

## 基于感知推理和外部空间先验特征的图像修复

吴鹏, 张孙杰, 王永雄, 陈远峰, 覃海旺

(上海理工大学光电信息与计算机工程学院, 上海 200093)

**摘要:** 在基于深度学习的图像修复算法中, 当存在大面积掩码时, 由于缺乏合理的先验信息指导, 修复结果往往会出现伪影和模糊纹理等现象。针对此问题, 提出将先验特征与图像预测滤波相结合的图像修复算法。该算法包含两个分支: 图像滤波核预测分支和特征推理与图像滤波分支。从图像滤波核预测分支的解码器部分提取特征, 利用多尺度外部空间特征融合对掩码区域特征进行重建, 并传递给另一分支的解码阶段作为先验特征, 为图像修复提供更为丰富的语义信息。然后, 在特征推理和图像滤波分支部分引入空间特征感知推理块, 它能够过滤掉分散注意力的特征, 同时捕捉信息丰富的远距离图像上下文进行推理。最后, 使用图像预测滤波核进行过滤消除伪影。在 CelebA 和 Places2 数据集上与其他修复网络进行对比实验, 证明了该方法在修复质量上的优越性。

**关键词:** 图像修复; 先验特征; 图像预测滤波; 特征感知推理; 外部空间特征融合

中图分类号: TP391.41

文献标志码: A

### Image Inpainting Based on Perceptual Inference and External Spatial Prior Features

WU Peng, ZHANG Sunjie, WANG Yongxiong, CHEN Yuanfeng, QIN Haiwang

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** Image inpainting based on deep learning has made a lot of remarkable progress. However, when there is a large area mask, due to the lack of reasonable prior information guidance, the repair results often appear artifacts and blurred textures. Therefore, we propose an image inpainting algorithm that combines prior features with image predictive filtering. It consists of two branches: Image filtering kernel prediction branch and feature inference and image filtering branch. The features are extracted from the decoder part of the image filter kernel prediction branch. The multi-scale external spatial feature fusion is used to reconstruct the mask region features, and the decoding stage is passed to another branch as a prior feature to provide richer semantic information for image inpainting. Then, a spatial feature-aware inference block is introduced in the feature inference and image filtering branches, which can filter out the distracting features and capture the informative long-distance image context for inference. Finally, the image prediction filter kernel is used to filter and eliminate artifacts. Compared with other repair networks on CelebA and Places2 datasets, the superiority of the method in repair quality is proved.

**Key words:** image inpainting; prior features; image predictive filtering; feature perception reasoning; external spatial feature fusion

基金项目: 上海市晨光学者基金(18CG52); 上海市自然科学基金(22ZR1443700)。

收稿日期: 2023-04-25; 修订日期: 2023-06-29

## 引言

图像修复旨在根据已知信息,用语义一致且纹理和谐的内容来填充图像缺失的像素。目前,图像修复算法在许多应用中发挥着重要作用,如物体去除、照片恢复等。早期的研究工作试图通过基于扩散<sup>[1]</sup>或基于补丁<sup>[2]</sup>的算法来解决图像修复问题,然而由于缺乏对复杂场景中缺失的内容和纹理的有力推理,他们在修复过程中往往无法生成高级的语义。深度特征学习<sup>[3]</sup>和生成对抗性网络(Generative adversarial network, GAN)<sup>[4]</sup>的出现使得图像修复的研究取得了重大进展。许多开创性的工作试图将先验信息引入到多阶段模型,该模型依次修复各种辅助信息和损坏的图像,以增强图像修复。这些方法学习特定领域的先验,如具有良好视觉可解释性的结构<sup>[5]</sup>或语义<sup>[6]</sup>,但在第一阶段丢弃了太多有用的信息,如颜色和纹理,使得它们的修复结果远不能令人满意。此外,一些学者提出了使用图像级预测滤波来减轻伪影<sup>[7]</sup>,图像级预测滤波通过待填补像素的相邻像素来重建像素。根据输入自适应地估计滤波内核,可以在避免伪影的同时恢复局部结构。然而,这些方法由于不能充分利用远距离依赖关系进行上下文推理,在大区域掩码上生成合理内容存在着困难。

