

CSWin-Transformer与可形变卷积相结合的 图像修复技术研究与实践

刘海洋, 胡永

(西藏民族大学信息工程学院, 陕西 咸阳 712082)

摘要: 针对现有图像修复模型修复大面积不规则缺损图像效果不佳、计算资源消耗大的问题, 提出了一种CSWin-Transformer与可形变卷积残差密集网络相结合的图像修复方法。首先, 构建一个由全局层网络和局部层网络组成的生成模型, 利用全局层CSWin-Transformer模块的条纹窗口在较低的计算复杂度下获取更大的感受野, 增强其图像特征提取能力; 其次, 在CSWin-Transformer中加入一种新的门控深度卷积前馈网络, 其能够进行有选择性的特征转换, 即过滤掉信息量不足的特征, 仅保留有价值的信息继续在网络的层级结构中流动; 再次, 通过并行局部层的可形变卷积残差密集块灵活对图像进行采样, 增强结构纹理修复的精确度, 同时, 在上述并行生成模型之间, 构建共享的注意力机制来促进全局和局部特征之间的信息交流; 最终, 采用谱归一化的马尔科夫判别模型进行对抗性训练。实验结果表明, 提出的方法相较于其他方法在PSNR和SSIM指标上分别提升了2.47dB和0.075 2, 在LPIPS指标上下降了0.092 4。

关键词: 深度学习; CSWin-Transformer; 门控深度卷积前馈网络; 可形变卷积残差密集网络

DOI: 10.11907/rjtk.241952

扫描二维码阅读全文:

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2026)001-0119-08



Research and Implementation of Image Restoration Technology Combining CSWin-Transformer with Deformable Convolution

LIU Haiyang, HU Yong

(School of Information Engineering, Xizang Minzu University, Xianyang 712082, China)

Abstract: Addressing the issues of poor performance in repairing large-area irregular defects in images and high consumption of computational resources, this paper proposes an image repair method that combines CSWin-Transformer with a deformable convolution residual dense network. First, a generative model is constructed, consisting of a global layer network and a local layer network. The global layer CSWin-Transformer module uses striped windows to obtain a larger receptive field at a lower computational complexity, enhancing its ability to extract image features. After the CSWin-Transformer, a new gated-dconv feed-forward network is added, which can selectively transform features, filtering out features with insufficient information and retaining only valuable information to continue flowing through the network's hierarchical structure. Then, parallel local layers with deformable convolution residual dense blocks flexibly sample the image, enhancing the precision of structural texture repair. Between the aforementioned parallel generative models, a shared attention mechanism is constructed to facilitate the exchange of information between global and local features. Finally, a Markov discriminative model with spectral normalization is used for adversarial training. Experimental results show that the proposed method achieves an improvement of 2.47 dB in the PSNR index and 0.075 2 in the SSIM index compared with other methods, while there is a decrease of 0.092 4 in the LPIPS index.

Key Words: deep learning; CSWin-Transformer; gated-dconv feed-forward network; deformable convolution residual dense network

收稿日期: 2024-11-11

基金项目: 国家自然科学基金项目(62062061)

作者简介: 刘海洋(2000-), 男, 西藏民族大学信息工程学院硕士研究生, 研究方向为图像检测、图像修复; 胡永(1980-), 男, 硕士, 西藏民族大学信息工程学院副教授、硕士生导师, 研究方向为智能交通、图形图像处理。本文通讯作者: 胡永。

0 引言

近年来,许多研究者对图像修复技术进行了深入研究。在图像修复的过程中,其主要目标是补全图像中缺失的真实像素,这在物体的去除和图片的恢复上具有极高的实用价值。为了实现预期的效果,需要考虑以下2个要素:①相邻纹理之间的连贯性;②人类视觉上的合理性。如今流行的修复技术大致可分为2类:传统修复手段和基于深度学习的修复手段。

传统的图像修复方法主要依赖像素扩散算法^[1]和补丁匹配^[2]技术,以生成相似的纹理或色彩信息,从而补充图像中的缺失部分。传统的修复方法只适用于破损较小的图片,随着破损区域的增大,其修复效果会出现结构模糊和边界不明确的问题^[3]。

应用于图像修复的深度学习方法主要包括基于生成对抗网络(GAN)、自编码器(AE)以及卷积神经网络(CNN)。GAN由生成器和判别器组成,生成器的目标是从随机噪声中生成逼真的图像,而判别器则判断图像的真实性。通过对抗训练,生成器学习生成越来越逼真的图像。在图像修复中,GAN被用来自动检测和填补图像中的缺失部分,从而改善图像质量,降低人工修复的工作量。近年来,基于GAN的方法在图像修复领域占据主导地位,其性能优于传统CNN方法。AE是一种无监督学习模型,用于数据的降维、特征提取或数据去噪,其由编码器和解码器组成:编码器将输入数据压缩成低维的隐藏表示;解码器则将低维的隐藏表示还原回原始数据的近似值。AE可以用于图像去噪、异常检测等任务,但在生成新样本的能力上有限,不具备对输入数据的生成建模能力。CNN的不同架构和学习策略被用来从海量训练数据中提取重要信息,以获得更好的图像修复解决方案。一些具体的CNN架构,如全卷积网络(Fully Convolutional Network)和U-Net网络,已被广泛应用于图像修复任务中。虽然基于卷积神经网络(CNN)的图像修复方法在许多方面表现出色,但也存在一些缺点和局限性:①CNN在训练过程中容易陷入局部最优,导致生成图像质量不佳;②需要大量的训练数据,在数据稀缺的任务中,CNN的性能可能会受到限制;③CNN假设图像中的特征在空间上是不变的,这可能导致它无法很好地处理图像中的变形、旋转或缩放等情况。

