

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2024)12-3699-13

论文引用格式: Yan J B, Tan Z W, Wu K C, Liu X L and Fang Y M. 2024. Viewport-independent and deformation-unaware no-reference omnidirectional image quality assessment. Journal of Image and Graphics, 29(12):3699-3711(鄢杰斌, 谭淄文, 吴康诚, 刘学林, 方玉明. 2024. 非视口依赖的抗畸变无参考全景图像质量评价. 中国图象图形学报, 29(12):3699-3711)[DOI:10.11834/jig.240188]

非视口依赖的抗畸变无参考全景图像质量评价

鄢杰斌, 谭淄文, 吴康诚, 刘学林, 方玉明*

江西财经大学计算机与人工智能学院, 南昌 330032

摘要: 目的 全景图像质量评价(omnidirectional image quality assessment, OIQA)旨在定量描述全景图像降质情况,对于算法提升和系统优化起着重要的作用。早期的OIQA方法设计思想主要是结合全景图像的几何特性(如两级畸变和语义分布不均匀)和2D-IQA方法,这类方法并未考虑用户的观看行为,因而性能一般;现有的OIQA方法主要通过模拟用户的观看行为,提取观看视口序列;进一步,计算视口序列失真情况,然后融合视口失真得到全景图像的全局质量。然而,观看视口序列预测较为困难,且预测模型的实时性和鲁棒性难以保证。为了解决上述问题,提出一种非视口依赖的抗畸变无参考(no reference, NR)OIQA(NR-OIQA)模型。针对全景图像等距柱状投影(equirectangular projection, ERP)所带来的规律性几何畸变问题,提出一种可同时处理不规则语义和规律性畸变的新型卷积方法,称为等矩形可变形卷积方法,并基于该卷积方法构建NR-OIQA模型。**方法** 该模型主要由先验指导的图像块采样(prior-guided patch sampling, PPS)模块、抗畸变特征提取(deformation-unaware feature extraction, DUFE)模块和块内一块间注意力聚集(intra-inter patch attention aggregation, A-EPAA)模块3个部件组成。其中,PPS模块根据先验概率分布从高分辨率的全景图像采样提取相同分辨率的图像块;DUFE模块通过等矩形可变形卷积渐进式地提取输入图像块质量相关特征;A-EPAA模块旨在调整单个图像块内部特征以及各图像块对整体质量评价的影响程度,以提升模型对全景图像质量的评价准确度。**结果** 在3个公开数据集上将本文模型与其他IQA和OIQA模型进行性能比较,与性能第1的Assessor360相比,参数量减少了93.7%,计算量减少了95.4%;与模型规模近似的MC360IQA相比,在CVIQ、OIQA和JUFU数据集上的斯皮尔曼相关系数分别提升了1.9%、1.7%和4.3%。**结论** 本文所提出的NR-OIQA模型,充分考虑了全景图像的特点,能够以不依赖视口的方式高效提取具有失真特性的质量特征,对全景图像进行准确质量评价,并具有计算量低的优点。

关键词: 图像质量评价(IQA);全景图像;可变形卷积;注意力机制;无参考;视口

Viewport-independent and deformation-unaware no-reference omnidirectional image quality assessment

Yan Jiebin, Tan Ziwen, Wu Kangcheng, Liu Xuelin, Fang Yuming*

School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang 330032, China

Abstract: Objective With the rapid development of the virtual reality (VR) industry, the omnidirectional image acts as an important medium of visual representation of VR and may degrade in the procedure of acquisition, transmission, pro-

收稿日期:2024-04-01;修回日期:2024-05-22;预印本日期:2024-05-29

*通信作者:方玉明 leo.fangyuming@foxmail.com

基金项目:国家重点研发计划资助(2023YFE0210700);江西省自然科学基金项目(20232BAB202001, 20243BCE51139)

Supported by: National Key R&D Program of China (2023YFE0210700); Natural Science Foundation of Jiangxi Province, China (20232BAB202001, 20243BCE51139)

cessing, and storage. Omnidirectional image quality assessment (OIQA) is an evaluation technique that aims to quantitatively describe the degradation of omnidirectional images and plays a crucial role in algorithm improvement and system optimization. Generally, the omnidirectional image has some inherent characteristics, i. e., geometric deformation in the polar region and semantic information more concentrated on the equatorial region. The viewing behavior can conspicuously affect the perceptual quality of an omnidirectional image. Early OIQA methods that simply fuse this inherent characteristic in 2D-IQA do not consider the significant user viewing behavior, thus obtaining suboptimal performance. Considering the viewport representation that is in line with the user viewing behavior, some deep learning-based OIQA methods have recently achieved promising performance by taking the predicted viewport sequence as the model input and computing the degradation. However, the prediction of the viewport sequence is difficult and viewport extraction needs a series of pixel-wise computations, thus leading to a significant computation load and hampering the application in the industry environment. To address the above problems, we proposed a new no-reference OIQA model, which introduces an equirectangular modulated deformable convolution (EquiMdconv) that can deal with the irregular semantics and the regular deformation caused by equirectangular projection simultaneously without the predicted viewport sequence. **Method** We propose a viewport-independent and deformation-unaware no-reference OIQA model for omnidirectional image quality assessment. Our model is composed of three parts: a prior-guided patch sampling (PPS) module, a deformable-unaware feature extraction (DUFE) module, and an intra-interpatch attention aggregation (A-EPAA) module. The PPS module samples a set of patch images on the basis of prior probability distribution in a slice-based manner to represent the complete image quality information. DUFE aims to extract the perceptual quality features of the input patch images, considering the irregular semantics and regular deformation in this process. It contains eight blocks, and each block comprises an EquiMdconv layer, a 1×1 convolutional layer, a batch normalization layer, and a 3×3 max pooling layer. The EquiMdconv layer employs a modulated deformable convolution layer that introduces learnable offset parameters to model distortions in the images more accurately. Furthermore, we incorporate fixed offsets based on distortion regularity factors into the deformable convolution's offset to effectively eliminate the regular deformation. The A-EPAA comprises a convolutional block attention module (CBAM) and a patch attention module (PA). The CBAM assigns weights to each channel to adjust perceptual quality features in both channel and spatial dimensions. The PA adjusts the contribution weights between patch images for an overall quality assessment. We train the proposed model on the CVIQ, OIQA, and JUFEE databases. In the training stage, we split each database into two parts: 80% for training and 20% for testing. We sample 10 patch images from each omnidirectional image, and the size of the patch image is set to 224×224 . All experiments are implemented on a server with an NVIDIA GTX A5000 GPU. Adaptive moment estimation optimizer (Adam) is utilized to optimize our model. We train the model for 300 epochs on the CVIQ and OIQA databases and 20 epochs on the JUFEE database; the learning rate is 0.0001 and the batch size is 16. **Result** We conduct experiments covering three databases, namely, CVIQ, OIQA, and JUFEE. We demonstrate the performance of the proposed model by comparing it with nine viewport-independent models and five viewport-dependent models. To ensure a persuasive comparison result, we select the Pearson linear correlation coefficient and Spearman's rank correlation coefficient (SRCC) as performance evaluation standards. The results indicate that compared with those of the state-of-the-art viewport-dependent model, i. e., Assessor360, the parameters of our model are reduced by 93.7% and the floating point operations are reduced by 95.4%. Compared with the MC360IQA, which has a similar model size, the SRCC is increased by 1.9%, 1.7%, and 4.3% on the CVIQ, OIQA, and JUFEE databases, respectively. **Conclusion** Our proposed viewport-independent and deformation-unaware no-reference OIQA model thoroughly considers the characteristics of the omnidirectional image. It can effectively extract quality features and accurately assess the quality of omnidirectional images with limited computational cost.

