



多层特征融合与语义增强的盲图像质量评价

赵文清, 许丽娇, 陈昊阳, 李梦伟

引用本文:

赵文清, 许丽娇, 陈昊阳, 李梦伟. 多层特征融合与语义增强的盲图像质量评价[J]. *智能系统学报*, 2024, 19(1): 132–141.

ZHAO Wenqing, XU Lijiao, CHEN Haoyang, et al. Blind image quality assessment based on multi-level feature fusion and semantic enhancement[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(1): 132–141.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202301007>

您可能感兴趣的其他文章

结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics
智能系统学报. 2021, 16(4): 729–738 <https://dx.doi.org/10.11992/tis.202008007>

改进MobileNet的图像分类方法研究

Research on the improved image classification method of MobileNet
智能系统学报. 2021, 16(1): 11–20 <https://dx.doi.org/10.11992/tis.202012034>

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN
智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection
智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene
智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network
智能系统学报. 2019, 14(6): 1152–1162 <https://dx.doi.org/10.11992/tis.201812003>

DOI: 10.11992/tis.202301007

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20230728.1406.002>

多层特征融合与语义增强的盲图像质量评价

赵文清^{1,2}, 许丽娇¹, 陈昊阳¹, 李梦伟¹

(1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003)

摘要: 针对现有盲图像质量评价算法在面对真实失真图像时性能较差的问题, 本文提出多层特征融合和语义信息增强相结合的无参考图像质量评价算法。提取图像的局部和全局失真特征, 利用特征融合模块对特征进行多层融合; 利用多层扩张卷积增强语义信息, 进而指导失真图像到质量分数的映射过程; 考虑预测分数和主观分数之间的相对排名关系, 对 L_1 损失函数和三元组排名损失函数进行融合, 构建新的损失函数 L_{mix} 。为了验证本文方法的有效性, 在野生图像质量挑战数据集上进行了验证和对比实验, 该算法的斯皮尔曼等级相关系数与皮尔逊线性相关系数指标相比原算法分别提升 2.3% 和 2.3%; 在康斯坦茨真实图像质量数据集和野生图像质量挑战数据集上进行了跨数据集实验, 该算法在面对真实失真图像时表现出了良好的泛化性能。

关键词: 深度学习; 图像质量; 卷积神经网络; 特征提取; 通道注意力结构; 多层次特征融合; 扩张卷积; 三元组损失函数

中图分类号: TP391.41 文献标志码: A 文章编号: 1673-4785(2024)01-0132-10

中文引用格式: 赵文清, 许丽娇, 陈昊阳, 等. 多层特征融合与语义增强的盲图像质量评价 [J]. 智能系统学报, 2024, 19(1): 132-141.

英文引用格式: ZHAO Wenqing, XU Lijiao, CHEN Haoyang, et al. Blind image quality assessment based on multi-level feature fusion and semantic enhancement[J]. CAAI transactions on intelligent systems, 2024, 19(1): 132-141.

Blind image quality assessment based on multi-level feature fusion and semantic enhancement

ZHAO Wenqing^{1,2}, XU Lijiao¹, CHEN Haoyang¹, LI Mengwei¹

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. Engineering Research Center of the Ministry of Education for Intelligent Computing of Complex Energy System Department, Baoding 071003, China)

Abstract: Aiming at the low performance of the existing blind image quality assessment algorithm when facing the real distorted images, the paper proposes a new no-reference image quality assessment algorithm, namely multi-level feature fusion and semantic enhancement for NR (MFFSE-NR), which combines multi-level feature fusion and semantic information enhancement. The local and global distortion features of an image are extracted, then a feature fusion module is used to fuse the features in layers. The multi-layer dilated convolution is employed to enhance semantic information and further direct the mapping process from distorted image to quality fraction. Finally, a novel loss function called L_{mix} is created by combining the triplet ranking loss function and the L_1 loss function, taking account of the relative ranking relationship between the predicted score and the subjective score. Validation and comparison experiments carried out on LIVEC dataset show that both the SROCC and PLCC index are improved respectively by 2.3% than the original algorithm; cross-dataset validation on the KonIQ-10k dataset and LIVEC dataset confirm that the proposed algorithm has good generalization ability when dealing with the real distorted images.

Keywords: deep learning; image quality; convolution neural network; feature extraction; channel attention structure; multi-level feature fusion; dilated convolution; triplet loss function

收稿日期: 2023-01-09. 网络出版日期: 2023-07-31.

基金项目: 国家自然科学基金项目 (61773160, 61871182); 河北省自然科学基金项目 (F2021502013); 中央高校基本科研业务费项目 (2020MS153, 2021PT018).

通信作者: 赵文清. E-mail: jbwzq@126.com.

图像在采集、存储、传输中可能存在着模糊、噪声等失真问题^[1]。图像质量评价 (image quality assessment, IQA) 旨在为各种失真图像进行质量的

评级打分, 对多种图像处理算法的评估、改善图像的视觉体验具有重要意义。最直接的图像质量评价方法是根据人类的视觉感知进行评分, 但是这种方法会被人的主观因素所影响, 且耗时费力。随着计算机技术的不断发展, 在没有人工干预的情况下, 可自动进行图像的质量评价^[2]。全参考图像质量评价算法要求利用高清晰度、无失真图像作为参照, 传统的全参考图像质量评价算法有峰值信噪比算法 (peak signal to noise ratio, PSNR)^[3] 和文献 [4] 提出的结构相似度算法 (structural similarity index measure, SSIM) 等。半参考图像质量评价算法仅利用无失真图像的部分特征即可评估图像的质量, 如文献 [5] 实现的综合多尺度几何分析算法。但是, 现实中的失真图像很难找出对应的无失真图像。无参考图像质量评价 (no reference image quality assessment, NR-IQA) 算法不需要使用参考图像, 通过提取图像的失真特征对图像进行评价, 在实际中有着广泛的应用^[6]。