尽管这些技术能够通过分析图像数据中的高级特征和上下文信息来实现更精确自然的修复效果,但仍需要大量参数和计算资源,特别是在处理高分辨率图像时。Li等^[4]提出了一种高分辨率图像修复方法,结合了多头自注意力和堆叠卷积技术。改进的Transformer块去除了层归一化和残差连接,以增强长距离交互和训练稳定性。使用Conv-U-Net细化细节,发挥CNN在局部纹理捕捉和效率上的优势,并设计风格操作模块以增加结果多样性。但该方法在语义理解和推理速度上仍有提升空间。Wan等^[5]提

出了一种双向transformer图像修复算法(ICT),模仿BERT^[6]的掩码语言模型和双向注意力,优化仅解码器的transformer。修复过程具体分为2步:首先,对有缺失区域的图像进行数据压缩和离散化,用改进的transformer采样低分辨率补全结果,作为粗糙外观先验;其次,利用上采样CNN模型在先验信息和已知像素指导下,对缺失区域进行高保真细节渲染,确保与周围像素一致。Zheng等^[7]将图像修复视为无方向序列—序列预测任务。粗修复阶段用限制性卷积模块提取token,并通过加权自注意力层的transformer生成初步图像;细修复阶段设计了自适应注意平衡层,分别计算已知和掩码区域的注意力,从可见区域传输高质量信息。这些设计有助于预测图像细节,但在高层语义理解和想象方面仍有不足,并且Transformer对计算机内存的需求是巨大的,这也限制了其应用范围和推广程度。

具体而言,当前的图像修复技术存在以下3个问题:①单阶段图像修复技术由于网络结构相对简单,在处理背景复杂或纹理细节丰富的图像时,往往会出现纹理错乱和形状失真的问题;②虽然渐进式图像修复技术能够提升纹理的细节表现,但在修复不规则的大范围缺失时,图像可能会出现对象变形和语义不一致的问题;③现有的先进模型往往具有较大的参数量。

针对现有先进图像修复方法存在的参数量过大、处理大面积缺损效果欠佳,以及在颜色复杂纹理细致图像上重建不足等问题,本文提出如下改进方法:

(1)为了降低计算复杂度并减少训练时间,本文选择了Transformer的一个变体——CSWin-Transformer。它通过使用条纹窗口自注意力机制来替代传统的完全自注意力,实现了更高效的计算。条纹窗口自注意力机制计算的是水平和垂直条纹交叉窗口的自注意力,其中每个条纹都是通过输入特征划分为等宽的条纹来获得的^[8]。这种方法可以在有限的计算成本下,快速且高效地实现对受损区域的修复。

(2)为了输出更高质量的修复图像,本文在CSWin-Transformer块中添加了一个特殊的前馈网络Gated-Dconv Feed-Forward Network。它执行受控制的特征转换,即抑制信息较少的特征,并只允许有用的信息进一步通过网络层次结构,从而导致高质量的输出^[9]。

(3)为了更精确地修复图像的细节和纹理,本文设计了一种名为DCRDNet(Deformable Convolutional Residual Dense Net)的可形变卷积残差密集网络,该网络对若干可形变卷积残差块进行多次叠加并连接,以增强图像结构纹理修复精度与品质。

1 本文方法

1.1 网络模型

针对现有图像修复方法在处理大面积缺损、复杂颜色

分布和精细纹理图像时存在的补全效果不足问题,提出一种结合 CSWin-Transformer 和可形变卷积残差密集网络的图像修复方法,网络模型如图 1 所示。以生成对抗网络为

主干网络,生成网络由全局层(改进的 CSWin-Transformer 网络)与局部层(自行设计的可形变卷积残差密集网络)组成,判别网络由谱归一化马尔科夫判别模型组成。