为了克服上述困难,本文提出一种将先验特征与图像预测滤波相结合的交互式图像修复方法,它包括两个分支:图像滤波核预测分支(Image filtering kernel prediction branch, IFKPB),以及特征推理和图像滤波分支(Feature inference and image filtering branch, FIIFB)。FIIFB为IFKPB提供多级特征,IFKPB为FIIFB提供先验语义特征和动态预测内核。为了使FIIFB能够获得足够丰富的语义,本文提出了多尺度外部空间特征融合模块(Multi-scale external spatial feature fusion module, MESFF),它对受损图像的上下文及其特征之间的长距离依赖性进行建模,通过在空间维度上操作的轻量级外部注意力机制实现空间信息传播,并采用多尺度聚合方式来突出来自遥远空间位置的线索,以渲染更精细的细节,从而提高语义引导和特征的质量。同时针对大掩码区域修复问题,本文还提出了空间特征感知推理块(Spatial feature-aware inference block, SFI),它在利用远距离上下文信息的同时,通过过滤掉分散注意力的特征来抑制空间噪声,实现空间自适应地关注掩码区域进行推理。最后通过图像预测滤波来进行细化和去伪影的工作,从而完成图像的高质量修复。

本文的贡献总结如下:(1)本文构造将先验特征与图像预测滤波相结合的交互式模型,利用丰富语义和像素级填充的优势互补进行细节重建;(2)本文提出多尺度外部空间特征融合模块,该模块将受损图像的上下文和编码特征集成在一起,并使用轻量级的外部注意力机制实现空间信息传播,以提高修复质量;(3)本文设计空间特征感知推理块作为网络的基本块,它能够自动聚焦掩码区域,允许捕获信息丰富的远距离上下文进行推理的同时,过滤掉分散注意力的特征来抑制空间噪声。

## 1 相关工作

### 1.1 利用先验信息进行图像修复

在之前的工作中,很多利用先验信息进行修复研究已经取得良好的成果。这些先验信息包括边缘,线,梯度,以及其他信息<sup>[8]</sup>,甚至潜在先验<sup>[9]</sup>。这些方法总的来说可以分为两种类型:一种方法是一开始矫正先验信息,然后采用多阶段模型指导修复,虽然这些方法具有良好性能和可解释性,但由于利用的先验信息不够全面,故无法正确处理图像修复;另一种方法使用监督信息直接监督修复模型,以便能够引入更多的有效先验信息,但这些低级特征仍然不足以实现具有丰富语义的结果。

### 1.2 图像预测滤波

深度预测滤波是最近提出的用于图像处理的深度神经网络框架<sup>[10]</sup>,其基本思想是通过预先训练的深度学习网络预测所有像素的核,并使用核进行逐像素图像滤波。Guo等<sup>[11]</sup>将其扩展到去噪领域,而Fu

等<sup>[12]</sup>将其视为融合多个曝光图像以实现高效阴影去除的工具。与像噪声、雨条纹和阴影这样的退化相反,空穴(即缺失区域)使图像信息完全丢失。直观地说,小孔或大孔边界处缺失的像素可以通过它们的相邻像素来重建。然而,大洞中心的像素可能具有不同的可能性,并且不能从相邻像素中恢复。这些像素应该通过理解整个场景来推理。

### 1.3 注意力机制

注意力图像修复模型可以利用注意力机制从背景中借用特征,借用背景区域更逼真的纹理填补缺陷。例如多尺度注意力模块<sup>[13]</sup>、语义关注层<sup>[14]</sup>等,以此来保证交换特征的语义相关性。尽管这些方法提供了相当大的改进,但却面临一个明显的缺点,即由于大量的矩阵乘法运算,其计算复杂度太高。

## 2 本文方法

### 2.1 主干结构

本文的主要模型框架如图1所示,它由2个具有相似架构的分支组成:图像滤波核预测分支,以及特征推理和图像滤波分支。这2个分支是包含几个卷积块的编码-解码器网络。采用了交互链接的方式,即IFKPB分支以原始图像和FIIFB的特征为输入,并预测FIIFB的图像预测核。此外,从IFKPB的解码器阶段的第2层提取特征,由多尺度外部空间特征融合模块对其长距离依赖性进行建模,作为特征先验信息传递至FIIFB的编码器阶段。之后,再通过空间特征感知推理块进行大范围掩码修复的上下文推理和纹理合成,同时实现空间噪声去除以及特征增强。最后经过图像预测滤波核过滤后,得到修复完成的图像。