盲/无参考图像空间质量评估器 (blind/referenceless image spatial quality evaluator, BRISQUE)^[7] 算法将图像的高斯分布特征作为质量分数的回归特征。基于高阶统计聚合的盲图像质量评估 (blind image quality assessment based on high order statistics aggregation, HOSA) 算法^[8] 提取码本中图像的归一化均值、方差和协偏度等特征, 通过计算码本中图像和测试图像之间的特征差异来进行质量分数的回归。以上方法需要从失真图像中提取人为设计的图像特征, 要求人们具有足够的先验知识, 具有一定的局限性且性能提升较慢。

鉴于深度模型具有很强的特征抽取能力, 越来越多的学者将其应用到图像质量评价领域^[9]。用于无参考的加权平均深度图像质量评估 (weighted average deep image quality assessment metric-no reference, WaDIQaM-NR)^[10] 通过堆叠多个卷积层和池化层自动提取与畸变有关的特征, 并将图像分割成许多图像块, 对图像块的分数进行加权操作来得到整幅失真图像的质量分数。分级退化级联卷积神经网络 (cascaded convolutional neural network with hierarchical degradation concatenation, CaHDC)^[11] 考虑了人类视觉系统中分层感知机制, 利用卷积神经网络 (convolutional neural network, CNN) 学习退化特征, 实现质量的预测。深度双线性卷积神经网络 (deep bilinear convolutional neural network, DB-CNN)^[12] 分别设计针对合成失真和真实失真的 CNN 模型来提取失真特征, 采

用双线性池进行结合, 最后利用全连接层进行质量分数的回归。文献 [13] 利用数据驱动的方式, 将大量带有标记的畸变图像映射为质量分数, 使无参考图像质量评价 (no reference image quality assessment, NR-IQA) 算法在合成失真图像数据集上的性能得到了显著提高。然而, 以上方法考虑的是图像退化的全局信息, 而真实情况下, 大多数畸变都是局部畸变, 而且人的视觉系统对局部畸变十分敏感, 从而导致以上算法在面对真实失真图像时表现不佳。自适应超网络引导下的野外图像质量盲评估算法 (blindly assess image quality in the wild guided by a self-adaptive hyper network, HyperIQA)^[14] 首次将不同层次的局部失真特征进行提取聚合。多尺度特征逐层融合的深度神经网络 (deep neural network based on multi-scale features fusion layer-by-layer, MsFF-Net)^[15] 通过逐层融合相邻的尺度特征, 获得了更能精确表征图像质量的多尺度失真特征。从补丁到图片的盲图像质量测量算法 (from patches to pictures blind image quality measurement, P2P-BM)^[16] 构建了基于图像区域的深度架构, 用于学习生成局部的图像质量分数以及全部的图像质量分数。以上 3 种方法充分考虑到了图像细粒度的局部畸变在评价过程中的作用, 在野生图像质量挑战数据集 (live in the wild image quality challenge database, LIVEC)^[17] 和康斯坦茨真实图像质量数据库 (konstanz authentic image quality 10k database, KonIQ-10k)^[18] 上性能得到了明显的提升。

为了兼顾图像的局部和全局信息, 往往对相邻层的特征构建金字塔结构, 但是会造成语义跨度较大的问题, 而且简单的特征融合不能有效地提取失真图像的细节信息^[19]。在面对不同内容时, 人们观察图像的感知方式并不相同, IQA 任务应与图像的语义信息高度相关。真实失真图像中内容多种多样, 存在着各种局部和非局部的失真, 质量评价算法需要同时考虑失真图像中的语义信息和局部信息^[19]。尽管以上方法中的回归损失对于质量预测任务十分有效, 但是没有考虑图像块之间的排序关系^[20]。

因此, 本文针对以上存在的问题, 提出了多层特征融合和语义增强相结合的盲图像质量评价算法 (multi-level feature fusion and semantic enhancement for NR, MFFSE-NR), 将失真图像块输入到深层特征提取网络, 提取出局部和全局特征, 利用特征融合模块对各个特征进行分层融合, 然后对各个层次的特征进行自适应池化, 最后通过全连

接层得到每个输入图像块的分数；为了更好地利用图像语义信息，设计语义特征增强模块，捕获长距离的上下文关系，使用语义信息为全连接层生成参数。考虑预测分数和主观分数之间排名的一致性，设计带有相对排名信息的混合损失函数 L_{mix} 。

1 相关技术和理论

1.1 Res2Net50 网络

本文选取 Res2Net50^[21] 网络作为主干网络，

结构如图 1 所示。Res2Net50 将 ResNet50^[22] 的 bottleneck 中的 3×3 卷积进行多尺度解耦，以进行多尺度的特征提取。首先是 1×1 的卷积运算，之后对通道进行分组，图 1 中组数 scale 为 4，第 1 组的特征向下传递，第 2 组的特征经过一个 3×3 卷积进行特征提取，由此特征提取的感受野随之改变，以此类推，越到后面的组感受野越大，最后将各个组的特征进行拼接还原，再次使用 1×1 卷积融合通道信息来提取同一层次的多尺度特征。本文中 scale 取 4，每次拆分中通道数 width 取 26。