Key words: image quality assessment (IQA); omnidirectional image; deformable convolution; attention mechanism; no reference; viewport

0 引言

近年来虚拟现实技术得到迅速发展,全景图像作为一种重要的视觉载体,为用户提供沉浸式的观看体验。全景图像通常以球体的方式被用户通过头戴式显示设备观看,用户可以沉浸式地观看 $360^\circ \times 180^\circ$ 的图像内容,以提供更真实、更全面的视觉感知。然而,由于全景图像高分辨率的特性,在其生成、处理、传输和存储等一系列过程中图像质量都可能降低,极大影响终端用户的观看体验(Wang和Rehman, 2019; 鄢杰斌等, 2022)。如何有效评价全景图像的视觉质量,对于保障用户在观看过程中的高质量体验以及提高图像处理效率至关重要。

全景图像质量评价(omnidirectional image quality assessment, OIQA)方法可通过主观和客观方式对全景图像视觉质量进行定量描述。由于无参考(no reference, NR)OIQA(NR-OIQA)方法使用时仅需输入失真图像而无需参考图像信息,因此对该类方法的研究较多且更具实际价值。目前, OIQA方法可根据输入形式分为非视口依赖的方法和视口依赖的方法。非视口依赖方法以等距柱状投影(equirectangular projection, ERP)图像或经裁剪的图像块作为输入,在数据预处理阶段具有较少的计算负担。早期OIQA方法通常基于峰值信噪比(peak signal-to-noise ratio, PSNR)的2D-IQA方法进行改进,主要解决ERP图像两极区域出现的规律性扭曲失真问题(Yu等, 2015; Zakharchenko等, 2016; Sun等, 2017)。由于PSNR的方法没有考虑人眼视觉系统(human visual system, HVS)的一般特性,学者们开始研究基于SSIM的方法。Chen等人(2018)通过度量图像中亮度、对比度和结构信息来表征区域质量。Zhou等人(2018)通过基于位置权重图映射的结构相似性(structural similarity index, SSIM)解决几何畸变。以ERP图像作为输入时,难以对不同的区域使用自适应处理方式,基于图像块输入的评价方法被提出(Lim等, 2018, 2020),并取得较好的效果。但是,非视口依赖的OIQA方法往往无法完全处理两极畸变问题,并受到球面失真、均匀加权以及对视觉感知不足等问题的限制,因此难以满足现实需求。

由于视口图像与用户观看行为存在高度一致性(Li等, 2019),学者们开始研究视口依赖的OIQA方

法,以模仿用户实际观看过程并同时解决ERP图像中的两极畸变。Sun等人(2020)将全景图像投影为6个视口图像,并通过多通道卷积神经网络(multi-channel convolution neural network)进行质量预测。图像质量评价被视为一个局部细节到整体感知、纹理信息到语义信息相互影响的过程(李雷达等, 2022; Zhang和Liu, 2022)。Xu等人(2021)提出视口导向图卷积网络(viewport oriented graph convolution network, VGCN),该方法可以同时获取视口的局部质量分数和ERP图像的整体分数并将其聚合。Zhang和Liu(2022)考虑到用户观看行为,通过定义多条扫视路径来提取视口序列,并引入门控循环单元(gate recurrent unit, GRU)来解决序列中存在的延时效应(Hands和Avons, 2001)。Wu等人(2024)考虑用户观看行为的不确定性,通过结合图像熵信息和先验赤道偏置生成多条视口序列,计算每条视口序列的分数后融合得到最终质量。尽管视口依赖的OIQA方法往往能取得令人满意的性能,但由于需要进行额外的视口序列预测和提取,其设计复杂且计算量较高,阻碍了在工业的实用性。

近年来,以深度卷积神经网络(deep convolutional neural network, DNN)为代表的深度学习技术在图像视觉任务处理中发挥了关键作用(左一帆等, 2023; 冷佳旭等, 2023; 刘华峰等, 2023),并且与注意力机制结合可以进一步处理重要视觉信息(何伟和潘晨, 2022; 方玉明等, 2022),这些发展极大促进了OIQA的研究进程。尽管基于DNN的OIQA方法展示出较好的评估效果,但其固定的卷积采样策略应用在不规则语义上时容易引入冗余信息,如图1(a)所示,其中灰色区域代表有效区域。Dai等人(2017)通过在标准卷积中引入可学习的偏置偏移,提出可变形卷积(deformable convolutional network, DCN)以对有效区域动态采样,如图1(b)所示。为了处理全景图像中的两极畸变, Fernandez-Labrador等人(2020)提出等矩形卷积(equirectangular convolution, EquiConv),通过ERP图像的几何先验来改变卷积核的形状以适应像素拉伸。但是,该方法无法处理不规则语义,如图1(c)所示。如何同时解决OIQA中的不规则语义和两极畸变问题,避免在特征提取中引入冗余信息,提高特征建模效率,对OIQA至关重要。

本文提出一个非视口依赖的、能够同时解决两

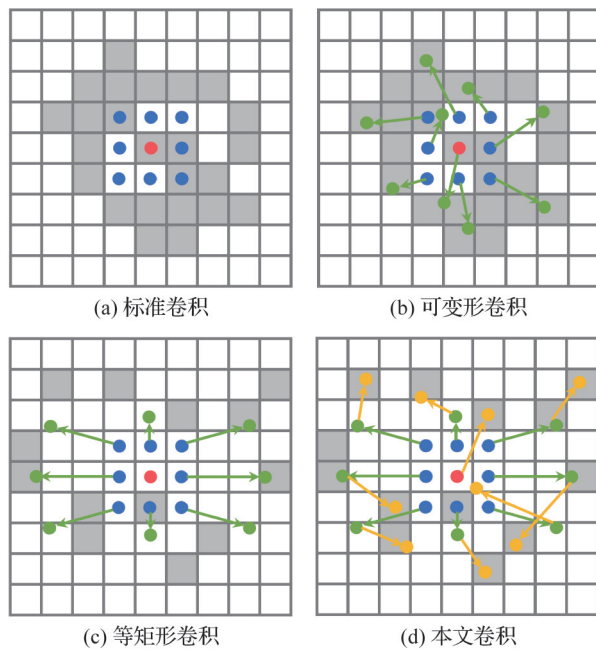


图1 各类型卷积的采样策略

Fig. 1 Sampling strategies of each type of convolution

((a) standard convolution; (b) deformable convolution; (c) equirectangular convolution; (d) proposed convolution)

极畸变问题和不规则语义问题的无参考 OIQA 方法,并在保证性能的同时进行模型轻量化。具体而言,为了避免视口序列预测与提取,本文提出基于先验指导的图像块采样 (prior-guided patch sampling, PPS) 模块。如图 2 所示,PPS 根据先验分布划分非均匀网格,使采样图像块聚集在赤道附近,和全景图像中的语义分布与用户观看行为 (Deng 等, 2021; 周玉 等, 2022) 保持一致。

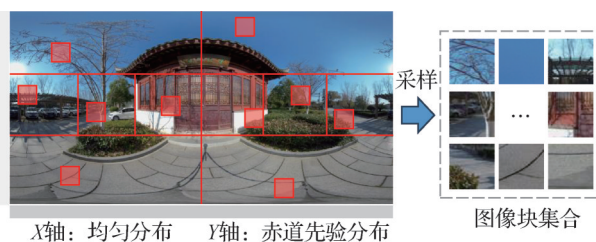


图2 先验指导的图像块采样

Fig. 2 Prior-guided patch sampling

为了解决图像块中的两级畸变和不规则语义,本文提出等矩形可变形卷积 (equirectangular modulated deformable convolution, EquiMdconv)。如图 1 (d) 所示,该卷积通过对采样位置的连续两次偏移,实现在解决两极畸变的基础上对不规则语义特征进行准确覆盖。进一步,用 EquiMdconv 构建抗畸变特