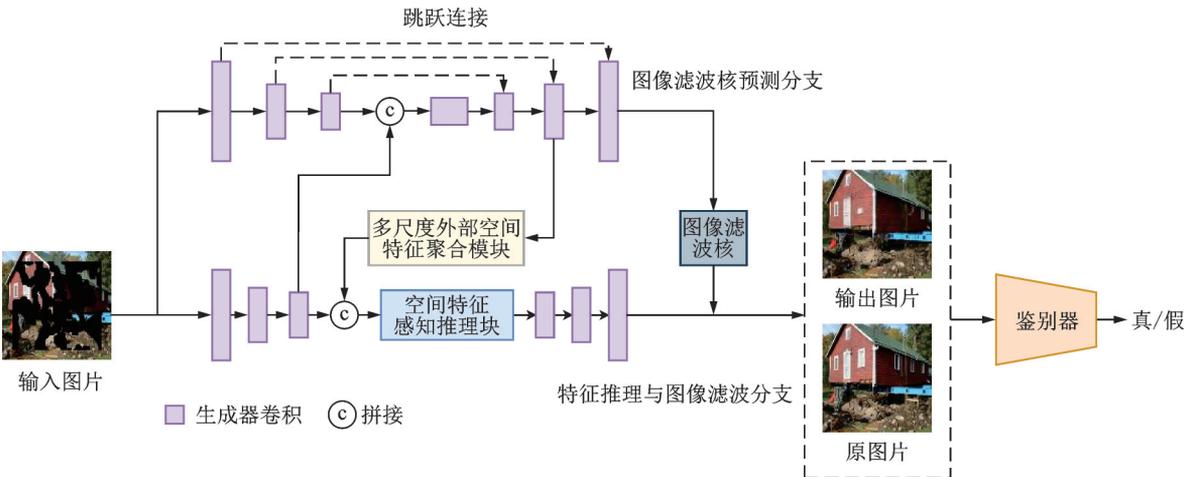


图1 本文模型整体框架

Fig.1 Overall architecture of the proposed model

### 2.2 图像级预测滤波

图像级预测滤波结合了经典图像滤波和深度学习的优点,在图像去噪和去阴影方面取得良好的效果。在图像修复之中使用,可以有效改善图像伪影等问题,将其化为逐像素预测滤波任务。其基本流程为用深度卷积神经网络预测的像素级核对图像进行滤波,在给定输入损坏图像 $I$ ,通过式(1)方式使用像素核级内核对其进行处理。

$$\hat{I}[p] = \sum_{q \in N_p} K_p[q-p] I[q] \tag{1}$$

式中: $p$ 为图像中的待填补像素点; $N_p$ 为像素 $p$ 的 $N$ 个相邻像素的集合; $K_p$ 为像素 $p$ 的专用内核。在图像的修复过程中,内核 $K_p$ 确定像素 $q$ 的组合权重,实现缺失的像素由其周围像素进行推导。

由于缺失结构的局部多样性,为了适应上下文变化,IFKPB分支以原始图像和FIIFB的特征为输入,训练输入图像所需要的所有像素的核,其过程为

$$K_p = \varphi_L \left( \cdots \varphi_{j+1} \left( [E_j, F_j] \right) \right) \quad (2)$$

式中: $\varphi_L(\cdots \varphi_{j+1}(\cdot))$ 表示采用UNet架构<sup>[15]</sup>从IFKPB分支的 $j+1$ 层逐层卷积到最后一层 $L$ 层,通过L1损失、生成对抗损失、风格损失和感知损失来对其进行训练; $E_j$ 为IFKPB分支的第 $j$ 层特征; $F_j$ 为FIIFB分支的第 $j$ 层特征,两个特征进行拼接操作。用于图像滤波的预测内核以及作为先验的语义特征都由输入图像 $I$ 和深度特征 $F_j$ 驱动,它们包含所有可用的空间细节和对整个场景的理解。预测滤波方法能够准确填充小区域掩码,也可以在图像修复最后阶段进行局部结构保留和伪影去除工作。