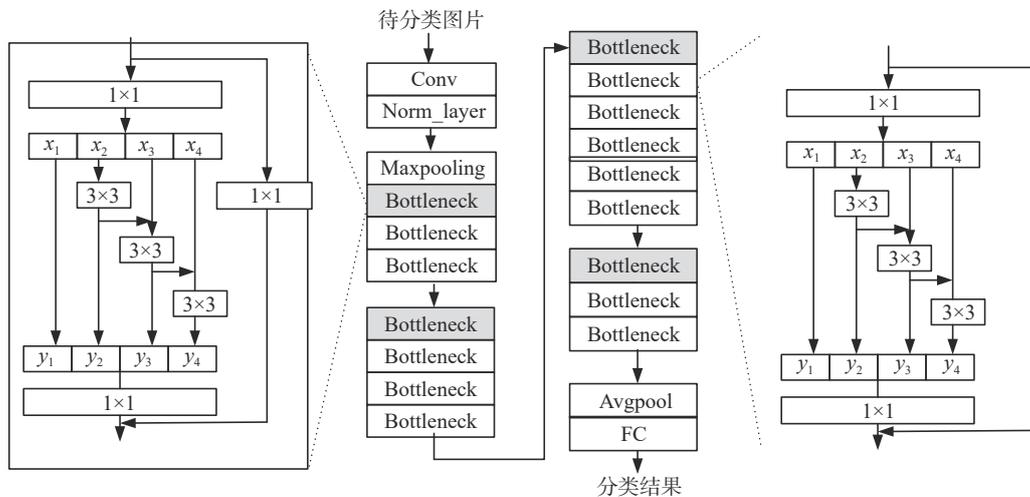


图 1 Res2Net50 整体架构
Fig. 1 Overall structure of Res2Net50

1.2 特征金字塔网络

为了有效识别目标检测任务中的多尺度目标，特征金字塔网络 (feature pyramid networks, FPN) 采用自顶向下的结构将上层具有高级语义信息的特征图进行上采样，与下层具有细节信息的特征图逐元素相加，把来自不同卷积层的不同表征信息融合在一起，实现多层特征的融合和增强。

以 ResNet 作为特征主干网络为例，FPN 选取 conv2、conv3、conv4、conv5 层的最后一个残差块的输出特征组成特征金字塔。每层特征图首先在横向上进行 1×1 卷积，然后将上一层特征图进行自顶向下的上采样，放大到上一层特征图一样的大小，将二者生成的特征图通过相加的方式进行融合，横向连接之后的特征图再次经过 3×3 的卷积得出最终的特征金字塔。

1.3 通道注意力模型

注意力机制可以为网络生成更具有辨别能力的特征，关注更需要关注的信息而忽略无关的信息。SENet 是通道注意力机制的代表模型，可以

嵌入在各种网络结构中改善性能。SENet 首先利用全局平均池化 (global average pooling, GAP) 对 $W \times H \times C$ 的特征图进行压缩，将每个通道上的空间特征进行编码，得到 $1 \times 1 \times C$ 的全局特征图；然后利用 2 个全连接层先降维再进行升维，实现通道之间的交互，得到 $1 \times 1 \times C$ 大小的向量；最后经过非线性激活函数 Sigmoid 生成每个通道的权重，将其与原始特征图相乘，得到加权后的通道特征。

1.4 扩张卷积

扩张卷积最早出现在 DeepLab 系列中，通过在卷积核的各像素点中间加入 0 值的像素点，增大卷积核的尺寸，可以在不牺牲特征图尺寸的前提下增加感受野，提取更多的语义信息，有效避免池化操作过程中造成的信息丢失。当多次叠加的扩张卷积具有相同的扩张率时，由于卷积核存在间隔，不是所有的输入都参与计算，会出现计算中心向外扩散的情况，为此，将扩张率设置为不同值的组合，可以很好地避免这个问题。

1.5 三元组损失函数

深度度量学习以使同类对象之间的距离比较近而不同类的对象之间的距离比较远作为目标。三元组损失最初由谷歌的研究团队提出, 在人脸识别领域被广泛使用, 用于实现对非同类极相似样本的区分^[23]。三元组损失函数为

$$L_{\text{triple}}(a, p, n) = \max\{0, |a, p| - |a, n| + A\} \quad (1)$$

式中: A 为常量, 通过学习, 模型会让锚点 a 和正例 p 的距离值更小, 同时让锚点 a 和负例 n 的距离值更大, 从而实现模型对细节的区分。

2 基于多层特征融合和语义增强的盲图像质量评价网络

本文提出的多层特征融合和语义增强相结合的盲图像质量评价算法 (multi-level feature fusion and semantic enhancement for NR, MFFSE-NR), 能够有效地处理真实失真场景中的各种复杂失真, 如图 2 所示, 模型包含 4 个模块, 分别为多尺度特征提取模块、多层特征融合模块、语义信息增强模块和语义信息指导的质量分数回归模块。