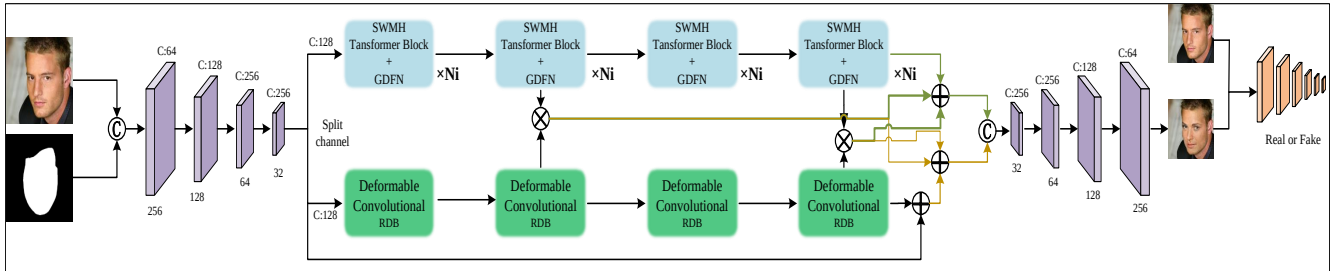


Fig. 1 Network architecture

图 1 网络结构

图像修复网络的执行流程如下:首先,对输入图像 I_{in} 进行一系列下采样卷积操作。这一步骤的核心目的在于高效地从图像中提取关键特征,为后续的修复过程奠定基础。通过下采样,网络能够捕捉到图像的高层抽象表示,同时减少计算复杂度,这对于处理大规模图像数据尤为重要。其次,将提取的特征分别送入全局和局部 2 个处理层面,以便进行更精细的操作。在这一过程中,全局层面负责捕捉图像的整体结构和上下文信息,而局部层面则专注于细节的恢复和增强。为了确保 2 个层面之间的有效信息交流,以及满足修复过程中的约束条件,网络采用了共享注意力机制。这种机制允许网络在不同尺度上动态地分配注意力资源,从而在保持全局一致性的同时,也能够精确地处理局部特征,实现高质量的图像修复效果。

在全局层,本文采用了 Cross-Shaped Window 方法,将图像分割为一系列不重叠的条纹窗口。每个窗口内,自注意力机制被应用于特征提取,以捕捉窗口内的局部依赖关系。通过整合来自所有条纹窗口的特征,全局层能够全面捕获图像内容的整体细节。随后,将这些整合后的特征进行受控的特征变换,从而有选择地抑制信息较少的特征,同时允许有价值的信息得以进一步通过网络传递。这种设计显著提高了破损图像整体细节修复的精度。

在局部层,本文设计了若干局部可形变卷积残差块,其能够灵活地适应图像破损区域周围的局部空间结构^[10]。通过堆叠这些残差块,网络能够从破损区域周围提取有效的局部特征。这些特征与输入特征通过残差连接的方式进行融合,其中权值由各残差块的输出动态调整^[11]。这种设计不仅增强了网络对有效特征的表达能力,还进一步提升了图像修复的准确性和鲁棒性。

随后,对修复后的图像执行一系列上采样卷积操作。这一过程旨在逐步恢复图像的分辨率,同时保留特征提取阶段所获得的详细信息。最终,将生成的图像与对应的真实图像一起输入至谱归一化马尔科夫判别模型中^[12]。在该模型中,生成图像与真实图像进行真假博弈对抗,通过这一对抗过程,模型能够进一步优化生成图像的质量,确保其在视觉上与真实图像无显著差异,从而实现图像修复

的目标。

1.2 改进的 CSWin-Transformer 模块

待修复图像通常展现出错综复杂的纹理和丰富的细节信息,在对破损图像进行修复时需充分利用这些有用的信息,以提高图像的修复效果。然而,破损区域往往伴随大量纹理和细节信息的丢失,导致修复区域的纹理与周围区域不匹配,从而造成视觉上的不连贯性,进而影响整体视觉效果。此外,大量纹理信息的缺失还会导致修复结果出现结构失真、伪影和模糊等问题。针对上述问题,本文引用一种基于改变条纹窗口与多头注意力机制相结合的特征提取方法,由于其十字形窗口自注意力的特殊性,只需两层即可获得全局感受野^[13]。因此,该方法能够在有限的计算成本下实现最优的修复效果。

