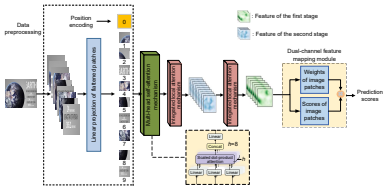
Opto-Electronic Engineering
光 电 工 程

Article

2025 年, 第 52 卷, 第 4 期

DOI: 10.12086/oee.2025.240309 CSTR: 32245.14.oee.2025.240309

基于多任务注意力机制的无参考屏幕内容图像质量评价算法



周子镱，董　武*，陆利坤，马　倩，侯国鹏，张二青

北京印刷学院高端印刷装备信号与信息处理北京市重点实验室，北京 102600

摘要：提出一种基于多任务注意力机制的无参考屏幕内容图像质量评价算法 (multi-task attention mechanism based no reference quality assessment algorithm for screen content images, MTA-SCI)。MTA-SCI 首先使用自注意力机制提取屏幕内容图像的全局特征，增强对屏幕内容图像整体信息的表征能力；然后使用综合局部注意力机制提取屏幕内容图像的局部特征，使局部特征能够聚焦于屏幕内容图像中更吸引人注意的细节部分；最后使用双通道特征映射模块预测屏幕内容图像的质量分数。在 SCID 和 SIQAD 数据集上，MTA-SCI 的斯皮尔曼秩序相关系数 (Spearman's rank order correlation coefficient, SRCC) 分别达到 0.9602 和 0.9233，皮尔森线性相关系数 (Pearson linear correlation coefficient, PLCC) 分别达到 0.9609 和 0.9294。实验结果表明，MTA-SCI 在预测屏幕内容图像质量任务中具有较高的准确性。

关键词：屏幕内容图像；无参考图像质量评价；vision transformer；多级视觉感知特性；注意力机制

中图分类号：TP391.9　　**文献标志码：**A

周子镱，董武，陆利坤，等. 基于多任务注意力机制的无参考屏幕内容图像质量评价算法 [J]. 光电工程，2025，**52**(4): 240309
Zhou Z Y, Dong W, Lu L K, et al. Multi-task attention mechanism based no reference quality assessment algorithm for screen content images[J]. *Opto-Electron Eng*, 2025, **52**(4): 240309

Multi-task attention mechanism based no reference quality assessment algorithm for screen content images

Zhou Ziyi, Dong Wu*, Lu Likun, Ma Qian, Hou Guopeng, Zhang Erqing

Beijing Key Laboratory of Signal and Information Processing for High-end Printing Equipment, Beijing Institute of Graphic Communication, Beijing 102600, China

Abstract: This paper proposed a multi-task attention mechanism-based no-reference quality assessment algorithm for screen content images (MTA-SCI). The MTA-SCI first used a self-attention mechanism to extract global features from screen content images, enhancing the representation of overall image information. It then applied an integrated local attention mechanism to extract local features, allowing the focus to be on attention-grabbing details

收稿日期：2024−12−30；修回日期：2025−02−24；录用日期：2025−02−25

基金项目：北京市数字教育研究重点课题 (BDEC2022619027)；北京市高等教育学会 2023 年立项面上课题 (MS2023168)；北京印刷学院校级科研项目 (Ec202303, Ea202301, E6202405)；北京印刷学院学科建设和研究生教育专项 (21090224002, 21090323009, 21090124013)；北京印刷学院出版学新兴交叉学科平台建设项目 (04190123001/003)；北京邮电大学网络与交换技术全国重点实验室开放课题资助项目 (SKLNST-2023-1-12)

*通信作者：董武，dongwu@bigc.edu.cn。

within the image. Finally, a dual-channel feature mapping module predicted the quality score of the screen content image. On the SCID and SIQAD datasets, MTA-SCI achieves Spearman's rank-order correlation coefficients (SROCC) of 0.9602 and 0.9233, and Pearson linear correlation coefficients (PLCC) of 0.9609 and 0.9294, respectively. The experimental results show that the MTA-SCI achieves high accuracy in predicting screen content image quality.

Keywords: screen content image; no-reference image quality assessment; vision transformer; multi-level visual system of human; attention mechanism

1 引　言

随着计算机远程交互技术的快速发展，在多媒体交互过程中产生了大量不同于自然图像的屏幕内容图像[1-2]。这些图像在获取、压缩、传输和显示等处理过程中，会经历各种不同类型、不同程度的失真。为了衡量屏幕内容图像的失真程度，研究人员提出了度量屏幕内容图像的质量评价方法。在屏幕内容图像的处理过程中，这些评价方法能够提供理论指导和技术支撑。此外，由于屏幕内容图像与自然图像在特性上存在显著差异，其失真问题也呈现出独特的挑战。例如，屏幕内容图像通常包含高对比度的文字、图表以及锐利的边缘细节，而这些特性在常见的失真类型下容易受到明显影响，具体表现形式包括文字边缘模糊、颜色失真、细节丢失以及内容结构的破坏，这些失真显著损害了用户的视觉体验。目前已有的自然图像质量评价算法[3]不能有效地应用于屏幕内容图像[4]，因此非常有必要开展屏幕内容图像质量评价的研究工作。

根据不同的特征提取方式，屏幕内容图像客观质量评价方法分为两类：基于手工提取特征的方法和基于深度学习自动提取特征的方法[5]，如表 1 所示。上述两类方法又可以细分为全参考 (full reference, FR)、无参考 (no reference, NR)、半参考 (reduced reference, RR) 屏幕内容图像质量评价方法。

在第一类方法中，依据人眼对屏幕内容图像具有不同的视觉感受进行质量评价。Yang 等[6]提出了 SPQA 算法，该算法同时使用图像亮度和清晰度的相似程度作为衡量文本区域的指标，并只使用清晰度的相似程度作为衡量图像区域的指标。BLIQUP-SCI[7]算法使用 BRISQUE 方法和局部二值模式 (local binary pattern, LBP) 分别提取屏幕内容图像中图像块和文本块的局部稀疏特征。由于屏幕内容图像中边缘特征的失真更容易影响图像的质量评价，Yang 等[8]使用梯度图的局部二值模式直方图来表示文本区域的边缘特征。由于边缘特征能准确地反映屏幕内容图像的失真情况，Ni 等[9]提出边缘相似度 (edge similarity, ESIM) 算法，该算法提取边缘的对比度、边缘的宽度和边缘的方向作为特征。基于人眼对屏幕内容图像中色彩和纹理的敏感特性，Huang 等[10]从屏幕内容图像的色度通道中提取梯度幅值、归一化定向直方图和局部离散余弦变换系数作为屏幕内容图像的统计特征。手工提取特征的方法获得的特征表达能力有限，所以这类方法很难获得较好的质量评价效果。

在第二类方法中，SIQA-DF-II[11]算法首先分割图像得到图像块，然后使用 FR 屏幕内容图像质量评价算法生成图像块的训练标签，用来辅助 NR 屏幕内容图像的质量评价。在此基础上，Jiang 等[12]提出了 QODCNN 算法，该算法筛除了预测分数与差分平均主观值 (differential mean opinion score, DMOS) 差异过大的图像块，以此减少对他们图像质量评价算法的影响。Zuo 等[13]提出了 MIC-CNN 算法，该算法将人眼视觉系统 (human visual system, HVS) 的特性与深度学习方法相结合，充分考虑了文本内容和图像内容不同的视觉特征。SR-CNN[10]算法将局部特征和全局特征融合为深层特征，使用深层特征评价算法，其性能明显优于仅使用局部特征评价算法的性能。基于区域的屏幕内容图像评价算法 (region image quality assessment, RIQA)[14]在提取图像局部特征的基础上，引入图像质量分数排序任务和噪声分类任务等多任务分支，能够更好地符合 HVS 的感知特性。Gao 等[15]提出新颖的双通道卷积神经网络 (convolutional neural network, CNN) 算法，该算法使用方向梯度直方图辅助 CNN 获得屏幕内容图像的视觉感知质量。在此基础上，Zhang 等[16]使用了双分支模块和残差连接的方式对神经网络特征提取模块加以改进，这些改进一方面增强了模型对屏幕内容图像视觉感知特征的描述能力，另一方面降低了模型的复杂度。MTDL[17]算法通