征提取 (deformable-unaware feature extraction, DUFE) 模块。本文结合注意力机制提出块内一块间注意力聚集 (intra-inter patch attention aggregation, A-EPAA) 模块,以调整图像块内部像素权重分布和块间对整体质量评价的影响权重。

本文的主要贡献可归纳为: 1) 设计了一种基于先验偏置的图像块采样方法 (PPS), 获得的图像块集合可有效表征全景图像的关键内容。2) 提出等矩形可变形卷积 (EquiMdconv), 可同时处理图像中的不规则物体形状和规律性两级畸变; 设计基于 EquiMdconv 的抗畸变特征提取 (DUFE) 模块作为质量特征提取的骨干网络。3) 提出块内一块间注意力聚集 (A-EPAA) 模块用于调整图像块内部和图像块之间对质量评价的影响分布。4) 由 PPS、DUFE 和 A-EPAA 模块组成本文模型, 该模型不需要依赖视口, 并且在 3 个公开 OIQA 数据集上取得显著的优势, 同时具备轻量化优点。

1 提出的方法

1.1 模型概述

本文模型的网络架构如图 3 所示, 主要包含 3 个部分, 即赤道先验指导的图像块采样 (PPS) 模块、抗畸变特征提取 (DUFE) 模块和块内一块间注意力聚集 (A-EPAA) 模块。具体而言, 首先通过 PPS 模块对输入的 ERP 图像提取表征图像关键区域的图像块集合, 然后通过 DUFE 模块对图像块使用具有双重采样位置偏移的可变形卷积层进行感知质量特征提取, 以有效处理全景图像中特有的两级畸变和不规则语义。然后使用 A-EPAA 模块逐步动态调整图像块内部及图像块之间的感知质量特征注意力权重, 以聚合感知质量相关的重要特征。最后通过 3 个全连接层 (full connection layer, FC) 将每个图像块对应的质量特征回归成局部质量分数, 将所有局部质量分数求均值得到最终预测质量分数。该网络通过块采样、语义特征提取和注意力机制的协同作用能够实现直接对 ERP 图像进行质量预测, 而无需依赖视口, 为 OIQA 提供了一种有效而高泛用性的评价架构。

1.2 图像块采样模块

为了避免视口序列预测和提取带来的额外复杂操作, 本文采用图像块的形式作为模型输入, 以降低

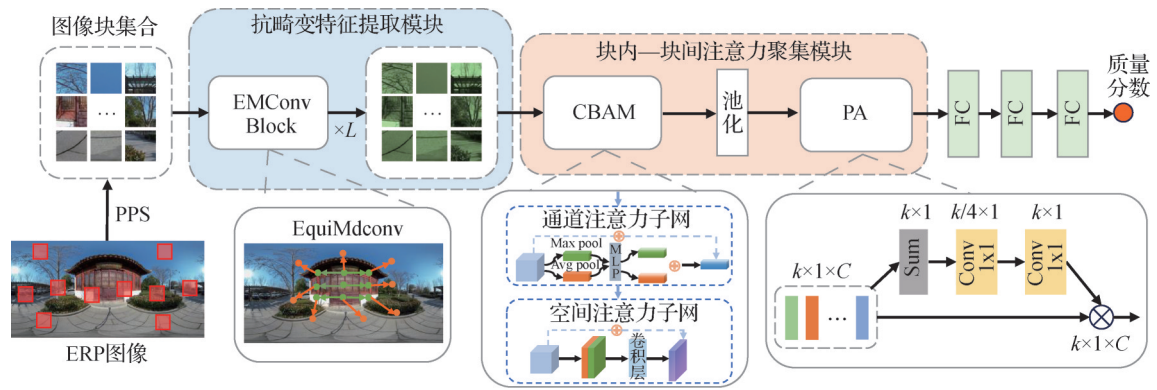


图3 本文模型的网络架构

Fig. 3 The architecture of the proposed model

在图像预处理阶段的计算复杂度。考虑到全景图像的语义信息集中在低纬区域,以及用户观看在纬度方向存在赤道偏置 (prior equator bias, PEB) (Djilali 等, 2021), 在经度方向上符合均匀分布。因此, 本文使用两种先验概率对 ERP 图像进行采样。一般地, 给定观测坐标 (x, y) , 则该点在纬度和经度上分别服从赤道先验分布以及均匀分布, 即 $X \sim PEB(\mu, \lambda)$ 和 $Y \sim U(-180^\circ, 180^\circ)$, 其中 μ 和 λ 分别为位置参数和尺度参数, 通常为 $\mu = 91.3^\circ$ 和 $\lambda = 18.58^\circ$; 设置纬度划分因子 θ_r , 以及采样图像块数量 K , 则在低纬度区域以及高纬度区域可分别划分以下数量的网格块

$$K_{\text{low}} = K \times \text{Floor} \left(\int_{-\theta_r}^{\theta_r} f_{\text{PEB}}(x) dx \right) \quad (1)$$

$$K_{\text{high}} = K - K_{\text{low}} \quad (2)$$

式中, Floor 表示向下取整, f_{PEB} 表示服从 PEB 的概率分布函数, K_{low} 和 K_{high} 分别表示低纬区域的网格数量和高纬区域的网格数量。由于全景图像沿纬度方向具有对称性, 因此在南北两级区域存在 $K_{\text{high}}/2$ 个网格块。随后, 在每个网格块内部进行 x 和 y 方向上的随机采样, 进一步得到图像块坐标集合 $\{(x_1, y_1), \dots, (x_k, y_k)\}$ 。获得坐标后, 根据每个坐标进行切片采样以获得对应的图像块集合 $\{I_1, \dots, I_k\}$ 。上述操作旨在确保对于富含语义信息的区域有更充分的覆盖, 以促进后续感知质量特征的有效提取, 并且没有引入额外的参数和复杂的计算过程。

1.3 抗畸变特征提取模块

传统 DNN 中的标准卷积在处理不规则语义特征时容易引入冗余信息, 而 DCN 可通过位置偏移 (offset) 动态地对不规则语义进行采样, 但无法直接应用于全景图像。等矩形卷积则可有效解决 ERP

图像中固有的两极畸变, 但它忽视了不规则语义的影响。本文通过结合可变形卷积和等矩形卷积的优点, 提出可同时处理不规则语义和两极畸变的等矩形可变形卷积 (EquiMdconv), 其工作流程如图 5 所示。具体而言, 使用改进的 DCN (Zhu 等, 2019) 作为 EquiMdconv 的主体, DCN 分别通过卷积层生成可学习的 offset 和掩码 (mask), 其中 offset 存储着每个像素点 x 和 y 方向上的偏移量, 用于偏移卷积核采样位置以便卷积区域始终覆盖在关键区域, mask 用于控制每个像素值的权重以进一步提高卷积的重要信息提取能力。受 Fernandez-Labrador 等人 (2020) 工作启发, ERP 图像中存在的两极畸变具有规律性, 即沿赤道向两极区域畸变程度逐渐加剧。因此, 通过在可变形卷积可学习的 offset 基础上引入用于修正两极畸变的固定 offset, 以此解决 ERP 图像中规律性失真的问题。如图 4 所示, 本文中提出的 EquiMdconv 所使用的偏移量由一个固定的偏移量和一个可学习的偏移量构成

$$\Delta p = \Delta p_f + \Delta p_l \quad (3)$$

式中, Δp_f 表示固定的偏移量, 由计算得到的 ERP 偏移在卷积对应位置采样得到, Δp_l 表示可学习的偏移量。最后, EquiMdconv 操作可表示为

$$f_{\text{out}}(p_0) = \sum_{p_n} w(p_n) f_{\text{in}}(p_0 + \Delta p) \quad (4)$$