本文所采用的 CSWin-Transformer 结构如图 2 所示。该结构在水平和垂直条纹窗口上同时执行自注意力机制,

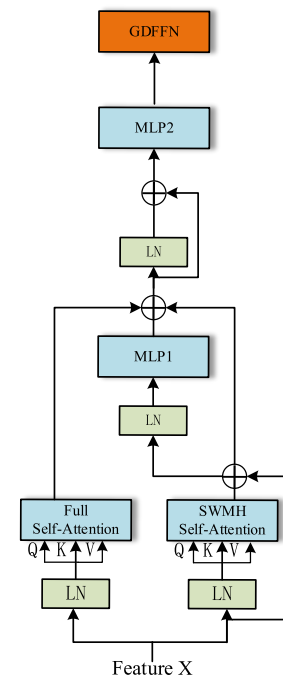


Fig. 2 Improved CSWin-Transformer module structure

图 2 改进的 CSWin-Transformer 模块结构

并逐步扩大条纹窗口的大小,使其能实现从局部到整体的特征提取。与一般的MHSA(Multi-Head Self-Attention)相比,本文所使用的SWMH-SA(Stripe Window Multi-Head Self-Attention)结合了多头和条纹窗口,大大降低了计算的复杂度,并且实现了更好的修复效果。首先,将输入特征 $X \in \mathbb{R}^{(H \times W) \times C}$ 经过线性投影转变为 k 个注意力头;其次,将 k 个注意力头平均分为2个平行组,第一组表现出横向条纹的自注意力,第二组表现为竖向条纹的自注意力;最后,连接这2组注意力头的注意力输出。对于横向条纹的注意力头,输入特征 X 被分为 $[X^1, \dots, X^M]$ 不重叠且宽度为 sw (stripe width)的横向等宽条纹。由此可得,每个条纹包含 $sw \times W$ 个标记。假设,第 k 个注意力头产生查询 $Q \in d_k$ 、键 $K \in d_k$ 和值 $V \in d_k$,则该注意力头的输出,如式(1)一式(3)所示:

$$X \in [X^1, X^2, \dots, X^M], \text{ Where } X^i \in \mathbb{R}^{(sw \times W) \times C} \text{ and } M = H/sw \quad (1)$$

$$Y_k^i = \text{Attention}(X^i W_k^Q, X^i W_k^K, X^i W_k^V), \text{ Where } i = 1, \dots, M \quad (2)$$

$$H\text{-Attention}_k(X) = [Y_k^1, \dots, Y_k^M], k = 1, \dots, K/2 \quad (3)$$

其中, H-Attention(Horizontal Stripes Self-Attention)为横向条纹自注意力。同理,竖向条纹自注意力的第 k 个注意力头输出,如式(4)所示:

$$V\text{-Attention}_k(X) = [Y_k^1, \dots, Y_k^M], k = K/2 + 1, \dots, K \quad (4)$$

最后,可以得出式(5)一式(6):

$$\text{head}_k = \begin{cases} H\text{-Attention}_k(X) \\ V\text{-Attention}_k(X) \end{cases} \quad (5)$$

$$\text{CSWin-Attention}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_k) W^0 \quad (6)$$

其中, $W^0 \in \mathbb{R}^{C \times C}$ 是一个投影矩阵,把自注意力结果投射成目标维度。公式(1)一式(6)均来源于文献[13]。这种分裂和组合的自注意力设计使得Transformer块中每一个标记的注意力作用范围都被扩大,为图像修复提供更全面和精确的特征信息。

为了产生一个层次表示,整个全局层网络由4个阶段组成。每两个相邻的阶段之间使用卷积操作,以减少token数量并使通道维数加倍。该操作使得网络前期阶段, H 和 W 远大于 C , 而在后期阶段则转变为 H 和 W 会远小于 C 。因此,在前期阶段可以选用较小的 sw (stripe width), 以实现局部特征的有效处理,而在后期阶段则使用较大的 sw (stripe width)用来捕捉更广泛的全局特征。为了达到更好的修复效果,文献[13]将4个阶段的注意力头设置为2、4、8、16,将条纹窗口的 sw (stripe width)设置为4、8、16、32。首先,由于输入特征在下采样时特征尺寸缩小为 32×32 ,前3个阶段对特征图进行局部条纹化处理,且随着 sw 增大,相应地得到不同尺度的局部条纹特征图;然后,将多头自注意力机制和条纹窗口自注意力机制分别施加于各条纹特征图,从而实现在不同的感受野上进行训练;最后,融合上述训练结果与原始特征图,获得局部融合特征。在最后阶段,由于条纹窗口的 sw (stripe width)设置为32,此时条

纹窗口能够覆盖整个特征图,进而实现了全局特征的处理。全局层网络在4个阶段中逐渐扩展注意力区域,以更好地捕获图像的纹理与内容。

此外,为了进一步降低计算量并提高图像修复的质量,在原始CSWin-Transformer模块中加入一种特殊的前馈神经网络,称为门控深度卷积前馈网络(GDFFN)。

1.3 GDFFN模块

为了进行特征转换,常规FFN(由2个 1×1 卷积构成)对每个像素进行独立处理。而GDFFN是一种经过改进的前馈神经网络,通过门控机制(2个并行分支的元素乘操作)和深度可分离卷积(3×3 深度卷积,用来学习局部结构信息)处理特征,同时,利用GELU非线性和层标准化提高模型的修复性能。GDFFN模块具体如图3所示。