过分析屏幕内容图像的失真类型和失真程度来预测屏幕内容图像的质量，同时结合注意力模块[18]去模拟人眼处理视觉信号的机制，从而提高评估模型的效果。

由表 1 可知，目前基于深度学习的自动特征提取方法主要依赖于 CNN[22]。由于 CNN 具有平移不变性和局部性，所以 CNN 善于提取图像的局部特征，但对空间位置关系等全局特征的建模能力比较差。此外，现有的屏幕内容图像客观质量评价方法还存在两个缺点：1) 由于人眼不易察觉屏幕内容图像中平坦区域的噪声，导致屏幕内容图像的质量预测分数与主观评价值出现较大差异；2) 现有算法未充分考虑人眼的多级视觉感知特性[23]，而这一特性在视觉质量评价中至关重要。人眼的多级感知特性反映了视觉信息处理的复杂性，即人眼对不同区域的敏感度存在一定差异。例如，在屏幕内容图像质量评价的过程中，人眼在初步扫视时往往会先获取布局信息等全局特征，再将注意力集中在文本标题和文章首句等局部特征；此外，屏幕内容图像通常会使用底纹掩膜来突出文本区域，使得文本部分的局部特征更容易吸引人眼的关注。因此，如果在客观的屏幕内容图像质量评价方法中考虑多级感知特性，就能更准确地模拟人眼对屏幕内容图像质量的主观评价特点。

针对上述分析，本文提出了一种基于多任务注意力机制的无参考屏幕内容图像客观质量评价算法 (multi-tasks attention mechanism based no reference quality assessment algorithm for screen content images, MTA-SCI)。在本文中，多任务注意力机制包括三部分：多头自注意力机制、空间分组注意力机制和非对称卷积通道注意力机制。这三种注意力机制在本文提出的 MTA-SCI 中承担着不同的任务，它们共同提升屏幕内容图像质量评价的性能。该算法同时提取屏幕内容图像的全局特征和局部特征，其中全局特征的提取可以模拟人眼多级视觉感知特点中识别宏观布局和结构的特点，局部特征的提取过程则模拟人眼多级视觉感知特点中感知图像中局部细节的特点。在此算法中，首先，由于视觉转换器 (vision transformer, ViT)[24-25] 可以建立视觉的长距离关系，所以本文使用 ViT 中的多头自注意力机制提取屏幕内容图像的全局特征，能更好地从全局特征的角度理解屏幕内容图像的质量，并能够把数值较大的注意力权重分配到文本的标题、文章首句、大面积的插图等相关区域。然后，为了增强局部特征对屏幕内容图像的表征能力，提出了综合局部注意力机制，该注意力机制由空间分组注意力机制和非对称卷积通道注意力机制组成。综合局部注意力机制具有两个功能：一方面此机制提取了屏幕内容图像中更加能引起人眼注意的局部特征，使模型更专注于屏幕内容图像中特有的局部特征的学习；

表 1　典型的屏幕内容图像质量评价算法

Table 1　Typical methods of screen content image quality assessment

Category	Method	Type	Feature
The first category	SPQA[6]	FR	Brightness and sharpness
	ESIM[9]	FR	Edge contrast
	MSDL[19]	FR	Feature extraction using log gabor filters
	BLIQUP-SCI[7]	NR	Natural scene statistics features and local texture
	Yang et al.[8]	NR	The amplitude, variance, entropy, and edge structure of wavelet coefficients
	Huang et al.[20]	RR	Oriented histogram, local discrete cosine transform coefficients, and gradient of amplitude in color channels
The second category	SR-CNN[10]	FR	Multi-level CNN features
	QODCNN[12]	FR/NR	CNN features
	Gao et al.[15]	NR	CNN features
	Zhang et al.[16]	NR	CNN features
	MIC-CNN[13]	NR	CNN features
	SIQA-DF-II[11]	NR	CNN features
	RIQA[14]	NR	CNN features
	DAMC[21]	FR	CNN features
	MTDL[17]	NR	CNN features

另一方面，此机制通过对不同通道的特征赋予不同的权重，能够抑制含有背景纹理噪声的通道，从而减少背景纹理噪声对图像质量评价产生的影响。最后，使用双通道特征映射模块得到图像块的质量分数与注意力权重，并使用加权求和的方法得到屏幕内容图像整体内容的质量分数。在本文提出的 MTA-SCI 中，ViT 深度模型起到了重要的作用。ViT 强大的特征提取能力使得模型能够更好地捕捉图像中的全局信息，从而提高了屏幕内容图像质量评价的准确性。

2 基于多任务注意力机制的算法

为了解决 CNN 在屏幕内容图像质量评价中的局限性，同时充分考虑人眼的视觉感知特点，本文提出的 MTA-SCI 利用多头自注意力机制和综合局部注意力机制分别提取屏幕内容图像的全局特征和局部特征，这些特征更符合人眼的视觉多级特性。这种多任务注意力机制的共同作用模拟了人眼的多级视觉感知特性，使模型能够兼顾局部细节和全局特征。

MTA-SCI 包含四个部分：数据预处理、多头自注意力机制、综合局部注意力机制和双通道特征映射模块，其中多头自注意力机制、综合局部注意力机制分别提取了屏幕内容图像的全局特征和局部特征，MTA-SCI 的网络结构如图 1 所示。此算法首先对失真图像进行分割，得到多个图像块，并对图像块进行线性投影操作从而得到对应的向量序列。线性投影操作包括两部分：一维展平和线性映射。然后，使用多头自注意力机制得到具有长距离相关性的全局特征向量。接着，使用综合局部注意力机制获得多尺度的局部特征。最后，使用双通道特征映射模块得到图像块的预测分数与预测权重，并对他们进行加权求和，从而得到失真图像的质量分数。

2.1 多头自注意力机制

在屏幕内容图像质量评价领域中，CNN 通常作为主干网络提取屏幕内容图像的特征。然而其局部归纳偏置阻碍了模型提取到屏幕内容图像所有区域的信息，而且它的平移不变性导致其无法有效地处理屏幕内容图像中的复杂组合特征。自注意力机制计算输入序列中的每个元素与整个序列中其他元素之间的相关性，并生成注意力特征图。本文的多头自注意力机制中包含并行的多个自注意力层，能够同时捕捉输入序列在不同子空间中的信息，从而增强模型的表征能力。本文使用 ViT-B/8 作为主干网络，其使用了 8 个自注意力层 (h=8) 提取屏幕内容图像的全局特征，如式 (1)、式 (2) 和式 (3) 所示。

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{K}}}}\right)\boldsymbol{V}, \tag{1}$$

$$\mathrm{Multihead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(\mathrm{head}_1,\cdots,\mathrm{head}_h)\boldsymbol{W}^{\mathrm{O}}, \tag{2}$$

$$\mathrm{head}_i=\mathrm{Attention}\left(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},\boldsymbol{V}\boldsymbol{W}_i^{V}\right), i=1,2,\cdots,h, \tag{3}$$

式中：$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示查询矩阵、键矩阵、值矩阵；$\boldsymbol{K}^{\mathrm{T}}$ 表示键矩阵的转置；d_{K} 表示键向量的维度；Softmax(·) 表示将矩阵中的数据进行归一化处理；