### 2.3 多尺度外部空间特征融合模块

如图2所示,多尺度外部空间特征融合模块从IFKPB分支的解码器阶段提取特征,通过外部空间注意力计算特征图的长距离依赖关系,并通过多尺度融合保持了整体图像的一致性,最后将其作为先验特征馈送到FIIFB的编码器阶段,为特征掩码部分提供丰富语义。与内部计算注意力权重的注意力机制不同,外部空间注意力利用两个外部可学习的Key矩阵 $K$ 和Value矩阵 $V$ ,分别沿维度 $H$ 和维度 $W$ 与输入查询相乘,以实现空间信息传播。外部Key矩阵和Value矩阵共享相同的结构,即简单建立在激活函数和两个线性层上,它们独立于输入查询特征并在样本之间共享,提高了泛化能力,并为注意力机制带来了线性复杂性。另外,该模块还采用多尺度特征融合来编码多尺度的丰富语义特征,通过提取不同尺度的特征,增强对不同特征的表达能力<sup>[16]</sup>,从而很好地平衡了准确性和复杂性。图中: $F$ 为通过卷积调整大小后的特征图; $P^1$ 、 $P^2$ 、 $P^3$ 为多尺度特征对应的权重; $F_0$ 为重建后的特征; $F_{y1}$ 、 $F_{y2}$ 、 $F_{y3}$ 为重构的特征图。

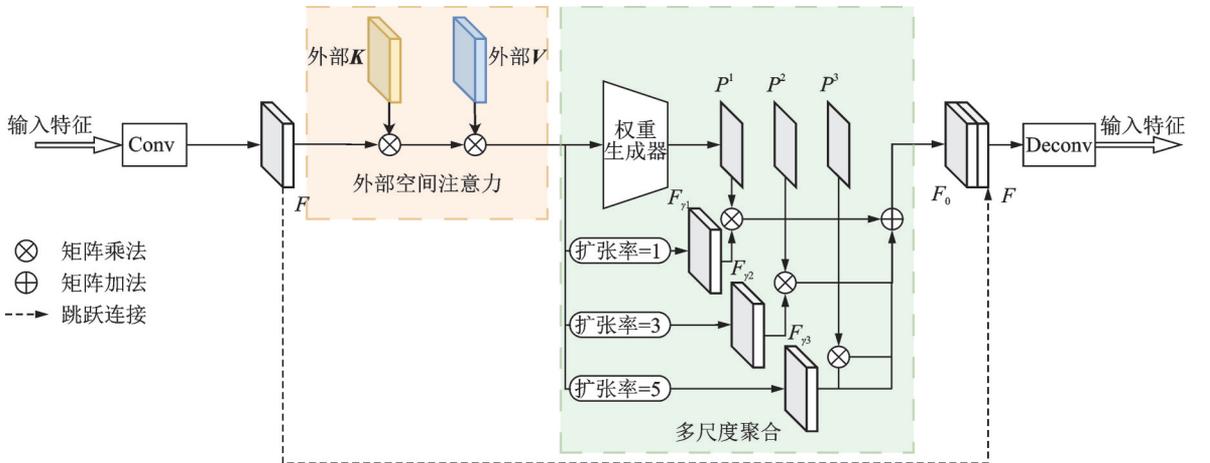


图2 多尺度外部空间特征融合模块

Fig.2 Structure of multi-scale external spatial feature fusion module

具体地说,该模块将输入特征图调整为256像素 $\times$ 256像素的大小,随后进行外部注意力的计算,其过程为

$$F_y = (Q^T \cdot K)^T \cdot V \quad (3)$$

式中: $F_y$ 为重构的特征图; $Q^T$ 为输入特征。

在重建特征图时,采用不同扩张率的扩张卷积层来获得多尺度特征 $F_y^k$ 为

$$F_y^k = \text{Conv}_k(F_y) \tag{4}$$

式中: $\text{Conv}_k$ 表示不同扩张率的卷积; $k$ 代表选取的不同的扩张率, $k \in \{1, 3, 5\}$ 。

之后采用了一个像素级权重图生成器 $T_w^{[17]}$ ,以此来生成不同尺度特征对应的权重,其内核分别为3和1的两个卷积层组成,每个卷积层后包含一个ReLU非线性激活,输出通道设置为3,其权重计算如下

$$P = \text{softmax}(T_w(F_y)) \tag{5}$$

$$P^1, P^2, P^3 = \text{Slice}(P) \tag{6}$$

式中: $\text{softmax}(\cdot)$ 为通道方向的softmax; $\text{Slice}(\cdot)$ 为通道方向的切片。最后通过元素加权的和融合多尺度特征得到重建结果 $F_0$ 为

$$F_0 = (F_y^1 \odot P^1) \oplus (F_y^2 \odot P^2) \oplus (F_y^3 \odot P^3) \tag{7}$$