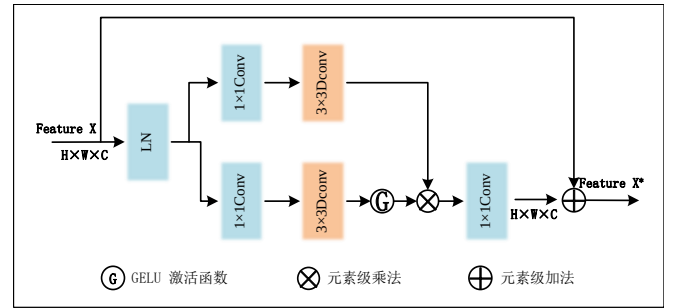


Fig. 3 GDFFN module

图3 GDFFN模块

给定一个输入张量 $X \in \mathbb{R}^{H \times W \times C}$, GDFFN的公式化如式(7)一式(8)所示:

$$\text{Gating}(X) = \Phi(W_d^1 W_p^1(\text{LN}(X))) \odot W_d^2 W_p^2(\text{LN}(X)) \quad (7)$$

$$\bar{X} = W_p^0 \text{Gating}(X) + X \quad (8)$$

其中, $\odot$ 为逐元素乘法, Φ 为GELU非线性激活函数, LN是层归一化。公式(7)一式(8)来源于文献[9]。GDFFN能够控制信息的流动,即通过抑制信息较少的特征,仅允许有用的信息通过网络层次结构进一步传递,从而使得每层专注于不同的细节信息。

尽管GDFFN通过引入门控分支和深度卷积增强了特征选择能力,但其参数量相较于常规FFN增加了约 $\gamma d \times k^2$ (其中 γ 为扩展因子, d 为输入特征的通道数, k 为卷积核尺寸)。为平衡计算成本,文献[9]将扩展因子 γ 从常规FFN的4降至2,使得GDFFN的总参数量控制在 $2\gamma d^2 + \gamma d k^2$ (与FFN的 $2\gamma d^2$ 相近),从而实现性能提升与计算负担的均衡。

1.4 可形变卷积残差密集网络

实际图像在经过时间和环境的影响后,往往会出现污损或损坏,导致有效像素的减少,甚至出现缺失。这些缺失区域在大多数情况下是不规则的。为了最大程度的利用未缺失区域的特征信息,本文引入可形变卷积方法,其卷积核的大小和位置可以根据当前需要动态的进行调整,从而适应各种缺失区域的形状,且能充分利用缺失区域周边的有效像素,更有效地提取图像的特征信息。可形变卷

积残差密集网络由多个 DCRDB (Deformable Convolutional Residual Dense Block) 模块组成。DCRDB 模块主要包含 3 个部分: 稠密连接 Dense Connection、局部特征融合 Local Feature Fusion (1x1 Conv) 以及局部残差学习 (Local Residual Learning)。这 3 个部分的相互配合形成了一种连续记忆机制^[14]。该网络的结构如图 4 所示。

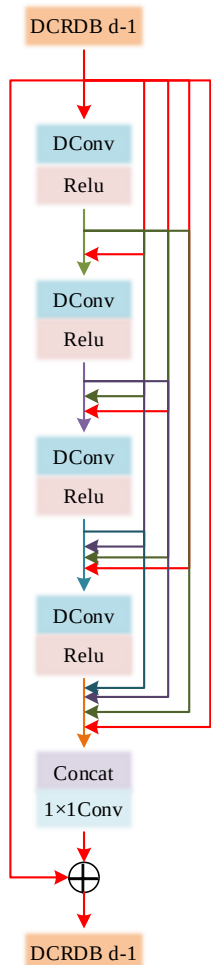


Fig. 4 DCRDB module

图 4 DCRDB 模块

通过将前一个 RDB 的状态传递给当前 RDB 的每一层, 实现了连续存储机制。设 F_{d-1} 和 F_d 分别为第 d 个 RDB 的输入和输出, 则第 d 个 RDB 的第 c 个 Conv 层的输出如式 (9) 所示:

$$F_{d,c} = \sigma(W_{d,c}[F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]) \quad (9)$$

其中, σ 表示 Relu 激活函数^[15]; $W_{d,c}$ 表示第 c 个卷积层的权重; $[F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]$ 表示第 $d-1$ 层的 DCRDB 特征图的连接; $1, \dots, c-1$ 是第 d 层的 DCRDB 的卷积层。这样前一层的 DCRDB 和各层输出都与全部后续层相连, 既保留前馈性质又提取局部密集特征。

本文提出的局部特征融合模块 (采用 1x1 卷积层) 能够自适应地整合先前 DCRDB 与当前 DCRDB 的所有卷积层特征, 从而实现输出信息的动态调控。具体如式 (10) 所示:

$$F_{d,LF} = H_{LFF}^d([F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]) \quad (10)$$

其中, H_{LFF}^d 为第 d 层 DCRDB 中的 1x1 卷积。

由于单个 DCRDB 中有多个卷积层, 为进一步提高信息流的质量, 将局部残差学习引入 DCRDB。第 d 层 DCRDB 的输出如式 (11) 所示:

$$F_d = F_{d-1} + F_{d,LF} \quad (11)$$

公式 (9) — (11) 来源于文献 [14]。

可形变卷积残差密集块 (DCRDB), 不仅可以连续记忆 (CM) 机制获取先前的 RDB 的状态, 还可以通过局部密集连接充分利用所有的卷积层^[16]。最终, 通过局部特征融合 (LFF) 自适应地保留累积的特征。