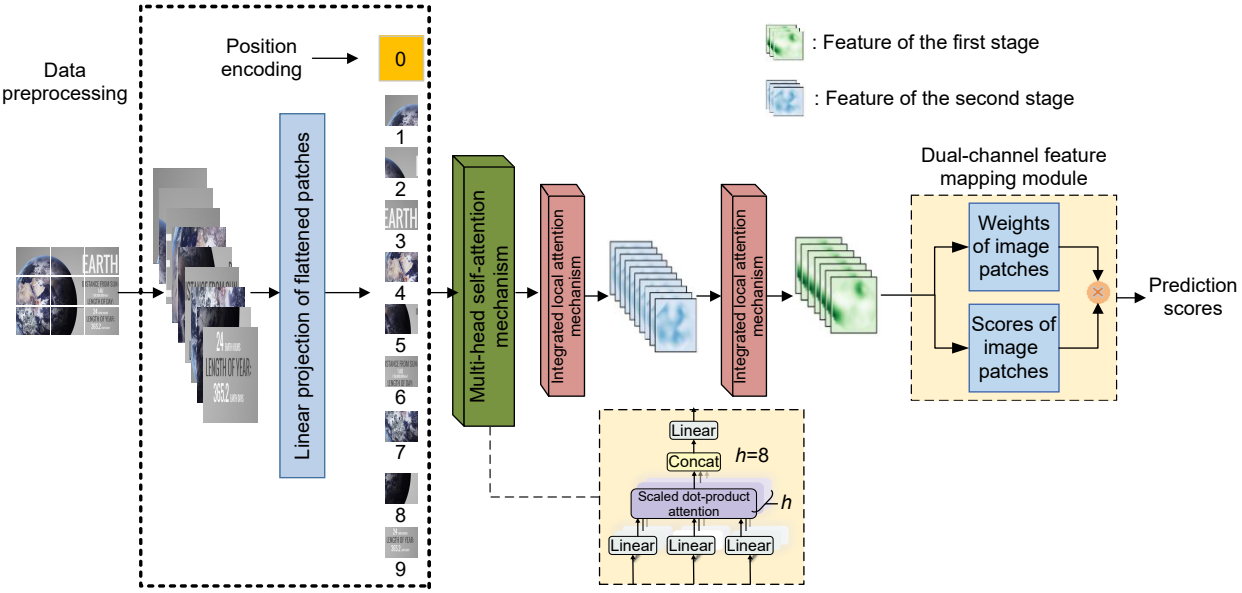


图 1 本文提出的 MTA-SCI 算法的网络结构

Fig. 1 Structure of the MTA-SCI proposed in the paper

Attention($\boldsymbol{Q}$,$\boldsymbol{K}$,$\boldsymbol{V}$) 表示缩放点积注意力的计算；head_h 表示 h 个自注意力层；Concat(·) 函数表示聚合 h 个自注意力层的操作；$\boldsymbol{W}^O$ 表示线性变换矩阵；Multihead ($\boldsymbol{Q}$,$\boldsymbol{K}$,$\boldsymbol{V}$) 表示多头自注意力机制的计算。

在屏幕内容图像的自注意力计算中，首先对失真图像进行预处理。具体来说，使用双三次插值方法将失真图像的尺寸调整为 224×224。然后，按照随机生成的起始位置和指定的裁剪尺寸从图像中裁剪出 28 个尺寸为 8×8 的图像块。最后，使用线性投影层获得对应图像块的一维展平序列和位置编码序列。

2.2 综合局部注意力机制

为了增强局部特征对屏幕内容图像的表征能力，本文提出了综合局部注意力机制，其结构如图 2 所示，其中 B 表示批量的大小。在图 2 中，该注意力机制包含两组空间分组注意力机制和一组非对称卷积通道注意力机制，并采用了残差连接的方式，该连接方式能够缓解梯度消失的问题。在本文提出的 MTA-SCI 网络结构示意图中 (图 1)，使用了两组综合局部注意力机制，能够在逐层的注意力计算[26-27] 中捕捉更复杂的层次关系。在图 1 中，使用第一组综合局部注意力机制得到了第一阶段的特征，把第一阶段的特征送给第二组综合局部注意力机制，从而得到第二阶段的特征。这种特征提取的方式能够全面增强模型的多层次特征表达能力。

空间分组注意力机制和非对称卷积通道注意力机制在 MTA-SCI 中承担着不同的任务，他们共同提升屏幕内容图像质量评价的性能。其中，空间分组注意力机制主要负责提取更能引起人眼视觉注意的高频信息，这些信息通常来源于文本、图标等细节丰富的内容，此机制模拟了人眼对显著区域的感知特点。另一方面，非对称卷积通道注意力机制专注于捕捉不同尺度的局部特征，它使用不同尺寸的卷积核去充分提取屏幕内容图像的多尺度结构特征。此外，非对称卷积通道注意力机制还承担着特征筛选的任务，通过为不同通道的特征赋予不同的权重，有效抑制包含背景纹理噪声或冗余信息的通道对屏幕内容图像质量评价的干扰。

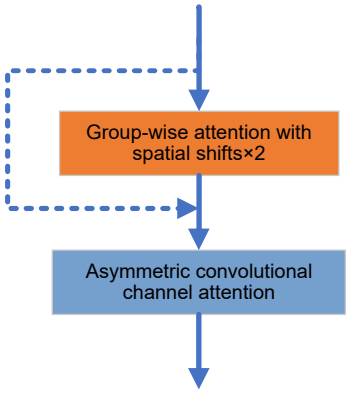


图 2 综合局部注意力机制的结构图

Fig. 2 Structure diagram of integrated local attention mechanism

2.2.1 空间分组注意力机制

针对屏幕内容图像的特点，空间分组注意力机制结合了空间平移和通道分组注意力的思想，旨在捕捉屏幕内容图像中不同对象所在位置的特征。屏幕内容图像通常包含大量的细节信息，例如文本、图标和复杂的几何形状，这些信息之间存在显著的空间关系。空间分组注意力机制能够根据屏幕内容图像中不同位置特征之间的相关性，动态调整注意力权重，从而突出关键区域的特征描述能力。这一机制不仅能够有效突出屏幕内容图像中显著区域的重要性，还能够增强模型对细节与整体布局的理解能力，从而显著提高屏幕内容图像质量评价算法的性能，其结构如图 3 所示。

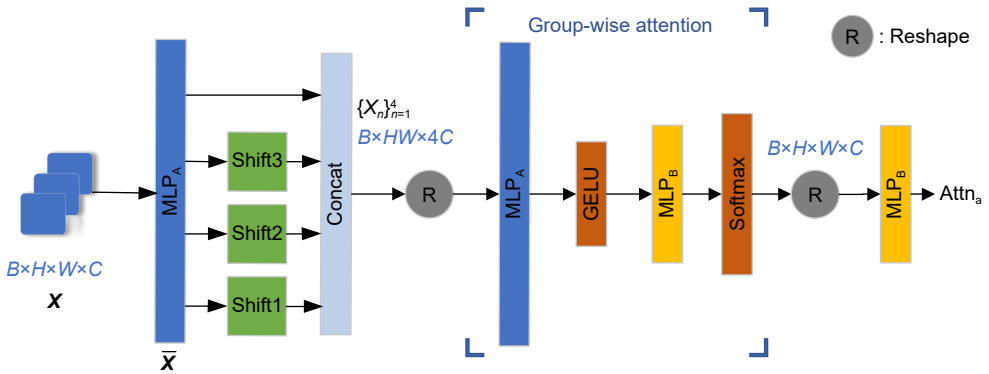


图 3 空间分组注意力机制的结构

Fig. 3 Structure of group-wise attention mechanism with spatial shifts

具体来说，首先将输入张量$\boldsymbol{X}\in\mathbb{R}^{W\times H\times C}$的通道数扩大到原来的四倍，得到$\overline{\boldsymbol{X}}$，如式 (4) 所示。

$$\overline{\boldsymbol{X}}=\mathrm{MLP_A}(\boldsymbol{X})\in\mathbb{R}^{W\times H\times 4C}, \tag{4}$$

式中：W 表示特征图的宽度；H 表示特征图的高度；C 表示通道的数量；$\mathrm{MLP_A}$ 表示全连接层 (multi-layer perceptron, MLP)；$\overline{\boldsymbol{X}}$表示经过处理后得到的新张量，其通道数是原始输入张量的四倍。

然后按照通道的数量，将张量$\overline{\boldsymbol{X}}$等分为四个子集$X_n$($n$=1、2、3、4)，每个子集表示一个分组，不同的分组分别执行不同的空间平移操作，并将空间平移操作后得到的四个子集进行拼接和形状调整，如表 2 所示。

表 2　基于分组的空间平移操作

Table 2　Group-based spatial shift operation

$\boldsymbol{X}_n$	Spatial shift
n=1	Shift1：move tensor x_1 down by one pixel vertically and right by one pixel horizontally
n=2	Shift2：move tensor x_2 down by two pixels vertically and right by two pixels horizontally
n=3	Shift3：move tensor x_3 right by one pixel horizontally and down by one pixel vertically
n=4	Without any processing

接着，计算分组注意力 (group-wise attention, GA)，依次使用两种不同的多层感知机 ($\mathrm{MLP_A}$、$\mathrm{MLP_B}$) 和高斯误差激活函数 (Gaussian error linear unit, GELU) 对输入张量进行处理，从而计算出注意力权重，如式 (5) 所示。

$$\mathrm{Attn_a}=\mathrm{MLP_B}(\mathrm{GA}(\{\boldsymbol{X}_n\}_{n=1}^{4})), \tag{5}$$