在最后的输出阶段采用跳越连接来防止补丁移位操作造成的语义损失。

### 2.4 空间特征感知推理块

为了推断合理的内容并为大的自由形式缺失区域生成清晰的纹理,本文引入了聚合上下文变换块(Aggregated contextual transformation, AOT)<sup>[18]</sup>。AOT块能够捕获信息丰富的远距离上下文和丰富的兴趣模式,以增强上下文推理。其基本结构如图3所示,AOT块将标准卷积的内核拆分为多个子内核,每个子内核的输出通道较少。图中: $r$ 为膨胀率; $f_1$ 为聚合输入特征; $f_2$ 为学习到的残差特征。每个子内核通过使用不同的膨胀率来执行输入特征 $f_1$ 的不同变换。使用较大的膨胀率可以使子内核“看到”输入图像的较大区域,而使用较小膨胀率的子内核则专注于较小感受野的局部模式。来自不同感受野的上下文转换最终通过级联进行聚合,然后进行标准卷积进行特征融合。这样的设计允许AOT块通过不同的视图来预测每个输出像素。通过以上3个步骤,AOT块能够聚合多个上下文转换,以增强上下文推理。为了缓解修复图像中的颜色差异问题,在构建块中使用一种新的门控残差连接。门控残差连接首先通过标准卷积和Sigmoid运算从 $f_1$ 计算空间变化的门值 $k$ ,然后AOT块通过与 $k$ 的加权和来聚合输入特征 $f_1$ 和学习的残差特征 $f_2$ ,表示为

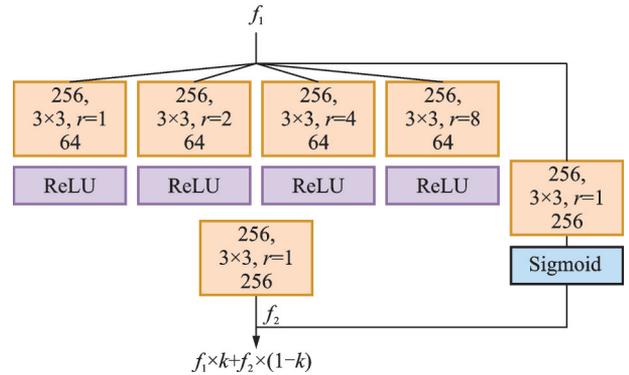


图3 AOT块结构

Fig.3 Structure of AOT block

$$f_k = f_1 \times k + f_2 \times (1 - k) \tag{8}$$

但AOT块的上下文捕获能力在空间域缺乏表现,为了进一步使掩码区域能够捕获潜在的有用语义,本文提出了一种基于AOT的空间特征感知推理块,预测了带有空间噪声的单通道空间映射。然后,将其与输入特征 $F$ 连接起来,输入到AOT块中,以生成感知图像特征 $F_q$ ,该特征捕捉了掩码区域有可能导致伪影的无意义特征。再从输入特征 $F$ 中减去 $F_q$ ,以消除 $F_q$ 特征对修复结果的负面影响,最后通过AOT残差块来增强局部特征。该过程实现了图像部分伪影的去除,并使模型专注于掩码区域的

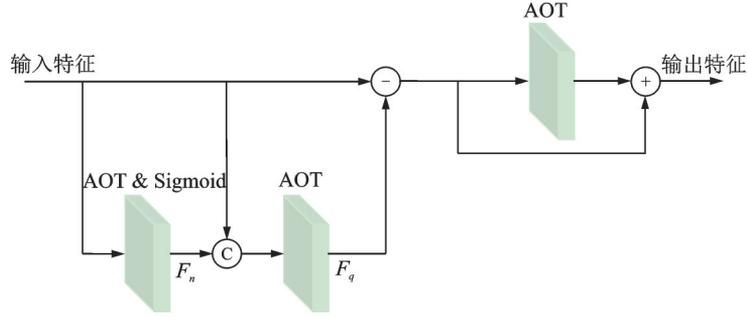


图4 空间特征感知推理块结构

Fig.4 Structure of spatial feature-aware inference block

修复。如图4所示,给定特征图  $F \in \mathbf{R}^{H \times W \times C}$ ,空间特征感知推理块首先通过AOT块和Sigmoid函数 $\delta$ 计算空间注意力图  $F_n$ ,表达式为

$$F_n = \delta(\text{AOT}(F)) \quad (9)$$

式中  $F_n$  为单通道空间映射。然后使用这个空间先验和原始特征来获得空间噪声,即

$$F_q = \text{AOT}(\text{concat}(F, F_n)) \quad (10)$$

去除空间噪声,并通过使用AOT块学习残差来增强特征,并获得输出特征  $F_t$ ,即

$$F_t = F - F_q + \text{AOT}(F - F_q) \quad (11)$$

本文堆叠了6个空间特征感知推理块在FIIFB的解码器与编码器之间作为bottleneck。

## 2.5 损失函数

本模型采用了一种联合损失,包括对抗性损失  $L_{\text{gan}}$ 、感知损失  $L_{\text{perc}}$ 、风格损失  $L_{\text{style}}$  和L1损失  $L_1$ 。