通过连续叠加多个具备可变形能力的卷积残差单元, 可以显著提升网络对特征的表达能。这些单元利用非线性处理, 能够捕获多维度和多尺度的特征, 从而更深入地解析图像。此外, 每个单元在处理数据时, 通过跳跃连接将特征进行叠加, 这不仅有助于在数据流经网络时保持特征的连贯性, 也增强了特征的表示力。这种多层堆叠的设计促进了特征的逐层提炼, 使模型能够从简单的局部特征逐渐向更为复杂和抽象的语义层面转变, 从而增强修复模型多层次特征的识别能力。同时, 深层的堆叠结构还增强了模型的稳定性, 提升了模型的泛化性能。

1.5 损失函数

损失函数的选择对于图像修复任务至关重要。通过选择合适的损失函数, 可以确保修复后的图像不仅视觉上更加逼真, 结构上更加连贯, 还能有效保留原始图像的重要特征。具体而言, 合理的损失函数能够确保修复区域在像素保真度、边缘和纹理一致性、色彩准确性、结构相似性以及整体感知质量方面与原始图像保持高度一致, 从而使修复后的图像在视觉上无缝且自然地融入未受损区域。损失函数的计算也不应过于复杂, 以免影响模型训练和推理的速度。本文通过设计一个混合型损失函数来完成图像修复任务, 具体包括感知损失^[17] (Perceptual_loss)、风格损失^[18] (style_loss)、边缘损失 (Edge_loss)、对抗损失^[19]、HS 损失 (HS_loss)^[20] 和 L1 损失。

2 实验结果与分析

2.1 实验设置

为了验证本文方法的有效性, 使用 Celeba-HQ 数据集、facade 数据集和自制西藏唐卡数据集进行训练和评估。其中, Celeba-HQ 是 CelebA 数据集的一个扩展版本, 包含了高质量和高分辨率的人脸图像。本文选取了 26 000 张图像用于训练, 2 000 张图像用于测试和验证。facade 数据集则包含了在机器视觉中心收集的 606 张正面建筑图像, 且已进行人工标注。这些建筑图像来源于全球多个城市, 展现了多样化的建筑风格。本文选取了 528 张图像用于训练, 78 张图像用于测试和验证。自制西藏唐卡数据集包

含佛教唐卡和护法唐卡^[21]。本文选取了1 000张图像用于训练,152张图像用于测试和验证。

本文的实验环境基于 Windows 10 操作系统,配备了 NVIDIA GeForce RTX 3090 显卡、16.0 GB 的内存以及 Intel (R) Core i7-11700 CPU,所有对比实验均在相同的硬件配置上执行。研究中采用 PyTorch 框架进行模型的构建,训练阶段利用 Adam 优化器进行参数优化,设置的批次大小为 8,生成器学习率为 1×10^{-3} 。为了进行定量分析,本文

采用了多种图像质量评价指标。这些指标包括经典的峰值信噪比 (PSNR)^[22]、结构相似性指数 (SSIM)^[23],以及可学习感知图像块相似性度量 (LPIPS)^[24]。

2.2 定量比较

为了更全面的评估本文修复方法的有效性,本文与近年来基于深度学习的图像修复方法进行对比分析,包括 CSWT、CTSDG^[25]、PUT^[26]和 ICT^[27]。定性比较结果如图 5 所示;定量比较结果如表 1 所示。