式中：$\mathrm{Attn_a}$ 表示空间分组注意力的数值；$\mathrm{MLP_B}$ 表示全连接层；GA 表示分组注意力的计算。

最后，对注意力权重进行指数归一化，并将其应用到每个分组的张量上。

本文使用了空间平移操作，使得模型能够在局部区域内捕捉到更广泛的空间信息，有助于增强模型对图像局部结构的表征能力。使用空间分组注意力机制能够有效地将输入特征分成多个组，增强了特征之间的交互性，进而提高模型对图像空间结构的表征能力。

2.2.2 非对称卷积通道注意力机制

为了提取屏幕内容图像中不同尺度的局部特征，非对称卷积通道注意力机制通过使用多个不同尺寸的卷积核，并结合 same 卷积方式，能够有效提取具有不同视野的特征信息。屏幕内容图像包含大量的文本、图标和线条等元素，这些元素往往有明确的方向性特征和边缘特征。非对称卷积[28] 通过在多个方向上对不同区域范围的特征进行建模，能够更好地捕捉这些方向性和边缘特征，而且有助于降低模型的复杂度。非对称卷积操作包括一个具有较大尺寸卷积核的初始卷积层 (Conv0)、多个非对称卷积层 (Conv0_1、Conv0_2、Conv1_1、Conv1_2、Conv2_1、Conv2_2) 和融合卷积层 (Conv3)，每个卷积核的尺寸和填充方式如表 3 所示。非对称卷积通道注意力机制的结构如图 4 所示。在此机制中，首先使用不同类型的卷积操作来提取特征，得到具有不同感受野的特征图。这些特征图捕捉了输入特征图中不同尺度的局部特征。接着，将融合后的注意力权重图与输入特征图进行逐元素点积运算。该机制对不同通道的特征赋予不同的权重，来选择性地增强或减弱不同通道特征重要性，其数学表达式为

$$\mathrm{Attn_b}=\mathrm{Concat}(\mathrm{attn_0}+\mathrm{attn_1}+\mathrm{attn_2}), \tag{6}$$

$$\mathit{Attention\_Map}=H_{\mathrm{Conv3}}(\mathrm{Attn_b}), \tag{7}$$

$$\boldsymbol{X}_1=\mathit{Attention\_Map}\otimes\boldsymbol{X}_0, \tag{8}$$

式中：$\boldsymbol{X}_0$ 表示输入特征图；$\boldsymbol{X}_1$ 表示输出特征图；H_{Conv3} 表示逐点卷积，它的作用主要是融合特征并调整通道数，为后续的加权操作提供适配的注意力权重；attn_i (i=0,1,2) 表示使用不同尺寸卷积核构成的局部注意力分支输出的特征；$\mathrm{Attn_b}$ 表示中间的注意力特征；*Attention_Map* 表示使用非对称卷积通道注意力机制生成的注意力权重。

表 3　卷积核的尺寸以及填充方式

Table 3　Convolutional kernel size and padding method

Convolution layer	Kernel size	Padding size
Conv0	5×5	2×2
Conv0_1	1×5	0×2
Conv0_2	5×1	2×0
Conv1_1	1×13	0×6
Conv1_2	13×1	6×0
Conv2_1	1×19	9×0
Conv2_2	19×1	0×9
Conv3	1×1	1×1

总的来说，非对称卷积通道注意力机制的主要作用是增强特征图的表征能力，它能够自动学习不同尺度的特征信息，从而更好地捕捉输入特征图中的关键信息，并且减少包含背景纹理噪声或冗余信息的通道

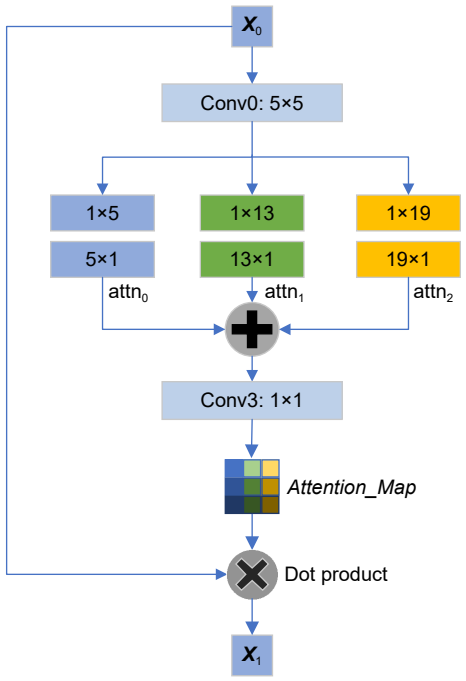


图 4 非对称卷积通道注意力机制的结构

Fig. 4 Structure of asymmetric convolutional channel attention mechanism

产生的影响。

2.3 双通道特征映射模块

在屏幕内容图像质量评价的过程中，每个图像块具有不同的重要性。因此，本文提出了双通道的特征映射模块，它使用了符合人眼视觉感知特点的自适应加权融合方法，其结构如图 5 所示。该模块包含两个通道，在第一个通道中，建立图像块的特征和主观质量评价分数之间的映射模型，得到图像块的质量分数；在第二个通道中，得到图像块特征对应的显著性权重。

在图 5 中，首先对输入张量进行形状的调整，将形状从 (B, C, H, W) 改变为 $(B, H\times W, C)$。这样处理的目的是将空间维度转换为特征维度，以便后续的 MLP 处理。特征映射模块的两个分支都包含 MLP 前馈神经网络，又分别使用了 ReLU 和 Sigmoid 这两个非线性激活函数。此外，在特征映射模块中，MLP 的前馈神经网络由两个线性层和指数线性单元 (exponential linear unit, ELU) 组成。

在图 5 中，对于每个输入图像块的特征 x_i，首先使用两个并行网络分别计算该图像块的质量分数 f_i 和显著性权重 ω_i；然后将此图像块的质量分数 f_i 与显著性权重 ω_i 相乘，并对所有图像块的乘积结果进行求和；最后除以权重的总和，从而获得失真图像的质量分数 Q_{A}，表达式为

$$Q_{\mathrm{A}}=\frac{\sum_{i=1}^{N} f_i\cdot\omega_i}{\sum_{i=1}^{N}\omega_i}, \tag{9}$$

式中：N 表示屏幕内容图像中图像块的数量。

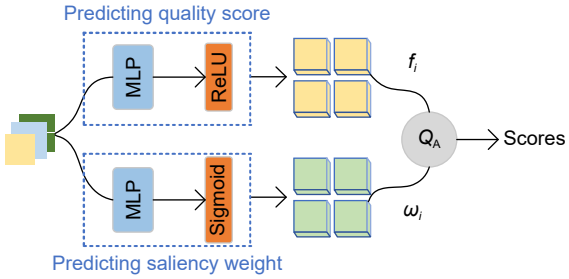


图 5 双通道特征映射模块的结构

Fig. 5 Structure of dual-channel feature mapping module

3 实验结果与分析

3.1 数据集

近年来，研究人员花费了大量人力创建屏幕内容图像数据集，例如 SCID[8]、SIQAD[5] 等。这两个数据集的基本特点如表 4 所示。在创建屏幕内容图像数据集时，为了减少主观因素带来的影响，使用了相同的显示设备、统一的环境光照等标准化条件。在获得屏幕内容图像的主观质量评价分数时，首先收集大量观察者的评分值，然后采用平均值或中位数等统计方法得到更为恰当的主观质量评价分数。SCID 数据集和 SIQAD 数据集分别采用平均主观分数 (mean opinion score, MOS) 和差分平均主观分数 (differential mean opinion score, DMOS) 作为主观质量评价值。当 MOS 数值越小时或 DMOS 数值越大时，屏幕内容图像的失真越严重，屏幕内容图像的质量越差；反之，屏幕内容图像的失真越轻微，则屏幕内容图像的质量越好。

表 4 常用的屏幕内容图像数据集

Table 4 Commonly used screen content image datasets

Dataset	Number of reference	Number of distortion	Distortion types count	Distortion levels count	Subjective score type
SCID	40	1800	9	5	MOS
SIQAD	20	980	7	7	DMOS

SCID 数据集包含 40 张参考图像和 1800 张失真图像。在 SIQAD 数据集包含的失真类型的基础上，SCID 数据集去掉了对比度失真这种失真类型，并增加了颜色饱和度变化 (color saturation change, CSC)、高效视频编码的屏幕内容图像压缩 (HEVC screen content compression, HEVC-SCC) 和抖动的颜色量化这三种失真类型。

