab.3 The generalization performance of the model on the QADS dataset

评价方法	在 SISRSet 上测试			在 MA 上测试		
	SROCC ↑	KROCC ↑	PLCC ↑	SROCC ↑	KROCC ↑	PLCC ↑
CNN <sup>[23]</sup>	<b>0.914 4</b>	<u>0.748 7</u>	<b>0.905 5</b>	0.660 3	0.484 3	<u>0.735 6</u>
Two-stream <sup>[25]</sup>	0.901 0	0.729 6	0.890 7	0.581 9	0.416 5	0.669 3
DBCNN <sup>[26]</sup>	0.676 9	0.487 6	0.653 0	0.653 3	0.459 5	0.696 3
HyperIQA <sup>[27]</sup>	0.807 5	0.616 0	0.797 2	<b>0.809 5</b>	<b>0.623 5</b>	<b>0.812 8</b>
TADSR <sup>[29]</sup>	0.885 9	0.714 1	0.868 5	<u>0.6643</u>	<u>0.491 9</u>	0.703 6
JCSAN <sup>[33]</sup>	0.903 6	0.734 5	<u>0.901 4</u>	0.630 5	0.462 9	0.728 8
本文方法	<u>0.903 9</u>	<b>0.820 1</b>	0.816 6	0.625 0	0.445 8	0.730 5

如表 4 所示, 在 MA 数据集上训练且在 SISR-Set 上测试时, 本文方法表现最好, CNN 次之。而在 QADS 数据集上测试, DBCNN 取得了最优结果, 这说明采用预训练模型对于提升模型的泛化

性能有一定的优势。UFACL 虽然考虑到尺度信息对于超分辨率图像质量评价的影响, 但在小数据集上训练的结果, 由于没有使用预训练模型, 因此并没有表现出明显优势, 处于中等水平。

表 4 验证 MA 数据集上模型的泛化性能实验

Tab.4 The generalization performance of the model on the MA dataset

评价方法	在 SISRSet 上测试			在 QADS 上测试		
	SROCC ↑	KROCC ↑	PLCC ↑	SROCC ↑	KROCC ↑	PLCC ↑
CNN <sup>[23]</sup>	<u>0.880 6</u>	<u>0.700 4</u>	<b>0.879 5</b>	<u>0.774 3</u>	<u>0.592 0</u>	<u>0.768 9</u>
Two-stream <sup>[25]</sup>	0.795 2	0.597 4	0.787 8	0.661 9	0.475 0	0.668 1
DBCNN <sup>[26]</sup>	0.837 2	0.643 0	<u>0.824 7</u>	<b>0.808 7</b>	<b>0.616 0</b>	<b>0.808 1</b>
HyperIQA <sup>[27]</sup>	0.774 9	0.575 6	0.753 6	0.768 7	0.573 0	0.745 7
TADSR <sup>[29]</sup>	0.802 1	0.602 5	0.807 1	0.698 9	0.518 8	0.714 6
JCSAN <sup>[33]</sup>	0.754 9	0.573 3	0.753 3	0.702 0	0.529 8	0.707 8
本文方法	<b>0.925 5</b>	<b>0.827 6</b>	0.763 0	0.658 9	0.499 8	0.645 8

### 3 结语

本文提出了一种基于尺度因子感知对比学习的超分辨率图像质量评价方法 UFACL。该方法主要分为尺度因子识别和质量分数预测 2 个分支。在尺度因子识别阶段, 采用有监督的对比学习来

学习不同尺度因子的特征表示, 提高模型识别尺度因子的准确性。在质量分数预测阶段, 通过频域注意模块获取图像的频域信息, 并使用轻量级网络提升模型训练效率, 然后使用回归网络预测图像的质量分数。实验结果表明, 与其他方法相比, 本文提出的方法在性能上表现更优。

尽管 UFACL 方法在基准数据集上取得了优

异的结果,但仍存在一些不可忽视的局限性。由于不同数据集的标签尺度不统一,导致模型在跨数据集验证时表现不佳,对特定数据集的依赖性限制了 UFACL 的应用范围。未来研究中,我们将探索把排序学习应用到图像质量分数预测模型中,以更准确地量化超分辨率图像的质量。同时,将 PLCC 指标作为损失函数约束来优化模型训练,进一步提高超分辨图像质量评价(SRIQA)的性能。