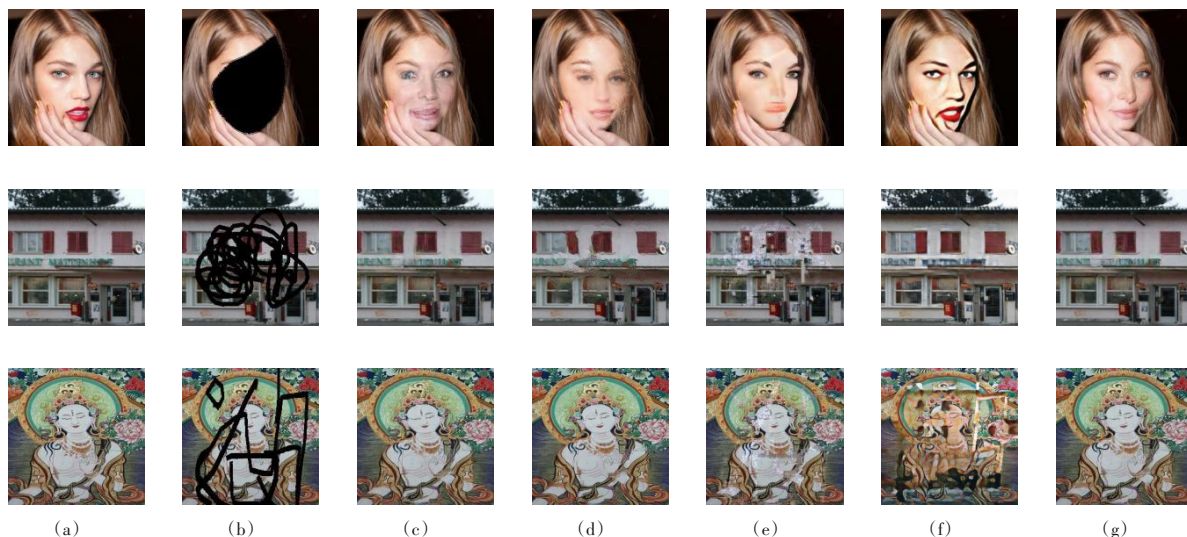


Fig. 5 Qualitative comparison results

图 5 定性比较结果

Table 1 Quantitative comparison results

表 1 定量比较结果

Method	Dataset	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
CSWT	Celeba	21.84	0.835 4	0.133 9
CTSDG		21.83	0.811 8	0.195 9
PUT		21.00	0.801 9	0.185 9
ICT		22.65	0.840 1	0.185 0
OURS		23.27	0.861 0	0.106 2
CSWT	Facade	28.44	0.905 5	0.112 5
CTSDG		25.60	0.850 5	0.183 8
PUT		22.96	0.821 9	0.217 1
ICT		22.99	0.800 8	0.261 4
OURS		28.48	0.906 5	0.104 4
CSWT	TK	23.18	0.862 1	0.095 5
CTSDG		22.28	0.856 1	0.101 7
PUT		23.56	0.835 3	0.147 6
ICT		13.60	0.538 7	0.393 1
OURS		23.16	0.864 7	0.095 4
平均提升	/	2.47	0.075 2	0.092 4

其中,图 5(a)为原始未破损图片,图 5(b)为添加掩码后的图片,图 5(c)、图 5(d)、图 5(e)、图 5(f)、图 5(g)分

别对应文献 CSWT、CTSDG、PUT、ICT 和本文方法的修复结果。实验结果表明,本文方法在客观评价指标上取得了全面提升:PSNR 值较对比方法提高 2.47 dB,SSIM 系数提升 0.075 2,同时感知质量指标 LPIPS 显著降低 0.092 4。定量指标的显著改进与主观视觉评估结果一致,生成的修复结果在纹理细节和结构合理性方面均展现出明显优势。

2.3 消融实验

为了验证本文提出的模块在新架构中的效果,进行了一系列消融实验。实验对比了未进行重新设计的 CSWin-transformer 原始版本和经过重新设计的 CSWin-transformer 版本,修复后的结果如图 6 所示。其中,图 6(a)为原始图像,图 6(b)为掩码图像,图 6(c)为原始方法修复后的图像,图 6(d)为使用可形变卷积残差密集块后修复的图像,图 6(e)为使用门控深度卷积残差密集块后所修复的图像,图 6(f)为使用本文方法修复后的图像。

消融实验结果如表 2 所示,记录了使用各个模块进行修复后的 PSNR、SSIM 和 LPIPS 指标。在本文中,PSNR、SSIM 和 LPIPS 指标采用多个图像修复后的平均数值。

根据图 6 和表 2 的结果可知,本文方法不仅在视觉上呈现最合理的修复效果,而且在 PSNR、SSIM 和 LPIPS 指标上均表现最佳。

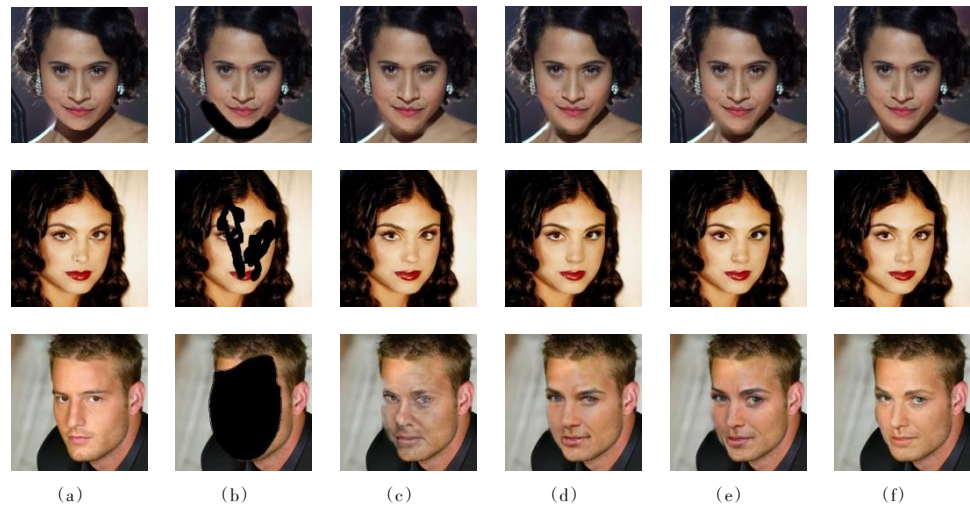


Fig. 6 Ablation study
图 6 消融实验

Table 2 Ablation study results
表 2 消融实验结果

SWMHT			SWMHT+DCRDNet			SWMHT+GDFFN			本文方法		
PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
28.55	0.916	0.044	28.88	0.917	0.043	28.62	0.918	0.045	29.44	0.921	0.034