式中, f_{in} 和 f_{out} 分别表示输入特征和输出特征, $w(p_n)$ 表示权重矩阵, p_0 表示特征图上的任意一点坐标, p_n 表示以 p_0 为中心的邻域, Δp 表示卷积核中每一个点相对于中心点的偏移量。

进一步地, 通过简单堆叠等矩形可形变卷积块 (EMConv block), 构建具有较低计算复杂度的抗畸

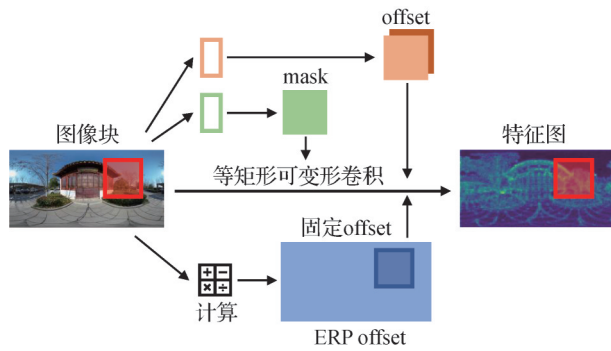


图4 EquiMdconv的工作流程

Fig. 4 The pipeline of EquiMdconv

变特征提取模块(DUFE)作为主干网络,用于提取图像中的感知质量特征。具体来说,为了模型的轻便性,DUFE仅由8个矩形可形变卷积块构成,EquiMdconv块由必要的卷积层和激活函数组成,其中包括一个 1×1 普通卷积、一个EquiMdconv、一个batch normalization(BN)层、一个ReLU层以及一个最大池化层。 1×1 卷积用于处理输入特征图的空间维度以及输入特征图的大小,由于此操作不涉及空间感受野表征,为减少计算量,仅使用普通卷积。BN层能降低网络过拟合的风险,同时保持网络层之间的稳定,防止出现梯度爆炸或者梯度消失的问题。ReLU用于让神经网络能够学习非线性关系,提高模型复杂特征表达能力。最大池化层用于调整特征图的空间维度,保留最显著的特征。每个块内部使用残差网络进行连接,每个层第1个块 1×1 卷积核会增加2倍通道数并且减少特征图大小,最大池化层也会改变特征图大小,最终 1×1 卷积层输出特征图会和池化的特征图进行相加,作为下一个块的输入,每层的第2个块则不对通道数和特征图大小作任何改变。

1.4 块内一块间注意力聚集模块

块内一块间注意力聚集模块(A-EPAA)主要由卷积注意力模块(convolutional block attention module, CBAM)(Woo等,2018)和块间注意力模块(patch attention, PA)构成。CBAM包含通道注意力子网和空间注意力子网,用于在通道纬度和空间纬度对单个图像块的质量特征信息进行注意力加权,以突出显著感知质量特征。PA用于调整ERP图像中不同图像块之间的注意力权重,强调显著图像块特征。具体而言,输入特征 f_n 经过通道注意力子网时,采用最大池化(max pooling)和平均池化(average

pooling)对输入特征在通道维度上进行聚合操作,得到密集特征表示 f_m 和 f_a 。然后将它们送入一个共享的多层感知机(multilayer perceptron, MLP)网络。经过该网络处理后对两个输出进行逐元素相加,然后通过sigmoid函数映射为通道注意力权重向量,初始特征与之相乘获得通道注意力加权后的特征图。主要过程可表示为

$$f_w = \text{MLP}(f_m) \oplus \text{MLP}(f_a) \quad (5)$$

$$f_c = \text{softmax}(f_w) \odot f_n \quad (6)$$

式中, f_w 表示权重向量, \oplus 表示特征逐元素相加, \odot 表示特征逐元素相乘, f_c 表示注意力加权后的特征图。随后, f_c 进入空间注意力子网沿通道方向进行平均池化和最大池化,生成两个2D的特征描述符号 S_a 和 S_m ,将两个特征描述符进行拼接后,使用一个 7×7 的卷积核对其进行卷积融合。之后将融合特征使用sigmoid函数映射得到空间注意力权重向量,与 f_c 相乘后得到空间注意力加权后的特征图。

将经过CBAM调整后的特征图进行全局平均池化(global average pooling, GAP),得到单个图像块的特征向量,作为PA模块输入。该模块不会关注图像块内部的质量特征分布,而是强调单个图像块在ERP图像中的感知质量显著性。具体而言,通过GAP层后会获得10个图像块的特征向量,然后对所有特征向量进行求和得到融合特征 f_c ,该特征包含整幅ERP图像的感知质量信息。通过两个全连接层将 f_c 映射为块间注意力权重向量,与初始特征向量相乘获得经过块间注意力加权后的输出向量。

输出特征向量经过3个全连接层后得到预测质量分数。最终,输入图像的质量分数由该图像对应的图像块分数取均值得到,即

$$\tilde{s}_n = \frac{1}{K} \sum_{k=1}^K s_n^k \quad (7)$$

1.5 损失函数

本文使用norm-in-norm loss(Li等人,2020)作为损失函数去优化所提出的模型,该方法能够使训练过程更快收敛的同时提高模型对参数的鲁棒性。给定一组预测质量分数 $\tilde{s} = (\tilde{s}_1, \dots, \tilde{s}_N)$ 以及一组主观质量分数 $\hat{s} = (\hat{s}_1, \dots, \hat{s}_N)$,则该损失函数定义为

$$\text{Loss}(\tilde{s}, \hat{s}) = \frac{1}{\varepsilon N} \|\tilde{q}_n - \hat{q}_n\|^\gamma \quad (8)$$

式中, ε 表示正则化因子, γ 表示超参数, \tilde{q}_n 和 \hat{q}_n 表示

正则化之后的质量分数,即

$$\tilde{q}_n = \frac{\tilde{s}_n - \tilde{\mu}}{\tilde{\sigma}} \quad (9)$$

$$\hat{q}_n = \frac{\hat{s}_n - \hat{\mu}}{\hat{\sigma}} \quad (10)$$

式中, \tilde{s}_n 和 \hat{s}_n 分别表示预测质量分数和主观质量分数, $\tilde{\mu}$ 和 $\hat{\mu}$ 分别表示预测质量分数和主观质量分数的均值, $\tilde{\sigma}$ 和 $\hat{\sigma}$ 分别表示预测质量分数和主观质量分数的标准差。

2 实验

2.1 数据集

本文实验在以下3个数据集上展开。

1) CVIQ 数据库(Sun等, 2018)。共包含16幅参考全景图像和对应的528幅压缩图像,这些图像由3种常见的编码技术压缩而成,包括JPEG压缩、H. 264/AVC和H. 265/HEVC,每个压缩类型包含11种压缩程度。

2) OIQA 数据库(Duan等, 2018)。共包含16幅参考全景图像和对应的320幅失真图像,这些失真图像由JPEG压缩、JPEG2000压缩、高斯模糊和高斯噪声4种常见编码技术压缩而成。

3) JUF E 数据库(Fang等, 2022)。共有258幅包含丰富场景类型的参考全景图像以及对应的1032幅失真图像,通过对鱼眼镜头进行单独处理形成非均匀失真,其中含有4种非均匀失真:高斯模糊、高斯噪声、亮度不连续和拼接失真。

2.2 实验设置

本文实验在Ubuntu22.0.4操作系统上进行,其配有NVIDIA RTX A5000(24 GB内存), Intel(R) Xeon(R) Gold 6326 CPU @ 2.90 GHz(260 GB RAM),提出的模型在PyTorch 2.1.0环境下搭建。