## 参考文献

- [1] 李云红,王珍,张凯兵,等. 基于学习的图像超分辨率重建方法综述[J]. 计算机工程与应用,2018,54(15):13-21.  
LI Y H, WANG Z, ZHANG K B, et al. Survey on example learning-based single image super-resolution technique[J]. Computer Engineering and Applications, 2018, 54(15): 13-21.
- [2] LI K, YANG S H, DONG R T, et al. Survey of single image super-resolution reconstruction[J]. IET Image Processing, 2020, 14(11): 2273-2290.
- [3] 崔琛,张凯兵. 基于双字典正则化的单帧图像超分辨率重建方法[J]. 西安工程大学学报,2021,35(2):66-72.  
CUI C, ZHANG K B. Learning dual dictionary regularization for single image super-resolution[J]. Journal of Xi'an Polytechnic University, 2021, 35(2): 66-72.
- [4] 张凯兵,朱丹妮,王珍,等. 超分辨图像质量评价综述[J]. 计算机工程与应用,2019,55(4):31-40.  
ZHANG K B, ZHU D N, WANG Z, et al. Survey of super-resolution images quality assessment[J]. Computer Engineering and Applications, 2019, 55(4): 31-40.
- [5] MANTIUK R K, TOMASZEWSKA A, MANTIUK R. Comparison of four subjective methods for image quality assessment[J]. Computer Graphics Forum, 2012, 31(8): 2478-2491.
- [6] HE L H, GAO F, HOU W L, et al. Objective image quality assessment: A survey[J]. International Journal of Computer Mathematics, 2014, 91(11): 2374-2388.
- [7] 张婷悦,张凯兵. 基于稀疏表示的无参考型超分辨图像质量评价方法[J]. 西安工程大学学报,2020,34(5):20-26.  
ZHANG T Y, ZHANG K B. No-reference super-resolution image quality assessment based on sparse representation[J]. Journal of Xi'an Polytechnic University, 2020, 34(5): 20-26.
- [8] 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.
- [9] YE P, KUMAR J, DOERMANN D. Beyond human opinion scores: Blind image quality assessment based on synthetic scores[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 4241-4248.
- [10] GU K, ZHAI G T, YANG X K, et al. Hybrid no-reference quality metric for singly and multiply distorted images[J]. IEEE Transactions on Broadcasting, 2014, 60(3): 555-567.
- [11] YANG J, LYU M J, QI Z Q, et al. Deep learning based image quality assessment: A survey[J]. Procedia Computer Science, 2023, 221: 1000-1005.
- [12] LIU X L, VAN DE WEIJER J, BAGDANOV A D. RankIQA: Learning from rankings for no-reference image quality assessment[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1040-1049.
- [13] 艾娜,彭进业,王珺,等. 基于低分辨率图像自举和双重稀疏性字典训练的单幅图像超分辨率重构[J]. 西北大学学报(自然科学版),2018,48(1):57-65.  
AI N, PENG J Y, WANG J, et al. Single image super-resolution based on bootstrapping method combining learned double sparsity dictionaries[J]. Journal of Northwest University(Natural Science Edition), 2018, 48(1): 57-65.
- [14] BOURBIA S, KARINE A, CHETOUANI A, et al. A multi-task convolutional neural network for blind stereoscopic image quality assessment using naturalness analysis[C]//2021 IEEE International Conference on Image Processing (ICIP). Anchorage: IEEE, 2021: 1434-1438.
- [15] LI F, ZHANG Y F, 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.
- [16] 李云红,刘杏瑞,谢蓉蓉,等. 基于超分辨重建和公共特征子空间的低分辨率人脸识别[J]. 西北大学学报(自然科学版),2023,53(2):241-247.  
LI Y H, LIU X R, XIE R R, et al. Low-resolution

- face recognition based on super-resolution reconstruction and common feature subspace [J]. *Journal of Northwest University (Natural Science Edition)*, 2023, 53(2): 241-247.
- [17] FU J. Scale guided hypernetwork for blind super-resolution image quality assessment[EB/OL]. (2023-06-04) [2024-09-20]. <https://arxiv.org/abs/2306.02398v1>.
- [18] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 4510-4520.
- [19] ZHANG H N, MENG F, HAN Y W. No-reference image quality assessment based on a multi-feature extraction network[C]//Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing. Singapore: ACM, 2020: 81-85.
- [20] WANG T, ZHANG T, ZHANG K B, et al. Context attention fusion network for crowd counting [J]. *Knowledge-Based Systems*, 2023, 271: 110541.
- [21] SAAD M A, BOVIK A C, CHARRIER C. DCT statistics model-based blind image quality assessment[C]//2011 18th IEEE International Conference on Image Processing. Brussels: IEEE, 2011: 3093-3096.
- [22] LIU L X, LIU B, HUANG H, et al. No-reference image quality assessment based on spatial and spectral entropies[J]. *Signal Processing: Image Communication*, 2014, 29(8): 856-863.
- [23] KANG L, YE P, LI Y, et al. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks [C]//2015 IEEE International Conference on Image Processing (ICIP). Quebec City: IEEE, 2015: 2791-2795.
- [24] KANG L, YE P, LI Y, et al. Convolutional neural networks for no-reference image quality assessment [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 1733-1740.
- [25] 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.
- [26] 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.
- [27] 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]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA: IEEE, 2020: 3667-3676.
- [28] LI H, ZHANG K B, NIU Z X, et al. C<sup>2</sup>MT: A credible and class-aware multi-task transformer for SR-IQA [J]. *IEEE Signal Processing Letters*, 2022, 29: 2662-2666.
- [29] QUAN X, ZHANG K B, LI H, et al. TADSRNet: A triple-attention dual-scale residual network for super-resolution image quality assessment[J]. *Applied Intelligence*, 2023, 53(22): 26708-26724.
- [30] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 6848-6856.
- [31] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2024-09-20]. <https://arxiv.org/abs/1704.04861v1>.
- [32] SHI G M, WAN W F, WU J J, et al. SISRSNet: Single image super-resolution subjective evaluation test and objective quality assessment [J]. *Neurocomputing*, 2019, 360: 37-51.
- [33] ZHANG T Y, ZHANG K B, XIAO C, et al. Joint channel-spatial attention network for super-resolution image quality assessment [J]. *Applied Intelligence*, 2022, 52(15): 17118-17132.

(编辑 李静)