3 结语

本文提出一种结合 CSWin-Transformer、GDFFN 和可形变卷积的图像修复方法。该模型的全局层由 CSWin-Transformer 和全局特征网络(GDFFN)构成,局部层由可变形卷积残差密集网络组成,两者并行运作,以实现破损图像的补全。在全局处理阶段,CSWin-Transformer 利用其特有的动态条纹窗口机制来提取图像的特征,此操作可以大幅减少计算量;GDFFN 通过门控机制和深度可分离卷积处理特征,利用 GELU 非线性和层标准化提高模型的修复性能。GDFFN 可以抑制信息较少的特征,只允许有用的信息进一步通过网络结构,从而提高模型的修复质量。局部层采用可形变卷积去自适应的提取图像的有用信息,并利用残差链接将原始输入与特征图相结合,以增强模型的学习能力和泛化性能。这种方法可以帮助网络更好地处理图像中的几何变化,提高对复杂场景的适应性。在全局层和局部层之间运用联合注意,将全局层提取的特征与局部层提取的特征混合以实现层间信息共享,使修复的图像在结构和细节上更加协调。实验数据证实,本文方法在图像修复领域展现出优异的性能,在常用的评估标准上也超越了现有技术。然而,本文方法需要同时输入受损图像及其对应的遮罩图像作为必要条件。但在实际应用中,精确获取损坏区域的信息通常是不现实的,且手动标记这些区域不仅耗时,还难以保证标记的精确度。因此,未来的研究将更加深入地探讨这些领域。

参考文献:

[1] XIA B, ZHANG Y, WANG S, et al. Diffir: efficient diffusion model for image restoration [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 13095–13105.

[2] YANG S, LIANG H, WANG Y, et al. Image inpainting based on multi-patch match with adaptive size [J]. Applied Sciences, 2020, 10 (14) : 4921.

[3] WEI G, GUO J, KE Y, et al. A three-stage GAN model based on edge and color prediction for image outpainting [J]. Expert Systems with Applications, 2022, 214: 119136.

[4] LI W, LIN Z, ZHOU K, et al. MAT: mask-aware transformer for large hole image inpainting [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 10758–10768.

[5] WAN Z, ZHANG J, CHEN D, et al. High-fidelity pluralistic image completion with transformers [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4692–4701.

[6] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [DB/OL]. <http://export.arxiv.org/abs/1810.04805>.

[7] ZHENG C, CHAM T J, CAI J, et al. Bridging global context interactions for highfidelity image completion [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11512–11522.

[8] DONG X, BAO J, CHEN D, et al. CSWin transformer: a general vision transformer backbone with cross-shaped windows [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 12124–12134.

[9] ZAMIR S W, ARORA A, KHAN S H, et al. Restormer: efficient transformer for high-resolution image restoration [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:

- 5718–5739.
- [10] QIU Z, YUAN L, LIU L, et al. Generative image inpainting with dilated deformable convolution[J]. *Journal of Circuits, Systems and Computers*, 2022, 31(6): 2250114.
- [11] SUN L, JIANG B, YANG C, et al. RePGAN: image inpainting via residual partial connection and mask discriminator[J]. *International Journal of Machine Learning and Cybernetics*, 2023, 14(9): 3193–3203.
- [12] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 4471–4480.
- [13] CHEN B W, LIU T J, LIU K H. Lightweight image inpainting by stripe window transformer with joint attention to CNN[C]//2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP), 2023: 1–6.
- [14] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 2472–2481.
- [15] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//*Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011: 315–323.
- [16] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700–4708.
- [17] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]//14th European Conference on Computer Vision, 2016: 694–711.
- [18] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2414–2423.
- [19] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139–144.
- [20] NIU Z, LI H, LI Y, et al. An adaptive face image inpainting algorithm based on feature symmetry[J]. *Symmetry*, 2020, 12(2): 190.
- [21] LIU D Q. Research and application of thangka image restoration based on deep learning[D]. Xi'ning: Qinghai Normal University, 2021.
刘丹青. 基于深度学习的唐卡图像修复研究与应用[D]. 西宁: 青海师范大学, 2021.
- [22] KORHONEN J, YOU J. Peak signal-to-noise ratio revisited: is simple beautiful?[C]//2012 Fourth International Workshop on Quality of Multimedia Experience, 2012: 37–38.
- [23] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600–612.
- [24] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 586–595.
- [25] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 14134–14143.
- [26] LIU Q, JIANG Y, TAN Z, et al. Transformer based pluralistic image completion with reduced information loss[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(10): 6652–6668.
- [27] WAN Z, ZHANG J, CHEN D, et al. High-fidelity pluralistic image completion with transformers[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 4692–4701.

(责任编辑:陈维捷)