在图像块采样阶段,纬度划分因子 θ_r 设置为 23° ,提取的图像块数量为10个,每个图像块的尺寸为 224×224 像素。采用自适应矩估计(adaptive moment estimation, Adam)优化算法对模型进行训练,初始学习率为 $1E-4$,学习率随迭代呈cosine形式逐渐下降为 $1E-6$,权重衰减(weight decay)设置为 $5E-4$ 。在训练阶段,样本批量大小(batch size)设置为16,在CVIQ、OIQA数据集上的训练周期为300个epoch,在JUF E数据集上的训练周期为20个epoch。

根据源图像将数据集分为训练集和测试集,训练集划分为总图像数量的80%,测试集占比20%。

2.3 评价指标

本文采用3种常用的评价指标,即皮尔森线性相关系数(Pearson linear correlation coefficient, PLCC)、斯皮尔曼相关系数(Spearman's rank correlation coefficient, SRCC)和均方根误差(root mean square error, RMSE)。其中,PLCC和RMSE用于评价预测的准确性,而SRCC用于评价预测单调性的一致性。除了SRCC外,PLCC和RMSE均按照五参数非线性映射计算,具体的映射函数为

$$F(x) = \rho_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\rho_2(x - \rho_3)}} \right) + \rho_4 x + \rho_5 \quad (11)$$

式中, x 和 $F(x)$ 分别表示预测的质量得分和映射的客观得分, $\rho_1, \rho_2, \dots, \rho_5$ 代表拟合参数。

2.4 性能比较

为了评估本文模型性能的有效性和优势,将提出的模型与目前主流的IQA和OIQA方法进行综合性性能比较,其中包括9种非视口依赖的模型:SSIM(Wang等, 2004)、S-PSNR(spherical PSNR)(Yu等, 2015)、WS-PSNR(weighted-to-spherically uniform PSNR)(Sun等, 2017)、CPP-PSNR(craster's parabolic projection PSNR)(Zakharchenko等, 2016)、WS-SSIM(Zhou等, 2018)、CONTRIQUE(contrastive image quality evaluator)(Madhusudana等, 2022)、MANIQA(multi-dimension attention network)(Yang等, 2022)、VCRNet(visual compensation restoration network)(Pan等, 2022)和LIQE(language-image quality evaluator)(Zhang等, 2023),以及5种视口依赖的模型:MC360IQA(multi-channel CNN)(Sun等, 2020)、VGCN(viewport oriented graph convolution network)(Xu等, 2021)、AHGCN(adaptive hypergraph convolutional network)(Fu等, 2022)、Fang22(Fang等, 2022)和Assessor360(Wu等, 2024)。不同方法的性能比较见表1。

此外,为了直观展示本文方法在轻量级方面的优势,对一些性能较好的评估方法进行模型参数(Params)和计算量(FLOPs)可视化,比较结果如图5所示,本文模型具有显著的轻量性。

从表1的比较结果中有如下发现。首先,以图像块作为输入的2D-IQA方法由于没有考虑全景图像中存在的两极畸变和语义分布特点,在3个数据

表 1 不同模型的性能比较

Table 1 Performance comparison of different models

	输入类型	Params/M ↓	FLOPs/G ↓	CVIQ 数据集		OIQA 数据集		JUFE 数据集	
				PLCC ↑	SRCC ↑	PLCC ↑	SRCC ↑	PLCC ↑	SRCC ↑
SSIM	ERP 图像	N/A	N/A	0.667 2	0.674 1	0.288 6	0.384 6	0.133 0	0.057 0
S-PSNR	ERP 图像	N/A	N/A	0.708 3	0.708 3	0.599 7	0.539 9	0.043 4	0.029 9
WS-PSNR	ERP 图像	N/A	N/A	0.672 9	0.610 7	0.581 9	0.526 3	0.044 4	0.029 9
CPP-PSNR	ERP 图像	N/A	N/A	0.687 1	0.626 5	0.568 3	0.514 9	0.044 0	0.029 8
WS-SSIM	ERP 图像	N/A	N/A	0.929 3	0.911 6	0.504 4	0.503 2	0.032 1	0.026 0
CONTRIQUE	图像块	28.0	26.6	0.456 2	0.467 3	0.172 4	0.170 0	0.104 6	0.095 1
MANIQA	图像块	135.7	108.6	0.287 2	0.279 5	0.367 5	0.308 2	0.126 2	0.118 2
VCRNet	图像块	16.7	10.3	0.390 3	0.392 7	0.425 1	0.382 8	0.146 7	0.120 7
LIQE	图像块	151.0	44.2	0.484 2	0.535 7	0.767 1	0.873 6	0.226 1	0.201 0
MC360IQA*	视口图像	22.4	30.3	0.950 6	0.913 9	0.924 7	0.918 7	0.633 1	0.589 7
VGCN*	视口图像	26.5	191.5	0.965 1	0.963 9	0.958 4	0.951 5	0.699 8	0.662 7
AHGCN	视口图像	12.2	47.8	0.964 3	0.962 3	0.964 9	0.959 0	-	-
Fang22	视口图像	25.2	174.3	-	-	-	-	0.642 2	0.617 1
Assessor360*	视口图像	88.2	230.5	0.976 9	0.964 4	0.964 4	0.964 4	0.812 5	0.807 8
本文	图像块	5.6	10.6	0.945 8	0.933 0	0.944 1	0.935 8	0.664 9	0.632 9

注:N/A表示无法计算该模型的参数量;加粗字体表示最优性能;↑表示值越大越优,↓表示值越小越优;*表示对模型进行修改以适配JUFE数据集,“-”表示未开展实验。

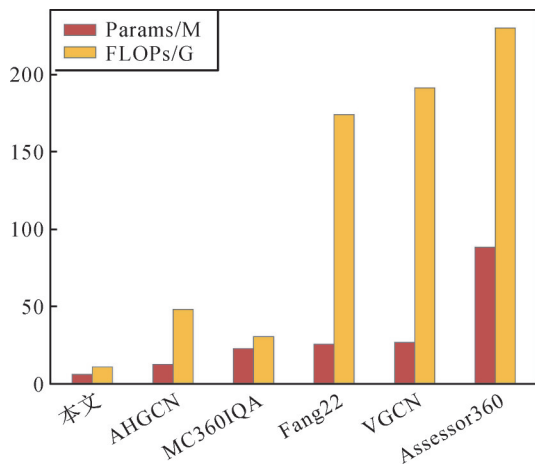


图5 本文方法与测试方法的参数量和计算量

Fig. 5 The Params and FLOPs of the proposed method and test methods

集上均取得了较低的评价性能,这表明当前2D-IQA方法无法将在平面图像上的质量评价能力转移至全景图像,从侧面反映了如何有效处理ERP图像中的两极畸变对于质量评价至关重要。其次,基于ERP图像特征对PSNR、SSIM加以改进的方法,在CVIQ

和OIQA数据集上取得了更高的性能。但是由于手工设计的特征提取难以适应复杂多变的失真环境,因此无法对失真全景图像进行精准评价。尤其是对于JUFE数据集中存在的非均匀失真全景图像,这类评价方法均无法处理图像中的局部降质而导致预测失效。此外,由于视口序列可有效表征用户观看行为相关信息,基于视口图像输入的方法在3个数据集上都取得了较好的评价效果,但这些方法在评价过程中需要进行额外的视口序列预测和提取操作,导致普遍具有较高的模型参数量和计算量。比如,VGCN方法需要通过SURF(speeded up robust features)局部特征检测来定位ERP图像中的关键点,然后通过高斯滤波和下采样获得显著性热力图(Heatmap),最后针对每个关键区域进行像素级坐标投影以提取该部分的视口图像序列,这造成了VGCN方法较高的Params和FLOPs。