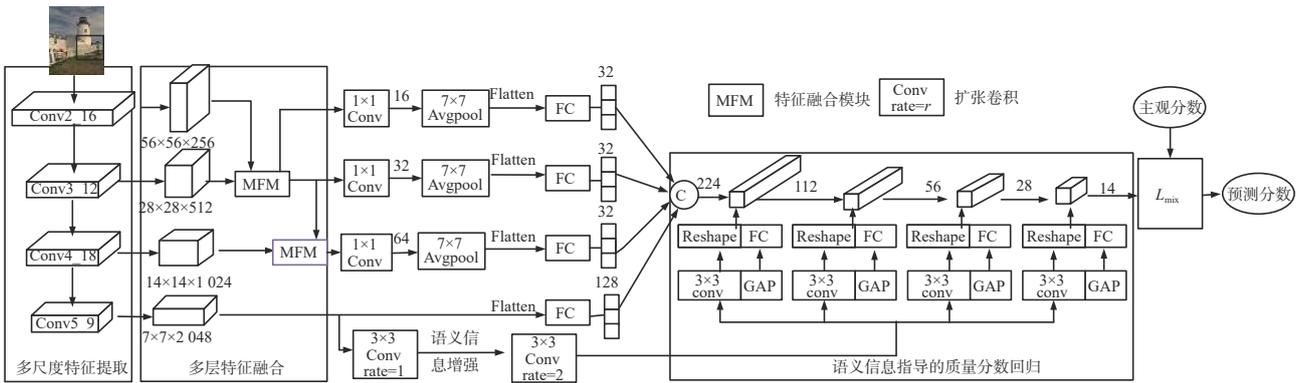


图 2 MFFSE-NR 整体架构

Fig. 2 Overall structure of MFFSE-NR

2.1 多尺度特征提取模块

残差图像含有与图像质量有关的重要信息^[24], 本文选取 Res2Net50 作为失真特征提取的主干网络。随着卷积层的不断叠加, 输出特征图的感受野越大、语义信息越丰富, 最后一层卷积生成的特征包含全局感受野, 具有最高级的语义信息^[25]。本文利用 Res2Net50 的 conv2_10、conv3_12、conv4_18、conv5_9 卷积层提取的特征构建特征金字塔, 捕获

多尺度的畸变信息。特征提取主干网络提取出不同层次的失真特征, 方便后续进行更细粒度的融合。

2.2 多层特征融合模块

由于输入的失真图像存在各种局部的失真, 同时为了获取不同退化层次的表征信息, 本文在特征金字塔结构的基础上, 设计了多层特征融合结构, 提出了更细粒度的特征融合模块, 其结构如图 3 所示。

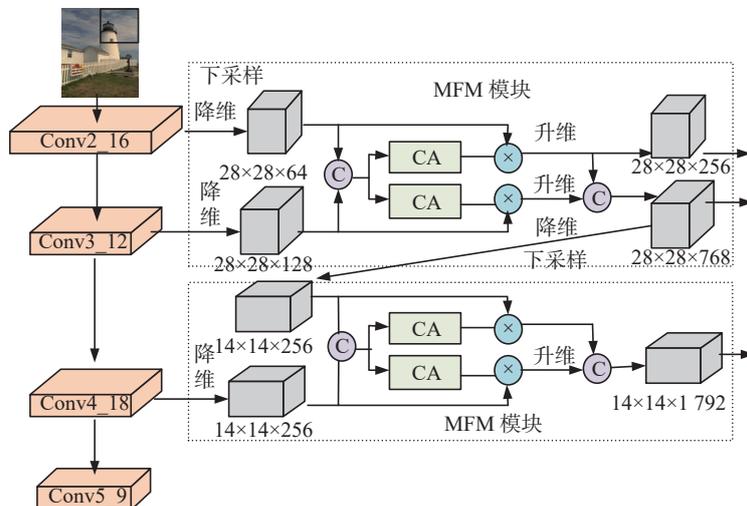


图 3 多层特征融合结构

Fig. 3 Structure of multi level feature fusion

与目标检测任务不同,上采样操作会给失真图像引入不必要的噪声,故本文通过横向连接和自下而上路径构建特征金字塔。首先使用步长为 2 的 3×3 卷积对低级语义特征进行下采样。为了减少计算量,利用 1×1 卷积进行横向的等比例降维,将第 1 层通道数降低至 64,第 2 层降低至 128,第 3 层降低至 256。随后,对第 1 层和第 2 层的特征进行拼接,得到一个融合的特征。为了减少特征融合过程中的冗余信息,通过通道注意力块得出每层特征的通道权重分数,将它们分别与低级语义特征和高级语义特征逐元素相乘,得到精细的特征图,再经过 1×1 大小的卷积核进行等比例升维,拼接生成最终的多尺度融合特征图,用于生成质量感知特征向量,进行质量分数的回归。这一过程为

$$\begin{cases} F_r = \text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(F_i)) \\ F_j = \text{Conv}_{1 \times 1}(F_j) \\ M_c = F_r \otimes F_j \\ W_i = \text{CA}(M_c) \\ W_j = \text{CA}(M_c) \\ F_o = [\text{Conv}_{1 \times 1}(W_i \times F_r)] \otimes [\text{Conv}_{1 \times 1}(W_j \times F_j)] \end{cases} \quad (2)$$

式中: F_i 和 F_j 为相邻层的特征图, M_c 为相邻特征拼接后的特征, W_i 和 W_j 分别为 F_r 和 F_j 的权重分数, CA 为独立的通道注意力模块, F_o 为自适应融合后的特征, \otimes 为拼接操作。

为了更加精确地识别真实失真中不同类型的复杂失真,减少特征融合过程中的冗余信息,本文在特征融合部分加入通道注意力模型,其结构如图 4 所示。

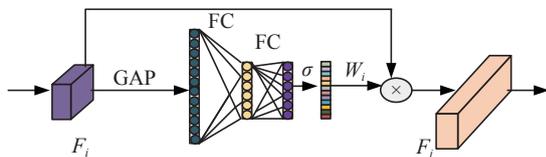


图 4 通道注意力模块
Fig. 4 Channel attention module

为了将每个通道的有利信息进行聚合,首先在特征图上进行全局平均池化处理,生成通道向量。然后利用一个全连接层把通道数缩小为原来的 1/16,再通过一个全连接层,实现跨通道之间的交互,最后经过非线性激活函数 Sigmoid 生成每个通道的权重。