(1)对抗损失:对抗损失由两部分组成:鉴别器损失和生成器损失。鉴别器损失用以衡量鉴别器的判断结果是否与事实相符。而生成器损失则是用来衡量生成图片与真实图片的接近程度。设  $G$  为生成器, $D$  为鉴别器,其表达式为

$$L_{\text{gan}} = \min_G \max_D E_{I_{\text{gt}}}(\log I_{\text{gt}}) + E_{I_{\text{out}}}(1 - \log I_{\text{out}}) \quad (12)$$

式中:  $I_{\text{gt}}$  表示真实图片;  $I_{\text{out}}$  表示生成器生成的图片。

(2)感知损失:由于一般性损失难以捕获高级语义,本文引入感知损失来评估图像的全局结构。它实际测量了ImageNet上预训练的VGG-16网络特征空间中  $I_{\text{out}}$  与  $I_{\text{gt}}$  之间的  $L_1$  距离,表达式为

$$L_{\text{perc}} = E \left[ \sum \left\| \phi_i(I_{\text{out}}) - \phi_i(I_{\text{gt}}) \right\|_1 \right] \quad (13)$$

式中  $\phi_i(\cdot)$  表示给定输入图像  $I_i$  的VGG-16的第  $i$  个池化层的激活图。

(3)风格损失:利用式(12)中的激活图可以计算图像风格损失,以测量激活图的协方差之间的差异,从而减轻“棋盘格”伪影,保证风格一致性。表达式为

$$L_{\text{style}} = E_j \left[ \sum \left\| G_j^\phi(I_{\text{out}}) - G_j^\phi(I_{\text{gt}}) \right\|_1 \right] \quad (14)$$

式中  $G_j^\phi(\cdot)$  为求取对应激活图Gram矩阵的函数。

(4)L1损失:即绝对平均误差(Mean absolute error, MAE),是预测结果值与实际真实值之间的绝对误差和。L1损失在图片修复过程中用于测量修复结果与真实结果之间的绝对差距,在训练过程中要尽可能减少其之间的绝对误差。

其联合损失为

$$L_{total} = \lambda_1 L_1 + \lambda_2 L_{gan} + \lambda_3 L_{perc} + \lambda_4 L_{style} \tag{15}$$

本文设置  $\lambda_1 = 1, \lambda_2 = 0.1, \lambda_3 = 0.1, \lambda_4 = 250$ 。

### 3 实验结果与分析

#### 3.1 实验环境和实验设置

本文方法在 Places2<sup>[19]</sup>和 CelebA<sup>[20]</sup>两个具有挑战性的数据集上进行了实验。为了进一步验证实验的有效性,本文还选取了实地拍摄图像作为验证手段。所有图像被调整为 256 像素 × 256 像素进行训练和测试,以与现有方法进行公平比较,并采用不规则掩模进行训练,使本文模型能够处理不同类型的受损图像。

评价指标选取了图像修复任务中常用的几个指标,如峰值信噪比(Peak signal-to-noise ratio, PSNR)、L1 损失  $L_1$  以及结构相似性指数(Structural Similarity, SSIM)。 $L_1$ 、PSNR 和 SSIM 用于测量修复后图片的质量。

在图片的掩码设置方面,采用了 3 类不规则掩码,即 0%~20%、20%~40% 和 40%~60%。在此基础上,该模型与几个最先进的网络性能上进行比较,其中包括 EdgeConnect<sup>[5]</sup>、AOT-GAN<sup>[18]</sup>、CTS-DG<sup>[17]</sup>和 JPGNet<sup>[7]</sup>。

该模型在 PyTorch 框架中实现。BatchSize 设置为 4,初始学习率为  $1 \times 10^{-4}$ ,使用 Adam 优化器。计算硬件使用了 NVIDIA GTX 1080 GPU(8 GB)。