尽管视口序列能够有效表征用户观看行为特征,但烦琐的数据预处理过程限制了其在实际工业中的应用。与之相反的是,本文所提出的模型

在 CVIQ、OIQA 和 JUF E 这 3 个数据集上的大部分指标均取得较好性能的前提下, 实现了对模型参数和计算量的高度压缩。具体来说, 本文模型与性能最好的 Assessor360 方法相比, 参数量减少了 93.7%, 计算量减少了 95.4%, 并且在 CVIQ 和 OIQA 上的各项指标综合差距控制在 3% 以内。与参数规模近似的 MC360IQ A 方法相比, 本文模型的 SRCC 指标在 CVIQ 数据集提升了 1.91%、OIQA 数据集提升了 1.71% 以及在 JUF E 数据集提升了

4.32%。

为了进一步验证本文模型在实际工业使用时的预测效果, 本文分别在实验的 3 个数据集上选取 4 幅质量不同的全景图像进行质量预测, 可视化结果如图 6 所示。预测结果表明, 本文模型在 CVIQ 和 OIQA 两个数据集上的预测结果较为准确, 与真实质量 (MOS) 相差较小; 在 JUF E 数据集上发现有部分图像的预测质量与 MOS 存在一定差距, 但是能够判断图像的大致质量。

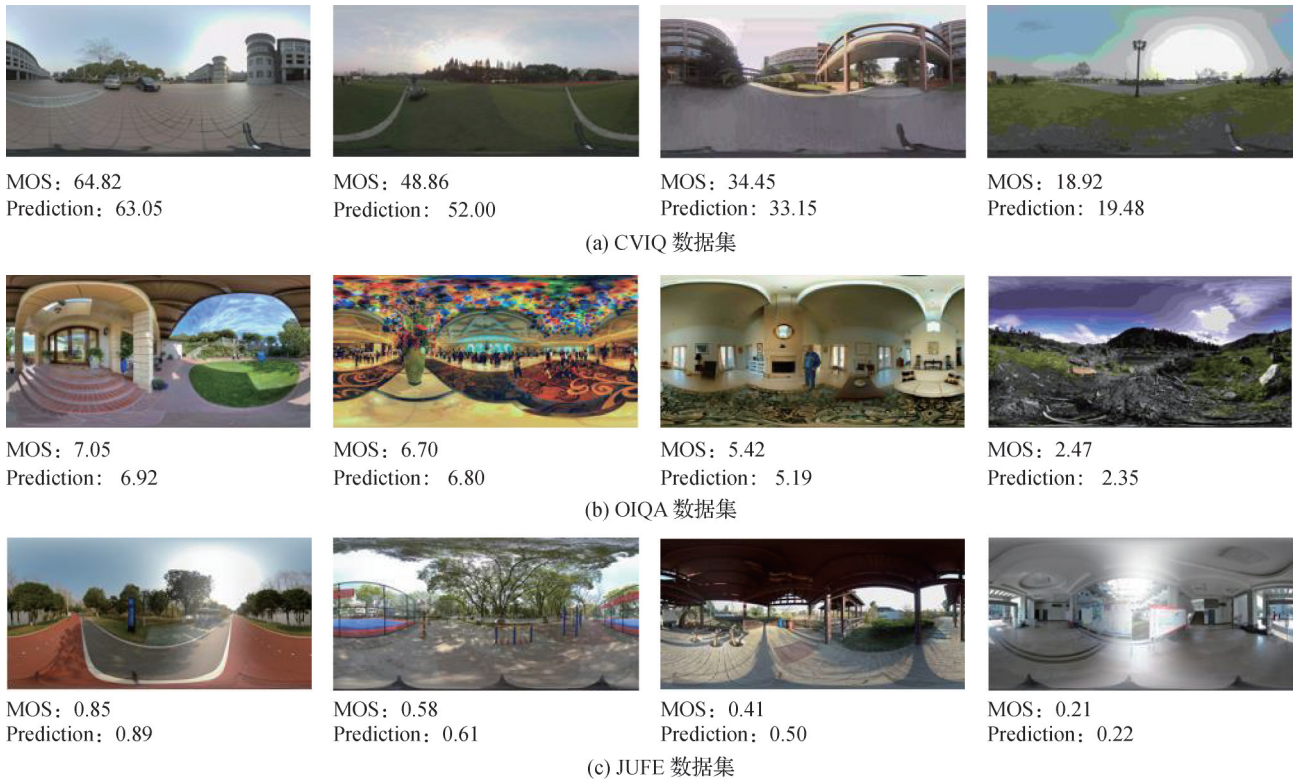


图 6 本文模型在 3 个数据集上的质量预测示例

Fig. 6 The examples of quality prediction of the proposed model on three databases
((a) CVIQ database; (b) OIQA database; (c) JUF E database)

2.5 消融实验

2.5.1 卷积方法

为了验证 DUFE 模块中等矩形可变形卷积的抗畸变的有效性, 在其他实验设置保持不变的情况下, 分别将该模块中的等矩形可变形卷积全部替换为标准卷积和可变形卷积, 以定量分析不同卷积方式给模型畸变特征自适应提取能力带来的影响程度, 在 CVIQ 数据集上进行该消融实验。实验结果如表 2 所示, 使用等矩形可变形卷积与使用标准卷积和可变形卷积相比, PLCC 分别提升了 6.51% 和 7.87%, SRCC 分别提升了 7.15% 和 10.42%, 证明了本文所

提出的卷积方法可有效解决不规则语义问题和两极畸变问题。

表 2 不同卷积操作对模型性能的影响

Table 2 Impact of different convolution operations

卷积	PLCC ↑	SRCC ↑	RMSE ↓
标准卷积	0.880 7	0.861 5	6.379 8
可变形卷积	0.867 1	0.828 8	6.709 6
等距可变形卷积	0.945 8	0.933 0	4.375 5

注: 加粗字体表示最优性能, ↑ 表示值越大越优, ↓ 表示值越小越优。

2.5.2 采样位置

为了探究多样图像块集合对模型学习的影响,实验过程中每个 epoch 都选取一组相同的采样位置进行训练。这意味着模型在整个训练过程中都依赖于相同的样本集合,而不考虑采样位置的变化。相反,在非固定采样位置实验中,每个 epoch 都对图像块集合进行重新随机采样,为模型提供多样化的图像块表征以及位置变化相关的信息。在模型的训练和评价过程中,本文确保两种采样方式下的其他超参数和设置相同,以消除潜在的干扰因素。在 CVIQ 数据集的结果如表 3 所示,在每个 epoch 不固定采样位置的性能显著优于固定采样位置的性能,这意味着在训练时引入多样的图像块集合有助于提高模型的质量特征感知能力以及预测精度。

表 3 固定采样位置以及非固定采样位置的性能对比

Table 3 Comparison of fixed and non-fixed sampling schemes

位置	PLCC ↑	SRCC ↑	RMSE ↓
固定位置采样	0.844 4	0.778 2	7.217 0
不固定位置采样	0.945 8	0.933 0	4.375 5

注:加粗字体表示最优性能,↑表示值越大越优,↓表示值越小越优。

2.5.3 纬度划分因子

为探究赤道先验分布与图像块采样密度之间的潜在关联,对纬度划分因子 θ_r 进行不同取值,以确定最优纬度区域划分。在 OIQA 数据集进行该消融实验,结果如表 4 所示。当 θ_r 为 0° ,划分线位于 ERP 图像的赤道位置,意味着对图像的上半部分和下半部分采样相同数量的图像块,且无法得到赤道位置的图像块,这与赤道先验分布存在矛盾,可见此时模型取得最差性能。随着 θ_r 的不断增大,更多图像块取自以赤道为中心的低纬区域,模型性能逐渐上升。当 θ_r 到达 23° ,即此时划定的低纬区域被用户观看到的概率为 60%,并从该区域采样占总图像块数量 60% 的图像块,模型性能取得最优。而当 θ_r 继续增加,由于赤道先验分布给采样位置带来的约束效果减弱,模型性能逐渐下降。该实验证明选取一个适合的纬度划分因子对本文模型获得最优质量评价结果至关重要。