2.3 语义信息增强模块

为了解决高层语义信息不足、感受野较小的问题,首次将多层扩张卷积应用于图像质量评价

任务中,使模型拥有更强的语义表达能力。语义信息增强的过程如图 5 所示。

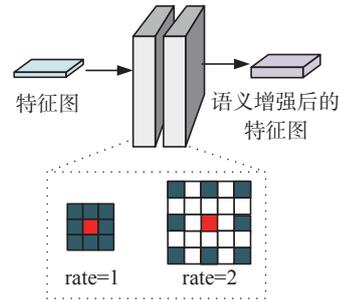


图 5 语义增强过程

Fig. 5 Process diagram of semantic enhancement

将扩张率设置为不同值的组合,可以很好地避免计算中心向外扩散的问题。本文探究了不同扩张率对野生图像质量挑战数据集 (live in the wild image quality challenge database, LIVEC) 测试集指标的影响,如表 1 所示。

表 1 扩张率对 LIVEC 测试集指标的影响
Table 1 Influence of expansion rate on LIVEC test set

扩张率组合	感受野大小	SROCC	PLCC
{1,2}	7×7	0.871	0.892
{1,2,3}	13×13	0.862	0.887
{1,2,4}	15×15	0.866	0.889

表 1 显示了不同扩张率组合对应的感受野大小及对各项指标的影响。由表 1 可以看出,当扩张率设置为 {1,2} 时,测试集上各指标均达到最高。在加入了多层扩张卷积的情况下,2 项指标均有提升,说明堆叠扩张卷积能在不缩小特征图的情况下,对模型性能的提升具有明显帮助。

2.4 语义信息指导的质量分数回归模块

人类视觉系统往往在理解图像内容的基础上对图像进行评分,因此本文引入超网络架构,利用全局语义特征作为质量分数回归模块的参数生成器,参数生成过程如图 6 所示。

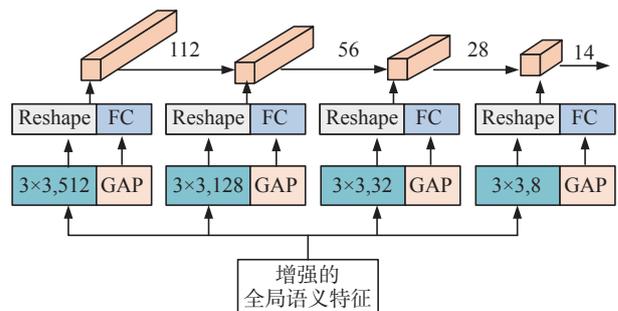


图 6 参数生成过程

Fig. 6 Parameter generation process

卷积和变形操作用于生成全连接层的权重系数 W ; 平均池化和全连接操作用于生成全连接层

的偏置系数 b 。卷积层和全连接层的输出通道数由回归网络中层的尺寸进行自适应的匹配。其中, 为了提高网络的拟合能力, 回归网络中的全连接层的节点数分别设为 112、56、28、14 和 1, 用于生成最终单幅图像的质量分数。

2.5 损失函数

为了提高主观质量分数和预测质量分数之间的等级相关性, 本文充分利用了每批图像块之间的相对距离信息, 融合了 L_1 损失函数和三元组损失函数 L_{triple} , 构建了混合损失函数 L_{mix} 。

L_1 损失函数有稳定的梯度, 不会出现梯度爆炸的问题, 常常被用于回归任务。

三元组损失函数 L_{triple} 可以对差异性较小的输入向量学习到更细微的特征。已知使用三元组损失函数的模型会让锚点 a 和正例 p 的距离值更小, 同时让锚点 a 和负例 n 的距离值更大, 即 $|a, p| \leq |a, n|$, 为了不让两边同时为 0, 增加了超参数 A , 得 $|a, p| \leq |a, n| - A$ 。

设 $s_{\text{max}}, s'_{\text{max}}, s'_{\text{min}}, s_{\text{min}}$ 分别表示 4 张图像块的主观质量得分, 按最高、次高、次低、最低分数的顺序排列, 即 $s_{\text{max}} < s'_{\text{max}} < s'_{\text{min}} < s_{\text{min}}$ 。 $p_{\text{max}}, p'_{\text{max}}, p'_{\text{min}}, p_{\text{min}}$ 为它们对应的预测得分, 如果预测得分 $p_{\text{max}} < p'_{\text{max}} < p'_{\text{min}} < p_{\text{min}}$, 则称模型预测分数和主观分数之间的相对排名信息保持一致。

为使模型预测分数和主观分数之间的相对排名信息保持一致, 可视作让 p_{max} 和 p_{min} 的距离更大, p_{max} 和 p'_{max} 的距离更小, 故可将 p_{max} 当做锚点, 将 p_{min} 当做负例, 将 p'_{max} 当做正例, 即为 $L_{\text{triple}}(p_{\text{max}}, p'_{\text{max}}, p_{\text{min}})$ 。并且, 让 p_{min} 和 p_{max} 的距离更大, p_{min} 和 p'_{min} 的距离更小, 故可将 p_{min} 当做锚点, 将 p_{max} 看当负例, 将 p'_{min} 当做正例, 即为 $L_{\text{triple}}(p_{\text{min}}, p'_{\text{min}}, p_{\text{max}})$ 。

故本文的混合损失函数为

$$L_{\text{mix}} = L_1 + \lambda [L_{\text{triple}}(p_{\text{max}}, p'_{\text{max}}, p_{\text{min}}) + L_{\text{triple}}(p_{\text{min}}, p'_{\text{min}}, p_{\text{max}})] = L_1 + \lambda [\max\{0, |p_{\text{max}}, p'_{\text{max}}| - |p_{\text{max}}, p_{\text{min}}| + A_1\} + \max\{0, |p_{\text{min}}, p'_{\text{min}}| - |p_{\text{min}}, p_{\text{max}}| + A_2\}] \quad (3)$$