#### 3.2 定量分析

Places2 和 CelebA 数据集上不同方法的客观性能对比和实拍图像上不同方法的客观性能对比如表 1 和表 2 所示。表 1,2 中符号“↓”表示越低越好,“↑”表示越高越好,其中,Places2 和 CelebA 选取 500 张图像进行测试,实地拍摄图像选取 50 张图像进行测试。在 2 个公共数据集以及实地拍摄图像的验证

表 1 Places2 和 CelebA 数据集上不同方法的客观性能对比

Table 1 Objective performance comparison of different methods on Places2 and CelebA dataset

评价指标	方法	Places2 数据集			CelebA 数据集		
		10%~20%	20%~40%	40%~60%	10%~20%	20%~40%	40%~60%
PSNR ↑	EdgeConnect	29.717	23.251	19.440	32.327	25.421	20.201
	AOT-GAN	21.384	20.873	18.474	18.032	17.980	16.263
	CTSDG	30.582	23.361	19.325	32.553	24.782	20.048
	JPGNet	29.913	23.166	19.400	35.816	28.217	23.072
	本文方法	<b>30.809</b>	<b>23.762</b>	<b>19.714</b>	<b>36.857</b>	<b>28.790</b>	<b>23.348</b>
SSIM ↑	EdgeConnect	0.939 2	0.829 6	0.674 2	0.975 4	0.910 9	0.775 3
	AOT-GAN	0.846 9	0.833 5	0.670 9	0.746 6	0.739 6	0.625 5
	CTSDG	0.942 2	0.826 5	0.673 6	0.958 1	0.869 7	0.738 6
	JPGNet	0.939 9	0.820 6	0.668 6	0.970 8	0.904 3	0.796 4
	本文方法	<b>0.947 5</b>	<b>0.835 5</b>	<b>0.685 0</b>	<b>0.975 6</b>	<b>0.912 3</b>	<b>0.808 2</b>
$L_1$ ↓	EdgeConnect	1.17	3.44	6.83	0.98	2.98	6.15
	AOT-GAN	7.51	6.50	9.07	9.78	9.61	11.9
	CTSDG	1.52	4.97	8.27	0.97	2.99	7.38
	JPGNet	0.85	2.64	5.31	0.40	1.38	3.26
	本文方法	<b>0.76</b>	<b>2.46</b>	<b>5.12</b>	<b>0.34</b>	<b>1.30</b>	<b>3.14</b>

表2 实拍图像上不同方法的客观性能对比

Table 2 Objective performance comparison of different methods on real shooting images

评价指标	方法	掩码比例		
		10%~20%	20%~40%	40%~60%
PSNR ↑	EdgeConnect	31.642	23.871	19.469
	AOT-GAN	21.351	21.671	16.214
	CTSDG	31.560	23.217	19.189
	JPGNet	32.084	24.788	20.376
	本文方法	<b>32.560</b>	<b>24.984</b>	<b>20.611</b>
SSIM ↑	EdgeConnect	0.953 4	0.846 5	0.677 8
	AOT-GAN	0.850 6	0.833 9	0.632 8
	CTSDG	0.961 6	0.843 0	0.681 7
	JPGNet	0.953 3	0.842 1	0.685 2
	本文方法	<b>0.966 9</b>	<b>0.857 3</b>	<b>0.694 3</b>
$L_1$ ↓	EdgeConnect	0.74	2.60	5.40
	AOT-GAN	5.84	5.10	10.23
	CTSDG	0.57	2.39	5.17
	JPGNet	0.51	2.21	4.74
	本文方法	<b>0.43</b>	<b>2.01</b>	<b>3.97</b>