SIQAD 数据集包括 20 张参考图像和 980 张失真图像，其宽度和高度大约为 600~900 个 pixel，这些图像来源于网页、幻灯片和电子杂志等。该数据集包含 7 种不同的失真类型：高斯噪声、高斯模糊、运动模糊、对比度失真、JPEG 压缩、JPGE2000 压缩和分层式压缩。

虽然上述失真类型具有全局性的特点，但它们仍然会在局部区域表现出特定特征。例如，运动模糊可能会产生在特定方向的模糊，JPEG 压缩可能会在边缘或细节处引入伪影。因此，本文提出的 MTA-SCI 提取了屏幕内容图像的局部特征，能够更好地评价屏幕内容图像的质量。

3.2 评价指标

如果屏幕内容图像质量评价算法性能越好，那么它的质量评测分数会与主观质量分数保持高度一致。国际上屏幕内容图像质量评价算法通用的三个性能评价指标如下所示。

皮尔森线性相关系数 (Pearson linear correlation coefficient, PLCC) 用于评估屏幕内容图像质量评价算法的预测准确性，其取值范围为-1~1。PLCC 的数学表达式为

$$PLCC=\frac{\sum_{i=1}^{N}(a_i-\bar{a})\left(b_i-\bar{b}\right)}{\sqrt{\sum_{i=1}^{N}(a_i-\bar{a})\times\sum_{i=1}^{N}\left(b_i-\bar{b}\right)}}, \tag{10}$$

式中：a_i 表示屏幕内容图像质量评价算法得到的客观质量评价分数；b_i 表示人工统计获得的主观质量评价分数；$\bar{a}$和$\bar{b}$分别表示客观质量评价分数和主观质量评价分数的平均值。

斯皮尔曼秩序相关系数 (Spearman rank order correlation coefficient, SRCC) 用于评价客观质量分数和主观质量分数之间的单调性，其取值范围为-1~1。SRCC 的数学表达式为

$$SRCC=1-\frac{6\sum_{i=1}^{N}(m_i-n_i)^2}{N(N^2-1)}, \tag{11}$$

式中：m_i 表示 a_i 的排序；n_i 表示 b_i 的排序。

均方根误差 (Root mean squared error, RMSE) 用于计算客观质量分数与主观质量分数之间的绝对误差。RMSE 的数学表达式为

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(a_i-b_i)^2}{N}}. \tag{12}$$

如果 SRCC、PLCC 的取值越靠近 1，同时 RMSE 的取值越小时，则屏幕内容图像客观质量评价算法的性能越好。

3.3 实验环境配置及其参数

实验环境配置和参数如表 5 所示，实验使用 Python 3.7.0 和 Pytorch-GPU 1.13.1。此外，实验还使用了具有 64.0 GB 内存的 Intel Core i7-13700 CPU 和具有 24.0 GB 内存的 NVIDIA RTX 4090 GPU。在实验中，输入图像的尺寸设置为 224×224，训练集和验证集的比例设置为 8∶2，学习率的初始值设置为 1×10^{-5}，并使用余弦退火优化器动态调整学习率，同时使用 RMSE 作为损失函数。本文提出的 MTA-SCI 网络模型训练了 200 轮，每轮大约需要 90 s。

表 5 实验的环境配置及其参数

Table 5 Environmental configuration and parameters of the experiment

Parameter	Value
Param count	307.94577 M
Compilation Environment	Python 3.7.0, Pytorch-GPU 1.13.1, and CUDA 11.3
CPU model	Intel Core i7-13700
GPU model	NVIDIA RTX 4090
Average time/epoch	90 s

为了解决小样本的问题，本文在训练过程中使用了动态随机裁剪方法进行数据增强。该方法没有增加原始数据集的样本数量，但是在每次训练过程中，把图像不同的裁剪区域输入到 MTA-SCI 模型中，从而增大了数据集的规模，并且增强了 MTA-SCI 模型的泛化能力。此外，本文采用了在 ImageNet-21k 数据集上已完成预训练的 ViT-B/8 模型作为基准模型，该模型已经学到了非常丰富的视觉特征。所以，在屏幕内容图像数据集具有较少数据的情况下，也能够对本文提出的 MTA-SCI 网络模型进行充分的训练。

3.4 实验性能对比

本文分别在 SCID 数据集、SIQAD 数据集上验证 MTA-SCI 的具体性能，如表 6 所示。在表 6 中，

使用粗体标注了无参考屏幕内容图像质量评价算法中最优的评价指标。从表 6 可以看出，本文提出的 MTA-SCI 在 SCID 数据集、SIQAD 数据集中均有良好的表现。

在 SCID 数据集上，与近 5 年主流的无参考屏幕内容图像质量评价算法相比，MTA-SCI 具有最优的性能，此算法的 PLCC 和 SRCC 分别高达 0.9602、0.9609。与全参考屏幕内容图像质量评价算法相比，此算法的 PLCC 和 SRCC 仅仅稍微低于 DAMC 算法，同时这两个指标明显大于 ESIM 算法和 SR-CNN 算法。MTA-SCI 与基于手工提取特征的全参考 ESIM 方法相比，其 PLCC 和 SRCC 分别提高了 11.34% 和 13.26%。

在 SIQAD 数据集上，MTA-SCI 的 PLCC 和 SRCC 分别为 0.9233、0.9294。与近 5 年主流的无参考屏幕内容图像质量评价算法相比，MTA-SCI 的 PLCC 值排名第一，同时此算法的 SRCC 值与 Zhang 等[16] 的 SRCC 值只有很小的差距，这表明 MTA-SCI 具有很强的竞争力。MTA-SCI 与基于手工提取特征的 ESIM 方法相比，其 PLCC 和 SRCC 分别提高了 6.96% 和 5.76%。

在 SCID 数据集和 SIQAD 数据集上，MTA-SCI 的 PLCC、SRCC 和损失值的变化曲线如图 6、图 7 所示。从这两张图可以看出，随着训练轮次的增加，PLCC 和 SRCC 先逐渐增加后趋于稳定。该现象表示本文提出的 MTA-SCI 不仅预测值与主观评价得分的线性相关性逐渐增强，而且预测值的顺序与主观评价

表 6　各类屏幕图像质量评价算法的性能对比

Table 6　Performance comparison of various screen content image quality assessment algorithms

Type	Method	SCID		SIQAD	
		SRCC	PLCC	SRCC	PLCC
FR	MIC-CNN[13]	-	-	0.9636	0.9669
	ESIM[9]	0.8478	0.8630	0.8632	0.8788
	DAMC[21]	0.9617	0.9617	0.9304	0.9373
	SR-CNN[10]	0.9400	0.9390	0.8943	0.9042
NR	Yang et al.[8]	0.7562	0.7867	0.8543	0.8738
	QODCNN[12]	0.8760	0.8820	0.8890	0.9010
	RIQA[14]	-		0.9000	0.9110
	Zhang et al.[16]	0.9050	0.9133	**0.9242**	0.9260
	BLIQUP-SCI[7]	-	-	0.7990	0.7705
	Yang et al.[8]	0.7562	0.7867	0.8543	0.8738
	SIQA-DF-II[11]	-	-	0.8880	0.9000
	Gao et al.[15]	0.8569	0.8613	0.8962	0.9000
	MTDL[17]	-	-	0.9233	0.9248
	DFSS-IQA[29]	0.8146	0.8138	0.8820	0.8818
	Zhang[30]	0.9445	0.9433	0.8640	0.8889
	MTA-SCI	**0.9602**	**0.9609**	0.9233	**0.9294**

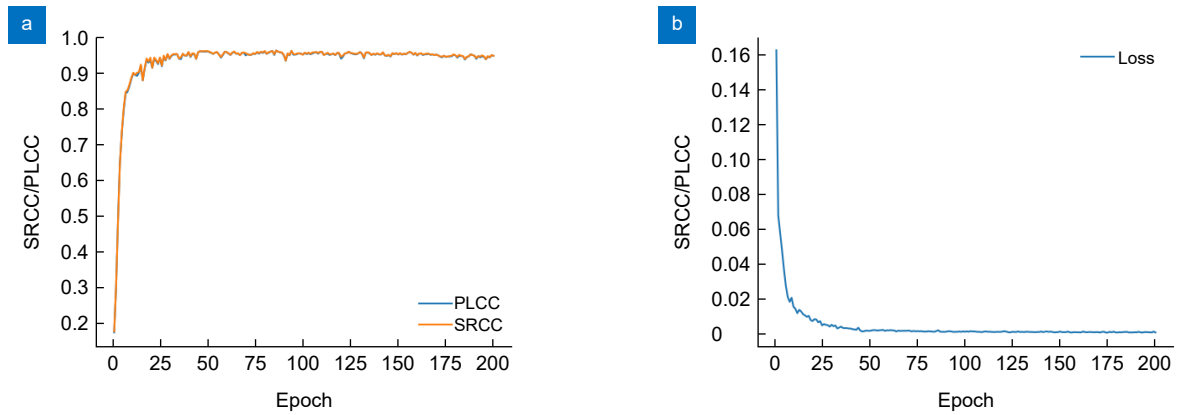


图 6　在 SCID 数据集上得到的 PLCC、SRCC 和损失值的变化曲线。(a) PLCC、SRCC 的变化曲线；(b) 损失值的变化曲线

Fig. 6　Variation curves of PLCC, SRCC, and loss values obtained on the SCID dataset. (a) Variation curves of PLCC and SRCC; (b) Variation curve of the loss value