式中 λ 为常量, 经实验验证, λ 取 0.2。

由于 $|p_{\text{max}}, p'_{\text{max}}| \leq |p_{\text{max}}, p_{\text{min}}| - A_1$, 即 $A_1 \leq p'_{\text{max}} - p_{\text{min}}$, 同时希望模型预测的质量分数和主观分数是一致的, 所以, 将 $s'_{\text{max}} - s_{\text{min}}$ 作为训练过程中 A_1 的上界, 设 $A_1 = s'_{\text{max}} - s_{\text{min}}$; 同理, 设 $A_2 = s_{\text{max}} - s'_{\text{min}}$ 。

3 实验结果

3.1 数据集和评估指标

为了验证本文所提模型 MFFSE-NR 的有效性, 本文在 2 个真实失真图像数据集上进行了实

验, 这 2 个数据集分别是 LIVEC 和康斯坦茨真实图像质量数据集 (konstanz authentic image quality 10k database, KonIQ-10k)。

LIVEC 数据集包含 1 162 张图像, 每张图像由各种类型的移动相机在高度多样化的条件下拍摄。KonIQ-10k 数据集包含 10 073 张图像, 每张图像从大型通用多媒体数据集 YFCC100M 中采样而来, 采样过程确保了图像内容的多样性和图像失真的真实性。

在测试阶段使用 2 个评估指标: 斯皮尔曼等级相关系数 (spearman rank-order correlation coefficient, SROCC)、皮尔逊线性相关系数 (pearson linear correlation coefficient, PLCC), 2 个指标越接近 1 说明模型性能越好, 公式为

$$S_{\text{ROCC}} = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (4)$$

$$P_{\text{LCC}} = \frac{\sum_{i=1}^N (s_i - \bar{s})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^N (s_i - \bar{s})^2 \sum_{i=1}^N (p_i - \bar{p})^2}} \quad (5)$$

式中: d_i^2 为第 i 个测试图像主观分数和预测分数的等级差异, s_i 和 p_i 分别为第 i 个图像的主观分数和预测分数, \bar{s} 和 \bar{p} 为它们的平均值, N 为测试图像数量。

3.2 参数设置

为了增强训练性能, 同时尽可能保留更多的上下文信息, 将数据集中的每幅图像都分割成若干个 224×224 的图像块。为了简单起见, 直接将图像块所在图像的主观分数赋给图像块。训练集、验证集、测试集根据失真图像按 8:1:1 划分。采用混合损失函数 L_{mix} 进行训练, 初始学习率为 10^{-6} , 使用 Adam 优化器, 权重衰减率为 0.000 5, 模型训练 40 个 epoch, 16 个图像块为一个 batch。

3.3 消融实验

为了验证多层特征融合、语义信息增强、混合损失函数 L_{mix} 的正确性, 本文选取 LIVEC 数据集进行了消融实验, 并使用 FPS 来评价各模型的计算效率, 实验结果如表 2 所示。其中, 基础模型由 Res2Net50 作为特征提取网络, 每层的失真特征直接经过平均池化、全连接层后进行聚合, 并利用最后一层的语义特征直接指导质量分数回归模块。

由表 2 可知, 本文提出的改进方法中, 在基础模型中分别加入特征融合方法、语义增强方法、混合损失函数后, LIVEC 数据集上的 SROCC 和

PLCC 均有一定的提升;同时加入 3 个模块后, SROCC 提高了 2.7%, PLCC 提高了 3.3%, 充分说明本文所提出方法能同时提高模型预测分数与人

类评分之间的准确性及等级相关性。同时,由表 2 可知,本文提出的改进方法在实现较高图像质量评价指标的同时,检测速度达到了 27 f/s。

表 2 消融实验
Table 2 Ablation experiment

序号	特征融合	语义增强	混合损失函数	斯皮尔曼等级相关系数	皮尔逊线性相关系数	每秒帧数/f/s
0	—	—	—	0.855	0.872	33
1	√	—	—	0.864	0.881	28
2	—	√	—	0.871	0.892	36
3	—	—	√	0.864	0.883	38
4	√	√	—	0.875	0.888	29
5	√	—	√	0.875	0.891	29
6	—	√	√	0.881	0.902	34
7	√	√	√	0.882	0.905	27

3.4 单一数据集上的对比实验

为了评估本文所提算法 MFFSE-NR 的有效性,将 MFFSE-NR 与 8 种性能较高的盲图像质量评价算法进行比较。其中,选取 6 种与 MFFSE-NR 相近的基于深度学习的经典算法:用于无参考的加权平均深度图像质量评估 (weighted average deep image quality assessment metric-no reference, WaDIQaM-NR)、分级退化级联卷积神经网络 (cascaded convolutional neural network with hierarchical degradation concatenation, CaHDC)、深度双线性卷积神经网络 (deep bilinear convolutional neural network, DB-CNN)、从补丁到图片的盲图像质量测量算法 (from patches to pictures blind image quality measurement, P2P-BM)、多尺度特征逐层融合的深度学习 (deep neural network based on multi-scale features fusion layer-by-layer, MsFF-Net)

和自适应超网络引导下的野外图像质量盲评估算法 (blindly assess image quality in the wild guided by a self-adaptive hyper network, HyperIQA);同时,为了保证对比范围的完整性,选取 2 种基于手工特征的算法:盲/无参考图像空间质量评估器 (blind/referenceless image spatial quality evaluator, BRISQUE) 和基于高阶统计聚合的盲图像质量评估算法 (blind image quality assessment based on high order statistics aggregation, HOSA)。实验结果如表 3 所示。

由表 3 可知, MFFSE-NR 在 KonIQ-10k 和 LIVEC 2 个真实失真图像数据集上 2 项指标均优于基于手工特征的方法,与其他 6 种基于深度学习的方法相比,同样实现了更好的性能,充分说明了在面对真实失真图像时 MFFSE-NR 能有效提高模型预测的准确性和等级相关性,有效识别各种场景的失真。