2.5.4 图像块个数和尺寸分析

为了研究图像块采样个数和大小对模型性能的

表 4 纬度划分因子影响

Table 4 Impact of the latitude division factor

θ_r	低纬占比/%	PLCC ↑	SRCC ↑	RMSE ↓
90	100	0.916 0	0.903 7	0.845 7
53	80	0.941 3	0.925 8	0.711 3
23	60	0.944 1	0.935 8	0.694 9
13	40	0.934 6	0.924 6	0.749 9
6	20	0.927 6	0.921 4	0.787 4
0	0	0.695 4	0.619 5	1.514 7

注:加粗字体表示最优性能,↑表示值越大越优,↓表示值越小越优。

影响,本文分别选择了 1 个、5 个、10 个和 20 个采样图像块作为模型的输入,并在每种数量级上依次设置了 28×28 、 56×56 、 112×112 以及 224×224 像素 4 种尺度的图像块尺寸。在 CVIQ 数据集上进行该实验,实验结果如图 7 所示。在不同图像块尺寸的情况下,随着图像块数量的不断增加,模型性能均得到提升,但数量过多时性能反而出现下降,可能是信息冗余造成模型优化困难。当从一幅 ERP 图像中提取 10 个图像块,并且每个图像块的尺寸为 224×224 像素时,模型取得最优性能,表明此时的图像块集合能够较好地表征全景图像的整体质量。

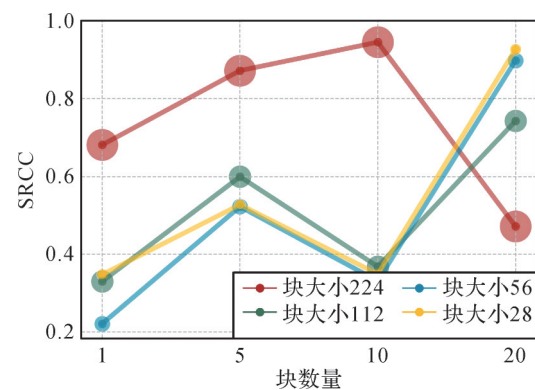


图 7 图像块数量和尺寸的影响

Fig. 7 Impact of the number of patches and their size

基于设计的 PPS、DUFE 和 A-EPAA 模块构建 OIQA 模型,并在构建过程中兼顾模型的有效性和轻便性。实验证明,本文模型与其他同类方法相比,和最优性能相差不大的同时模型参数数量和计算量显著较少,在 CVIQ 数据集(Sun 等,2018)上的模型比较可视化结果见图 8,图中圆的大小表示该模型的参数量。

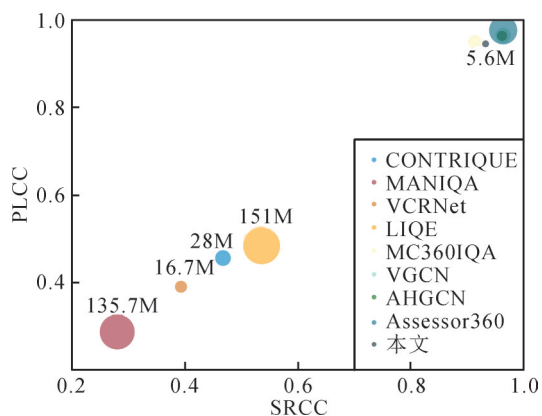


图8 模型性能与尺寸比较

Fig. 8 Comparison of model performance against size

3 结论

针对视口依赖模型中存在视口序列预测难度大、视口提取计算负担高的问题,本文提出一种非视口依赖的抗畸变NR-OIQA模型,无需对视口序列进行预测和提取即可对全景图像进行准确的质量评价。具体而言,本文通过设计的图像块采样模块(PPS),可获取全景图像的充分内容表征。在抗畸变特征提取模块(DUFE)中,结合DCN中可学习的位置偏移和ERP图像中固有规律性畸变偏移,提出了一种可同时解决不规则语义和规律性畸变的EquiMdconv,用于处理全景图像中复杂特征关系。通过块内一块间注意力聚集模块(A-EPAA)对提取的特征进行动态调整,以提高模型特征表达能力。为了验证模型性能,本文在3个公开发布的数据库上与9个非视口依赖方法和5个视口依赖方法进行了性能比较。实验结果表明,相比于现有的非视口依赖方法,本文所提出的模型在性能上具有明显的优势。在视口依赖方法比较中,本文模型具有较好的性能,并且在模型参数量与计算量方面,与性能最好的模型相比,参数量与计算量分别减小了15倍和21倍。

参考文献 (References)

Chen S J, Zhang Y X, Li Y M, Chen Z Z and Wang Z. 2018. Spherical structural similarity index for objective omnidirectional video quality assessment//Proceedings of 2018 IEEE International Conference on Multimedia and Expo. San Diego, USA: IEEE: 1-6 [DOI: 10.

1109/ICME.2018.8486584]

Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]

Deng X, Wang H, Xu M, Guo Y C, Song Y H and Yang L. 2021. LAU-Net: latitude adaptive upscaling network for omnidirectional image super-resolution//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9185-9194 [DOI: 10.1109/CVPR46437.2021.00907]

Djilali Y A D, McGuinness K and O'Connor N E. 2021. Simple baselines can fool 360° saliency metrics//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 3743-3749 [DOI: 10.1109/ICCVW54120.2021.00418]

Duan H Y, Zhai G T, Min X K, Zhu Y C, Fang Y and Yang X K. 2018. Perceptual quality assessment of omnidirectional images//Proceedings of 2018 IEEE International Symposium on Circuits and Systems. Florence, Italy: IEEE: 1-5 [DOI: 10.1109/ISCAS.2018.8351786]

Fang Y M, Huang L P, Yan J B, Liu X L and Liu Y. 2022. Perceptual quality assessment of omnidirectional images//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 580-588 [DOI: 10.1609/aaai.v36i1.19937]

Fang Y M, Zhong Y, Yan J B and Liu L X. 2022. Deep attention guided image cropping with fine-grained feature aggregation. *Journal of Image and Graphics*, 27(2): 586-601 (方玉明, 钟裕, 鄢杰斌, 刘丽霞. 2022. 聚合细粒度特征的深度注意力自动裁图. *中国图象图形学报*, 27(2): 586-601) [DOI: 10.11834/jig.210544]

Fernandez-Labrador C, Facil J, Perez-Yus A, Demonceaux C, Civera J and Guerrero J J. 2020. Corners for layout: end-to-end layout recovery from 360 images. *IEEE Robotics and Automation Letters*, 5(2): 1255-1262 [DOI: 10.1109/LRA.2020.2967274]

Fu J, Hou C, Zhou W, Xu J H and Chen Z B. 2022. Adaptive hypergraph convolutional network for no-reference 360-degree image quality assessment//Proceedings of the 30th ACM International Conference on Multimedia. Lisbon, Portugal: ACM: 961-969 [DOI: 10.1145/3503161.3548337]

Hands D S and Avons S E. 2001. Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology*, 15(6): 639-657 [DOI: 10.1002/acp.731]

He W and Pan C. 2022. The salient object detection based on attention-guided network. *Journal of Image and Graphics*, 27(4): 1176-1190 (何伟, 潘晨. 2022. 注意力引导网络的显著性目标检测. *中国图象图形学报*, 27(4): 1176-1190) [DOI: 10.11834/jig.200658]

Kim H G, Lim H T and Ro Y M. 2020. Deep virtual reality image quality assessment with human perception guider for omnidirectional image. *IEEE Transactions on Circuits and Systems for Video Tech-*