得分的顺序具有较强的一致性。此外，从这两个图也可以看出，损失值随着训练的进行而逐渐下降，这表明 MTA-SCI 的预测能力在逐渐增强。

3.5 算法预测结果

图 8 展示了来自 SCID 验证集的参考图像和经过不同程度失真的图像。其中，图 8(a) 是一张未经过压缩的参考图像，它没有任何程度的失真。图 8(b)、8(c)、8(d) 分别展示了对图 8(a) 进行不同程度 JPEG 压缩后的失真图像，其失真等级分别为 3、4、5，这表示压缩程度逐渐增加。随着失真等级的增加，图像的质量逐渐下降，图像中的细节和清晰度受到影响，失真效果更加明显。

使用本文提出的 MTA-SCI 对图 8 的失真图像进

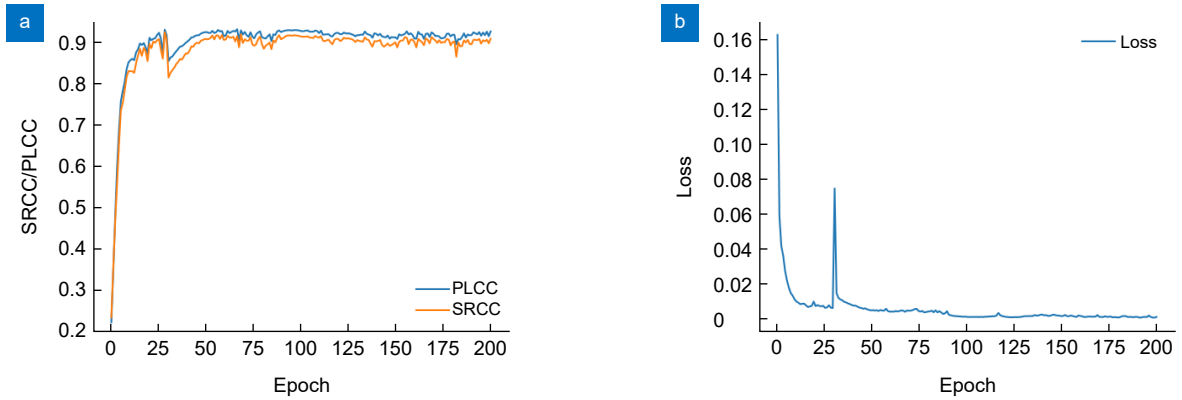
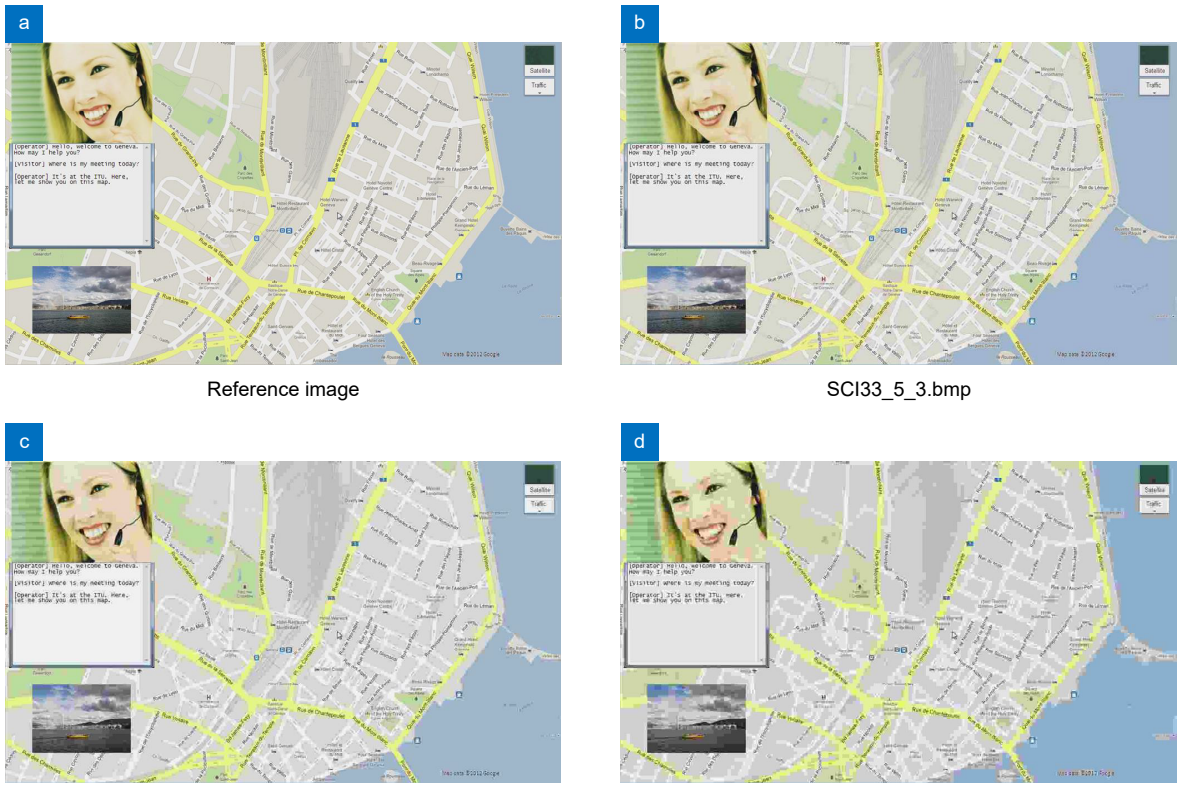


图 7 在 SIQAD 数据集上得到的 PLCC、SRCC 和损失值的变化曲线。(a) PLCC、SRCC 的变化曲线；(b) 损失值的变化曲线

Fig. 7 Variation curves of PLCC, SRCC, and loss values obtained on the SIQAD dataset. (a) Variation curves of PLCC and SRCC; (b) Variation curve of the loss value



Reference image　　SCI33_5_3.bmp

SCI33_5_4.bmp　　SCI33_5_5.bmp

图 8 失真的屏幕内容图像实例。(a) 参考图像；(b) SCI33_5_3.bmp；(c) SCI33_5_4.bmp；(d) SCI33_5_5.bmp

Fig. 8 Distorted screen content image. (a) Reference image; (b) SCI33_5_3.bmp; (c) SCI33_5_4.bmp; (d) SCI33_5_5.bmp

行质量评价，失真图像的质量分数如表 7 所示。从表 7 可以看出，待测图像的预测值与其 MOS 值在排序上具有一致性，这意味着本文提出的 MTA-SCI 能够有效地捕捉图像的失真程度，并与人类主观质量评价保持良好的一致性。此外，预测值与归一化后的 MOS 值之间也表现出了高度相关性，进一步验证了本文提出的 MTA-SCI 处理无参考图像质量评估任务中的准确性和鲁棒性。总之，MTA-SCI 不仅能够在不同程度的失真情况下准确评估图像的质量变化，还能较好地预测人眼主观感知的图像质量。这种客观评价与主观评价之间的一致性使得 MTA-SCI 在实际应用中具有较高的实用价值。

3.6 消融实验

为了验证本文提出的 MTA-SCI 的有效性，本文在 SCID 数据集上进行了两部分的消融实验。

在第一部分的消融实验中，本文通过实验来确定空间分组注意力机制最优的分组数量和非对称卷积注意力机制中非对称卷积核大小组合，从而使本文提出的 MTA-SCI 具有最佳的预测性能。不同的分组数量得到的实验结果如表 8 所示。从表 8 可以看出，当分组数量 k=4 时，算法效果更好。