表 3 不同方法的比较
Table 3 Comparison of different methods

算法	KonIQ-10k		LIVEC	
	斯皮尔曼等级相关系数	皮尔逊线性相关系数	斯皮尔曼等级相关系数	皮尔逊线性相关系数
BRISQUE	0.705	0.681	0.608	0.629
HOSA	0.671	0.694	0.640	0.678
WaDIQaM-NR	0.797	0.805	0.671	0.680
CaHDC	—	—	0.738	0.744
DB-CNN	0.875	0.884	0.851	0.869
P2P-BM	0.872	0.885	0.844	0.842
MsFF-Net	0.909	0.922	0.866	0.883
HyperIQA	0.906	0.917	0.859	0.882
MFFSE-NR	0.910	0.924	0.882	0.905

3.5 跨数据集上的对比实验

为了验证本文所提算法 MFFSE-NR 的泛化能力, 选取 6 种先进的基于深度学习的算法进行模型泛化能力的对比实验, 实验结果如表 4 所示。

由表 4 可知, 在 2 组跨数据集实验中, 针对真实失真图像测试的 SROCC 和 PLCC 指标均达到

了最高, 说明 MFFSE-NR 在面对现实生活中的失真时, 泛化性能具有一定的先进性。针对不同失真程度的特征进行了融合, 对语义信息进行了增强, 并利用了带有排名信息的损失函数, 在一定程度上能提升模型预测分数和主观质量分数之间的等级相关性和准确性, 提高模型的泛化能力。

表 4 跨数据集测试的 SROCC
Table 4 SROCC for cross dataset testing

训练集	测试集	斯皮尔曼等级相关系数						
		WaDIQaM-NR	CaHDC	DB-CNN	P2P-BM	MsFF-Net	HyperIQA	MFFSE-NR
KonIQ-10k	LIVEC	0.682	0.624	0.755	0.770	0.787	0.785	0.794
LIVEC	KonIQ-10k	0.711	—	0.754	0.740	0.760	0.772	0.775

3.6 预测结果散点图

本文对 KonIQ-10k 数据集和 LIVEC 数据集上的测试结果进行了可视化展示, 如图 7 所示。

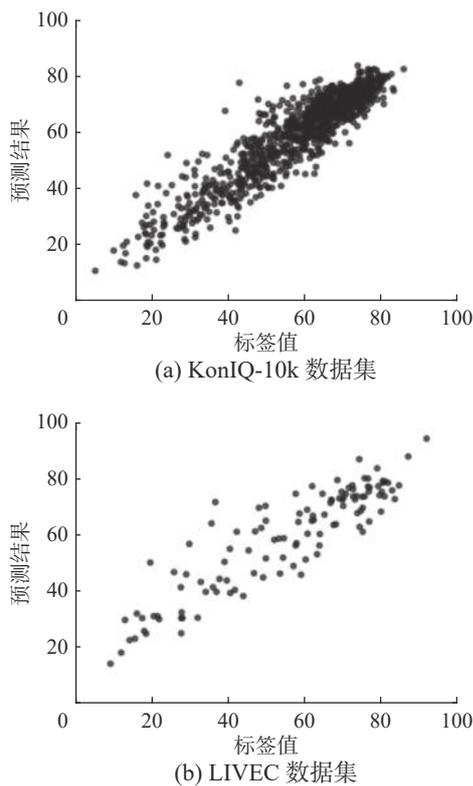


图 7 KonIQ-10k 和 LIVEC 数据集上测试结果的可视化展示

Fig. 7 Visualization of test results on the KonIQ-10k and LIVEC datasets

从可视化结果可看出, MFFSE-NR 的预测分数与人类评分标签之间呈现出高度的一致性, 说明本算法能对各类真实失真进行有效的识别, 在一定程度上能模拟人类视觉系统对真实失真图像进行评分的过程, 进而提升了真实失真图像质量评价的准确度。

4 结束语

为解决现有的无参考图像质量评价算法在面临真实失真图像时存在的各种问题, 本文提出了多层特征融合和语义增强的盲图像质量评价算法, 通过 Res2Net50 网络构建特征金字塔, 提出了细粒度的基于注意力的多层特征融合模块, 以应对真实失真图像中的局部和非均匀失真; 探究了不同扩张率叠加的扩张卷积对语义信息的增强效果, 提出了语义信息增强模块, 并利用增强的高级语义特征生成质量分数回归模块的参数; 最后考虑图像块质量分数之间的相对排名关系, 引入了三元组损失, 构建了混合损失函数 L_{mix} , 提高了模型对真实失真特征的学习能力。

但是, 本方法将图像分割成图像块, 并将图像的质量分数赋给图像块, 在一定程度上引入了标签噪声, 如何利用整幅图像进行质量分数的预测是本文未来的研究内容之一。此外, 随着各种大规模数据集的出现, 如何在复杂场景下对各种真实失真进行评价也是本文后续工作的研究重点。

参考文献:

- [1] 方玉明, 眭相杰, 鄢杰斌, 等. 无参考图像质量评价研究进展 [J]. 中国图象图形学报, 2021, 26(2): 265-286.
FANG Yuming, SUI Xiangjie, YAN Jiebin, et al. Progress in no-reference image quality assessment[J]. Journal of image and graphics, 2021, 26(2): 265-286.
- [2] 曹玉东, 刘海燕, 贾旭, 等. 基于深度学习的图像质量评价方法综述 [J]. 计算机工程与应用, 2021, 57(23): 27-36.
CAO Yudong, LIU Haiyan, JIA Xu, et al. Overview of image quality assessment method based on deep