- nology, 30(4): 917-928 [DOI: 10.1109/TCSVT.2019.2898732]
- Leng J X, Mo M J C, Zhou Y H, Ye Y M, Gao C Q and Gao X B. 2023. Recent advances in drone-view object detection. *Journal of Image and Graphics*, 28(9): 2563-2586 (冷佳旭, 莫梦竞成, 周应华, 叶永明, 高陈强, 高新波. 2023. 无人机视角下的目标检测研究进展. *中国图象图形学报*, 28(9): 2563-2586) [DOI: 10.11834/jig.220836]
- Li C, Xu M, Jiang L, Zhang S Y and Tao X M. 2019. Viewport proposal CNN for 360° video quality assessment//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE: 10169-10178 [DOI: 10.1109/CVPR.2019.01042]
- Li D Q, Jiang T T and Jiang M. 2020. Norm-in-norm loss with faster convergence and better performance for image quality assessment//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM: 789-797 [DOI: 10.1145/3394171.3413804]
- Li L D, Yin Y T, Wu J J, Dong W S and Shi G M. 2022. Mask-fused human face image quality assessment method. *Journal of Image and Graphics*, 27(12): 3476-3490 (李雷达, 殷杨涛, 吴金建, 董伟生, 石光明. 2022. 掩膜融合下的人脸图像质量评价方法. *中国图象图形学报*, 27(12): 3476-3490) [DOI: 10.11834/jig.210986]
- Lim H T, Kim H G and Ra Y M. 2018. VR IQA NET: deep virtual reality image quality assessment using adversarial learning//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE: 6737-6741 [DOI: 10.1109/ICASSP.2018.8461317]
- Liu H F, Chen J J, Li L, Bao B K, Li Z C, Liu J Y and Nie L Q. 2023. Cross-modal representation learning and generation. *Journal of Image and Graphics*, 28(6): 1608-1629 (刘华峰, 陈静静, 李亮, 鲍秉坤, 李泽超, 刘家瑛, 聂礼强. 2023. 跨模态表征与生成技术. *中国图象图形学报*, 28(6): 1608-1629) [DOI: 10.11834/jig.230035]
- Madhusudana P C, Birkbeck N, Wang Y L, Adsumilli B and Bovik A C. 2022. Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31: 4149-4161 [DOI: 10.1109/TIP.2022.3181496]
- Pan Z Q, Yuan F, Lei J J, Fang Y M, Shao X and Kwong S. 2022. VCRNet: visual compensation restoration network for no-reference image quality assessment. *IEEE Transactions on Image Processing*, 31: 1613-1627 [DOI: 10.1109/TIP.2022.3144892]
- Sun W, Gu K, Ma S W, Zhu W H, Liu N and Zhai G T. 2018. A large-scale compressed 360-degree spherical image database: from subjective quality evaluation to objective model comparison//Proceedings of the 20th IEEE International Workshop on Multimedia Signal Processing. Vancouver, Canada: IEEE: 1-6 [DOI: 10.1109/MMSP.2018.8547102]
- Sun W, Min X K, Zhai G T, Gu K, Duan H Y and Ma S W. 2020. MC360IQA: a multi-channel CNN for blind 360-degree image quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 14(1): 64-77 [DOI: 10.1109/JSTSP.2019.2955024]
- Sun Y L, Lu A and Yu L. 2017. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE Signal Processing Letters*, 24(9): 1408-1412 [DOI: 10.1109/LSP.2017.2720693]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Wang Z and Rehman A. 2019. Begin with the end in mind: a unified end-to-end quality-of-experience monitoring, optimization and management framework. *SMPTE Motion Imaging Journal*, 128(2): 1-8 [DOI: 10.5594/JMI.2018.2887288]
- Woo S, Park J, Lee J and Kweon I S. 2018. CBAM: Convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Wu T H, Shi S W, Cai H M, Cao M D, Xiao J, Zheng Y Q and Yang Y J. 2024. Assessor360: multi-sequence network for blind omnidirectional image quality assessment//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 64957-64970 [DOI: 10.5555/3666122.3668956]
- Xu J H, Zhou W and Chen Z B. 2021. Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5): 1724-1737 [DOI: 10.1109/TCSVT.2020.3015186]
- Yan J B, Fang Y M and Liu X L. 2022. The review of distortion-related image quality assessment. *Journal of Image and Graphics*, 27(5): 1430-1466 (鄢杰斌, 方玉明, 刘学林. 2022. 图像质量评价研究综述——从失真的角度. *中国图象图形学报*, 27(5): 1430-1466) [DOI: 10.11834/jig.210790]
- Yang S D, Wu T H, Shi S W, Lao S S, Gong Y, Cao M D, Wang J H and Yang Y J. 2022. MANIQA: multi-dimension attention network for no-reference image quality assessment//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 1190-1199 [DOI: 10.1109/CVPRW56347.2022.00126]
- Yu M, Lakshman H and Girod B. 2015. A framework to evaluate omnidirectional video coding schemes//Proceedings of 2015 IEEE International Symposium on Mixed and Augmented Reality. Fukuoka, Japan: IEEE: 31-36 [DOI: 10.1109/ISMAR.2015.12]
- Zakharchenko V, Choi K P and Park J H. 2016. Quality metric for spherical panoramic video//Proceedings of Optics and Photonics for Information Processing X. San Diego, US: SPIE: #99700 [DOI: 10.1117/12.2235885]
- Zhang C F and Liu S G. 2022. No-reference omnidirectional image quality assessment based on joint network//Proceedings of the 30th ACM International Conference on Multimedia. Lisbon, Portugal:

- ACM: 943-951 [DOI: 10.1145/3503161.3548175]
- Zhang W X, Zhai G T, Wei Y, Yang X K and Ma K D. 2023. Blind image quality assessment via vision-language correspondence: a multitask learning perspective//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 14071-14081 [DOI: 10.1109/CVPR52729.2023.01352]
- Zhou Y, Wang Y, Li L D, Gao C Q and Lu Z L. 2022. Research progress in objective quality evaluation of virtual reality images. *Journal of Image and Graphics*, 27(8): 2313-2328 (周玉, 汪一, 李雷达, 高陈强, 卢兆林. 2022. 虚拟现实图像客观质量评价研究进展. *中国图象图形学报*, 27(8): 2313-2328) [DOI: 10.11834/jig.210949]
- Zhou Y F, Yu M, Ma H L, Shao H and Jiang G Y. 2018. Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video//Proceedings of the 14th IEEE International Conference on Signal Processing. Beijing, China: IEEE: 54-57 [DOI: 10.1109/ICSP.2018.8652269]
- Zhu X Z, Hu H, Lin S and Dai J F. 2019. Deformable ConvNets V2: more deformable, better results//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 9300-9308 [DOI: 10.1109/CVPR.2019.00953]
- Zuo Y F, Fang Y M and Ma K D. 2023. The critical review of the growth of deep learning-based image fusion techniques. *Journal of Image and Graphics*, 28(1): 102-117 (左一帆, 方玉明, 马柯德. 2023. 深度学习时代图像融合技术进展. *中国图象图形学报*, 28(1): 102-117) [DOI: 10.11834/jig.220556]

作者简介

鄢杰斌,男,博士,讲师,主要研究方向为视觉质量评价。

E-mail:jiebinyan@foxmail.com

方玉明,通信作者,男,教授,主要研究方向为计算机视觉、多媒体信号处理、视觉质量评价。

E-mail:leo.fangyuming@foxmail.com

谭淄文,男,硕士研究生,主要研究方向为全景图像质量评价。E-mail:ziwentan@foxmail.com

吴康诚,男,硕士研究生,主要研究方向为全景图像质量评价。E-mail:kangchengwu-my@foxmail.com

刘学林,男,博士,主要研究方向为视觉质量评价。

E-mail:xuelinliu-bill@foxmail.com