表 9 展示了 6 组卷积核组合在 PLCC 和 SRCC 指标上的表现。从表 9 可以看出，*Kernal*=5、13、19 组合的性能最优，验证了适度的卷积核大小对 MTA-SCI 的正向作用。同时，过大的卷积核组合 (如 *Kernal*=15、23、27) 可能因感受野过大而忽略局部细节，导致算法性能下降。

此外，在第 2 部分的消融实验中，验证了本文提出的综合局部注意力机制 (integrated local attention mechanism, ILAM) 与双通道映射模块 (dual-channel feature mapping module, DFM) 的有效性，消融实验的结果如表 10 所示。在第 1 个实验中，使用了基准模型得到的 PLCC 和 SRCC 分别为 0.8792、0.8651。在第 2 个实验中，只使用了双通道映射模块，得到的 PLCC 和 SRCC 分别为 0.8832、0.8796。与基准算法相比 PLCC 和 SRCC 分别提高了 0.5% 和 1.7%。由于双通道映射模块综合考虑了图像块的特征及其在整体图像中的显著性，此消融实验结果验证了此模块能有

表 7　失真图像的预测分数

Table 7　Predicted scores of the distorted screen content images

Image ID	Prediction value	MOS	Normalized MOS
SCI33_5_3.bmp	0.1067	36.7569	0.2711
SCI33_5_4.bmp	0.0653	25.4741	0.0619
SCI33_5_5.bmp	0.0658	25.6376	0.0650

表 8　不同的分组数量对 MTA-SCI 性能的影响

Table 8　Impact of different numbers of groups on the MTA-SCI performance

Group count	PLCC	SRCC
k=2	0.9471	0.9520
k=3	0.9591	0.9587
k=4	0.9602	0.9609
k=5	0.9572	0.9487

表 9　不同非对称卷积核组合对 MTA-SCI 性能产生的影响

Table 9　Impact of different asymmetric convolution kernel combinations on the performance of the MTA-SCI

Kernel combination	PLCC	SRCC
Kernal=3, 15, 19	0.9568	0.9585
Kernal=5, 7, 9	0.9569	0.9584
***Kernal*=5, 13, 19**	**0.9602**	**0.9609**
Kernal=7, 11, 21	0.9575	0.9563
Kernal=9, 17, 25	0.9581	0.9589
Kernal=15, 23, 27	0.8967	0.9005

表 10 综合局部注意力机制、双通道特征映射模块和残差连接对模型性能的影响

Table 10 Impact of ILAM, DFM, and residual connection on the algorithm performance

No.	ILAM	DFM	RC	PLCC	SRCC
1	×	×	×	0.8792	0.8651
2	×	√	×	0.8832	0.8796
3	√	×	×	0.9481	0.9508
4	√	√	×	0.9571	0.9592
5	√	√	√	**0.9602**	**0.9609**

效地提高了对屏幕内容图像质量的评估精度。在第 3 个实验中，只使用综合局部注意力机制，得到的 PLCC 和 SRCC 分别为 0.9481、0.9508。与基准算法相比，PLCC 和 SRCC 分别提高了 7.8% 和 9.9%。由于综合局部注意力机制能够增强屏幕内容图像局部信息的表征能力，此消融实验结果验证了综合局部注意力机制能提高算法性能。在第 4 个实验中，同时使用综合局部注意力机制和双通道映射模块，得到的 PLCC 和 SRCC 分别为 0.9571、0.9592。与基准算法相比，PLCC 和 SRCC 分别提高了 8.9% 和 10.9%。在第 5 个实验中，在基准网络的基础上同时使用综合局部注意力机制、双通道映射模块和残差连接 (residual connection, RC) 得到本文提出的 MTA-SCI 模型，此模型得到的 PLCC 和 SRCC 比基准网络分别提高了 9.2% 和 11.1%。

4 结 论

基于人眼多级视觉感知特性，本文提出了一种基于多任务注意力机制的无参考屏幕内容图像质量评价算法。该算法提出了综合局部注意力机制，它由空间分组注意力机制和非对称卷积注意力机制组成。一方面，综合局部注意力机制中的空间分组注意力机制使得本文提出的 MTA-SCI 能够在局部区域内捕捉到更广泛的空间信息；另一方面，此机制中的非对称卷积注意力机制使模型更专注于屏幕内容图像中多尺度特征的学习，并减少了包含背景纹理噪声的通道对评价过程产生的影响。此外，MTA-SCI 利用双通道的特征映射模块，使用符合人眼视觉感知特点的自适应加权融合方法。实验结果表明，MTA-SCI 在预测性能上优于现有算法。MTA-SCI 还存在以下两个局限：1) 该算法使用了 ViT 模型，此模型对内存和计算能力的需求较高；2) 在屏幕内容图像局部特征和全局特征的融合方法上，需要进一步研究更加合理的融合机制。

屏幕内容图像质量评价未来的研究工作可从两方面入手：其一，结合 ViT 与 CNN 的优势[31]，增强全局特征与局部特征的综合建模能力；其二，把手工提取特征与深度学习特征相结合的方法也是以后非常重要的研究方向。这种方法既能避免图像拉伸带来的失真，又保留不同特征提取方法的互补优势。这些方向将进一步推动屏幕内容图像质量评价技术的发展。

利益冲突: 所有作者声明无利益冲突

参考文献

[1] Nizami I F, Rehman M U, Majid M, et al. Natural scene statistics model independent no-reference image quality assessment using patch based discrete cosine transform[J]. *Multimed Tools Appl*, 2020, **79**(35): 26285−26304.

[2] Yang J C, Bian Z L, Zhao Y, et al. Full-reference quality assessment for screen content images based on the concept of global-guidance and local-adjustment[J]. *IEEE Trans Broadcast*, 2021, **67**(3): 696−709.

[3] Wang B, Bai Y Q, Zhu Z J, et al. No-reference light field image quality assessment based on joint spatial-angular information[J]. *Opto-Electron Eng*, 2024, **51**(9): 69−81.
王斌, 白永强, 朱仲杰, 等. 联合空角信息的无参考光场图像质量评价[J]. 光电工程, 2024, **51**(9): 69−81.

[4] Bai Y Q, Zhu Z J, Zhu C H, et al. Blind image quality assessment of screen content images via fisher vector coding[J]. *IEEE Access*, 2022, **10**: 13174−13181.

[5] Yan J B, Fang Y M, Liu X L. The review of distortion-related image quality assessment[J]. *J Image Graphics*, 2022, **27**(5): 1430−1466.
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]. 中国图像图形学报, 2022, **27**(5): 1430−1466.

[6] Yang H, Fang Y M, Lin W S. Perceptual quality assessment of screen content images[J]. *IEEE Trans Image Process*, 2015, **24**(11): 4408−4421.

[7] Shao F, Gao Y, Li F C, et al. Toward a blind quality predictor for screen content images[J]. *IEEE Trans Syst Man Cybern Syst*, 2018, **48**(9): 1521−1530.

[8] Yang J C, Zhao Y, Liu J C, et al. No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions[J]. *IEEE Trans Cybern*, 2022, **52**(5): 2798−2810.

[9] Ni Z K, Ma L, Zeng H Q, et al. ESIM: edge similarity for screen content image quality assessment[J]. *IEEE Trans Image Process*, 2017, **26**(10): 4818−4831.

[10] Chen C L Z, Zhao H M, Yang H, et al. Full-reference screen

content image quality assessment by fusing multilevel structure similarity[J]. *ACM Trans Multimedia Comput Commun Appl*, 2021, **17**(3): 1−21.

[11] Jiang X H, Shen L Q, Ding Q, et al. Screen content image quality assessment based on convolutional neural networks[J]. *J Vis Commun Image Represent*, 2020, **67**: 102745.

[12] Jiang X H, Shen L Q, Feng G R, et al. An optimized CNN-based quality assessment model for screen content image[J]. *Signal Process Image Commun*, 2021, **94**: 116181.

[13] Zuo L X, Wang H L, Fu J. Screen content image quality assessment via convolutional neural network[C]//*23rd IEEE International Conference on Image Processing*, 2016: 2082–2086. https://doi.org/10.1109/ICIP.2016.7532725.

[14] Jiang X H, Shen L Q, Yu L W, et al. No-reference screen content image quality assessment based on multi-region features[J]. *Neurocomputing*, 2020, **386**: 30−41.