- learning[J]. *Computer engineering and applications*, 2021, 57(23): 27–36.
- [3] 毕晓君, 潘梦迪. 基于生成对抗网络的机载遥感图像超分辨率重建 [J]. *智能系统学报*, 2020, 15(1): 74–83.
BI Xiaojun, PAN Mengdi. Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks[J]. *CAAI transactions on intelligent systems*, 2020, 15(1): 74–83.
- [4] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE transactions on image processing: a publication of the IEEE signal processing society*, 2004, 13(4): 600–612.
- [5] GAO Xinbo, LU Wen, TAO Dacheng, et al. Image quality assessment based on multiscale geometric analysis[J]. *IEEE transactions on image processing*, 2009, 18(7): 1409–1423.
- [6] 王志明. 无参考图像质量评价综述 [J]. *自动化学报*, 2015, 41(6): 1062–1079.
WANG Zhiming. Review of no-reference image quality assessment[J]. *Acta automatica sinica*, 2015, 41(6): 1062–1079.
- [7] KANG Le, YE Peng, LI Yi, et al. Convolutional neural networks for no-reference image quality assessment[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1733–1740.
- [8] XU Jiangtao, YE Peng, LI Qiaohong, et al. Blind image quality assessment based on high order statistics aggregation[J]. *IEEE transactions on image processing: a publication of the IEEE signal processing society*, 2016, 25(9): 4444–4457.
- [9] 严浙平, 曲思瑜, 邢文. 水下图像增强方法研究综述 [J]. *智能系统学报*, 2022, 17(5): 860–873.
YAN Zheping, QU Siyu, XING Wen. An overview of underwater image enhancement methods[J]. *CAAI transactions on intelligent systems*, 2022, 17(5): 860–873.
- [10] 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.
- [11] WU Jinjian, MA Jupuo, LIANG Fuhu, et al. End-to-end blind image quality prediction with cascaded deep neural network[J]. *IEEE transactions on image processing*, 2020, 29: 7414–7426.
- [12] ZHANG Weixia, MA Kede, YAN Jia, et al. Blind image quality assessment using a deep bilinear convolutional neural network[J]. *IEEE transactions on circuits and systems for video technology*, 2020, 30(1): 36–47.
- [13] 鄢杰斌, 方玉明, 刘学林. 图像质量评价研究综述: 从失真的角度 [J]. *中国图象图形学报*, 2022, 27(5): 1430–1466.
YAN Jiebin, FANG Yuming, LIU Xuelin. The review of distortion-related image quality assessment[J]. *Journal of image and graphics*, 2022, 27(5): 1430–1466.
- [14] SU Shaolin, YAN Qingsen, ZHU Yu, 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. Piscataway: IEEE, 2020: 3664–3673.
- [15] 杨春玲, 杨雅静. 基于多尺度特征逐层融合深度神经网络的无参考图像质量评价方法 [J]. *华南理工大学学报 (自然科学版)*, 2022, 50(4): 81–89, 141.
YANG Chunling, YANG Yajing. A deep neural network based on layer-by-layer fusion of multi-scale features for no-reference image quality assessment[J]. *Journal of South China University of Technology (natural science edition)*, 2022, 50(4): 81–89, 141.
- [16] YING Zhenqiang, NIU Haoran, GUPTA P, et al. From patches to pictures (PaQ-2-PiQ): mapping the perceptual space of picture quality[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 3572–3582.
- [17] 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.
- [18] HOSU V, LIN Hanhe, 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.
- [19] 刘玉珍, 刘美怡, 林森等. 多尺度特征融合注意力网络的水下图像增强 [J]. *计算机辅助设计与图形学学报*, 2023, 35(5): 685–695.
LIU Yuzhen, LIU Meiyi, LIN Sen, et al. Underwater image enhancement based on multi-scale feature fusion and attention network[J]. *Journal of computer-aided design &*

- computer graphics, 2023, 35(5): 685–695.
- [20] GOLESTANEH S A, DADSETAN S, KITANI K M. No-reference image quality assessment via transformers, relative ranking, and self-consistency[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2022: 3989–3999.
- [21] GAO Shanghua, CHENG Mingming, ZHAO Kai, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(2): 652–662.
- [22] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition . Piscataway: IEEE, 2016: 770–778.
- [23] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition . Piscataway: IEEE, 2015: 815–823.
- [24] LI Dingquan, JIANG Tingting, JIANG Ming. Exploiting high-level semantics for no-reference image quality assessment of realistic blur images[C]//Proceedings of the 25th ACM international conference on Multimedia. New York: ACM, 2017: 378–386.
- [25] 高涛, 杨朝晨, 陈婷, 等. 深度多尺度融合注意力残差人

脸表情识别网络 [J]. *智能系统学报*, 2022, 17(2): 393–401.

GAO Tao, YANG Zhaochen, CHEN Ting, et al. Deep multiscale fusion attention residual network for facial expression recognition[J]. *CAAI transactions on intelligent systems*, 2022, 17(2): 393–401.

作者简介:



赵文清, 教授, 主要研究方向为人工智能与图像处理。主持和参与国家自然科学基金项目、河北省自然科学基金项目、中央高校基本科研业务费专项资金资助项目和国家电网科技项目等 10 余项。获河北省科技进步二等奖、三等奖各 1 项。发表学术论文 80 余篇, 出版专著 1 部。E-mai: jbzwq@126.com。



许丽娇, 硕士研究生, 主要研究方向为图像质量评价。E-mai: xulijiao2021@163.com。



陈昊阳, 硕士研究生, 主要研究方向为图像质量评价。E-mai: 870715478@qq.com。