[15] Gao R, Huang Z Q, Liu S G. Multi-task deep learning for no-reference screen content image quality assessment[C]//*27th International Conference on MultiMedia Modeling*, 2021: 213–226. https://doi.org/10.1007/978-3-030-67832-6_18.

[16] Zhang C F, Huang Z Q, Liu S G, et al. Dual-channel multi-task CNN for no-reference screen content image quality assessment[J]. *IEEE Trans Circuits Syst Video Technol*, 2022, **32**(8): 5011−5025.

[17] Yang J C, Bian Z L, Zhao Y, et al. Staged-learning: assessing the quality of screen content images from distortion information[J]. *IEEE Signal Process Lett*, 2021, **28**: 1480−1484.

[18] Pan L L, Shao J F. Multi-resolution point cloud completion fusing graph attention[J]. *Laser Technol*, 2023, **47**(5): 700−707.
潘李琳, 邵剑飞. 融合图注意力的多分辨率点云补全[J]. 激光技术, 2023, **47**(5): 700−707.

[19] Chang Y L, Li S M, Liu A Q, et al. Quality assessment of screen content images based on multi-stage dictionary learning[J]. *J Vis Commun Image Represent*, 2021, **79**: 103248.

[20] Huang Z Q, Liu S G. Perceptual hashing with visual content understanding for reduced-reference screen content image quality assessment[J]. *IEEE Trans Circuits Syst Video Technol*, 2021, **31**(7): 2808−2823.

[21] Yao Y, Hu J T, Yang W M, et al. Distortion-aware mutual constraint for screen content image quality assessment[C]//*12th International Conference on Image and Graphics*, 2023: 403–414. https://doi.org/10.1007/978-3-031-46305-1_33.

[22] Rehman M U, Nizami I F, Majid M. DeepRPN-BIQA: deep architectures with region proposal network for natural-scene and screen-content blind image quality assessment[J]. *Displays*, 2022, **71**: 102101.

[23] Min X K, Gu K, Zhai G T, et al. Screen content quality assessment: overview, benchmark, and beyond[J]. *ACM Comput Surv*, 2022, **54**(9): 187.

[24] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//*9th International Conference on Learning Representations*, 2021.

[25] Ren L H, Huang L W, Tian X, et al. Multivariate long-term series forecasting method with DFT-based frequency-sensitive dual-branch transformer[J]. *J Comput Appl*, 2024, **44**(9): 2739−2746.
任烈弘, 黄铝文, 田旭, 等. 基于 DFT 的频率敏感双分支 Transformer 多变量长时间序列预测方法[J]. 计算机应用, 2024, **44**(9): 2739−2746.

[26] Chen L, Zhang H W, Xiao J, et al. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6298–6306. https://doi.org/10.1109/CVPR.2017.667.

[27] Yu L L, Zhang X B, Wang K. CMAAC: combining multiattention and asymmetric convolution global learning framework for hyperspectral image classification[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5508518.

[28] Ding X H, Guo Y C, Ding G G, et al. ACNet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks[C]//*17th IEEE/CVF International Conference on Computer Vision*, 2019: 1911–1920. https://doi.org/10.1109/ICCV.2019.00200.

[29] Chen B L, Zhu H W, Zhu L Y, et al. Deep feature statistics mapping for generalized screen content image quality assessment[J]. *IEEE Trans Image Process*, 2024, **33**: 3227−3241.

[30] Zhang W. No-reference quality assessment method for screen content image based on multi-scale convolutional neural network[J]. *J Liaoning Univ Technol Nat Sci Ed*, 2024, **44**(5): 286−291.
张巍. 基于多尺度卷积神经网络屏幕内容图像无参考质量评价方法[J]. 辽宁工业大学学报 (自然科学版), 2024, **44**(5): 286−291.

[31] Guo J L, Zhi M, Yin Y J, et al. Review of research on CNN and visual Transformer hybrid models in image processing[J]. *J Front Comput Sci Technol*, 2025, **19**(1): 30−44.
郭佳霖, 智敏, 殷雁君, 等. 图像处理中 CNN 与视觉 Transformer 混合模型研究综述[J]. 计算机科学与探索, 2025, **19**(1): 30−44.

作者简介

周子镱 (2000-)，女，硕士研究生，主要研究方向为屏幕内容图像质量评价。

E-mail: 646750685@qq.com

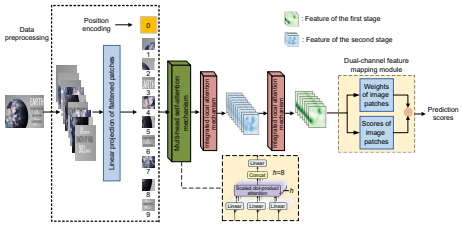
【通信作者】董武 (1980-)，男，博士，副教授，硕士生导师，主要研究方向为深度学习与人工智能、图像处理与机器视觉、数字印刷与数字出版等。

E-mail: dongwu@bigc.edu.cn

扫描二维码，获取PDF全文

Multi-task attention mechanism based no reference quality assessment algorithm for screen content images

Zhou Ziyi, Dong Wu*, Lu Likun, Ma Qian, Hou Guopeng, Zhang Erqing



Network structure of MTA-SCI

Overview: The previous screen content image quality assessment algorithms failed to fully consider the multi-level visual perception characteristics of the human eye. To address this limitation, we propose a multi-task attention mechanism-based no-reference quality assessment algorithm for screen content images (MTA-SCI), which better simulates human visual perception. The MTA-SCI combined the advantages of both global and local features of SCIs, enabling it to capture the overall structure while focusing on visually significant details. This approach significantly enhanced the SCI quality evaluation capability. Specifically, the MTA-SCI employed a self-attention mechanism to extract global features, improving the representation of overall information in SCIs. Subsequently, it utilized an integrated local attention mechanism to extract local features, allowing the algorithm to focus on more salient and attention-grabbing details in the images and suppressing channels containing background texture noise, reducing the impact of background texture noise on image quality assessment. The integrated local attention mechanism consists of the group-wise attention mechanism with spatial shifts and asymmetric convolutional channel attention mechanism. In the MTA-SCI algorithm, they perform different tasks, working together to improve the performance of screen content image quality assessment. Finally, a dual-channel feature mapping module is adopted to predict SCI quality scores. In the first channel, it predicted the quality score of image patches; in the second channel, it predicted the saliency weights of the image patches. The dual-channel feature mapping module effectively quantifies the importance of different image patches within the overall image, making the predictions more aligned with subjective human assessments. Experiments on the SCID dataset demonstrate that the proposed MTA-SCI achieves a Spearman's rank-order correlation coefficient (SROCC) of 0.9563 and a Pearson linear correlation coefficient (PLCC) of 0.9575. On the SIQAD dataset, it achieves an SROCC of 0.9274 and a PLCC of 0.9171. Overall, the multi-task attention mechanism consists of three components: multi-head self-attention mechanism, group-wise attention mechanism with spatial shifts, and asymmetric convolutional channel attention mechanism. These three attention mechanisms perform different tasks in the proposed MTA-SCI algorithm, working together to improve the performance of screen content image quality assessment. By integrating self-attention for global feature extraction, integrated local attention for detail refinement, and a dual-channel feature mapping module for prediction, MTA-SCI effectively captures the complex perceptual characteristics of the human visual system. The high performance achieved on benchmark datasets validates its accuracy and reliability, making it a promising solution for future applications in screen content image quality.

Zhou Z Y, Dong W, Lu L K, et al. Multi-task attention mechanism based no reference quality assessment algorithm for screen content images[J]. *Opto-Electron Eng*, 2025, **52**(4): 240309; DOI: 10.12086/oee.2025.240309

Foundation item: Beijing Digital Education Research Key Project (BDEC2022619027), Beijing Higher Education Society 2023 General Project (MS2023168), Beijing Institute of Graphic Communication University-level Scientific Research Projects (Ec202303, Ea202301, E6202405), Discipline Development and Graduate Education Special Fund of Beijing Institute of Graphic Communication (21090224002, 21090323009, 21090124013), Emerging Interdisciplinary Platform Construction Project for Publishing Studies of Beijing Institute of Graphic Communication (04190123001/003), and Open Research Fund Project of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications (SKLNST-2023-1-12)

Beijing Key Laboratory of Signal and Information Processing for High-end Printing Equipment, Beijing Institute of Graphic Communication, Beijing 102600, China

* E-mail: dongwu@bigc.edu.cn