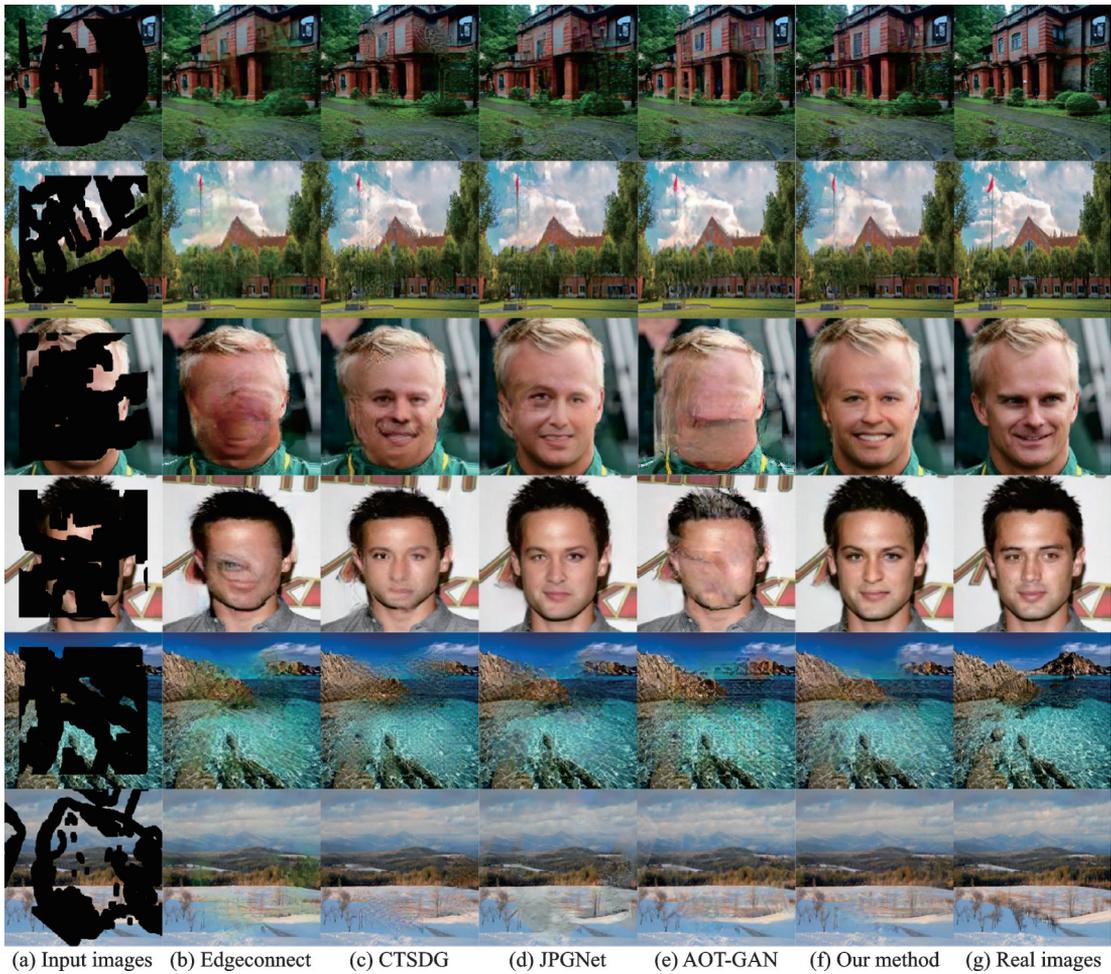
上相较于其他几种方法,本文方法实现了最佳的PSNR、SSIM和 $L_1$ 分数。在CelebA数据集上,由于人脸面部结构相对简单,在掩码为10%~20%与20%~40%这两个掩码比例较低的区间,采用结构先验方法的EdgeConnect能够准确地从非掩码区域中捕获丰富的结构信息帮助其进行修复,故在SSIM这一指标上取得了不错的效果。但本文方法在结合了多尺度外部空间特征先验和空间特征感知推理块的前提下,不仅在掩码为10%~20%与20%~40%区间取得了与EdgeConnect接近的效果,并在掩码40%~60%这一结构信息恢复困难的区间取得最佳的结果。此外,在PSNR与 $L_1$ 这两个指标上,本文方法相较于EdgeConnect分别取得了14%与65%的提升。

### 3.3 定性分析

图5为本文方法与目前最具代表性的几种方法的可视化结果对比。输入采用了不规则掩码生成的掩码破损图片,对比结果包含了Places2和CelebA两个数据集,以及实拍数据。可以观察到,EdgeConnect生成的结果在孔和背景之间产生了局部色差。在人脸修复方面,当掩码面积较大时,出现了扭曲的结构。CTSDG采用结构纹理相互引导的双流结构,但却生成了过于平滑的结果,结构不够清晰。AOT-GAN生成的结果未能很好地捕获人脸的结构纹理细节,在较为密集的掩码孔洞处也产生了棋盘式的伪影。JPGNet在面对同样密集的掩码空洞处也产生模糊现象,在人脸修复方面的细节(眼睛处)产生不够令人满意的视觉效果。与这些方法相比,本文方法在视觉上得到了令人满意的效果,在结构和纹理的恢复也较为详尽。

### 3.4 消融试验

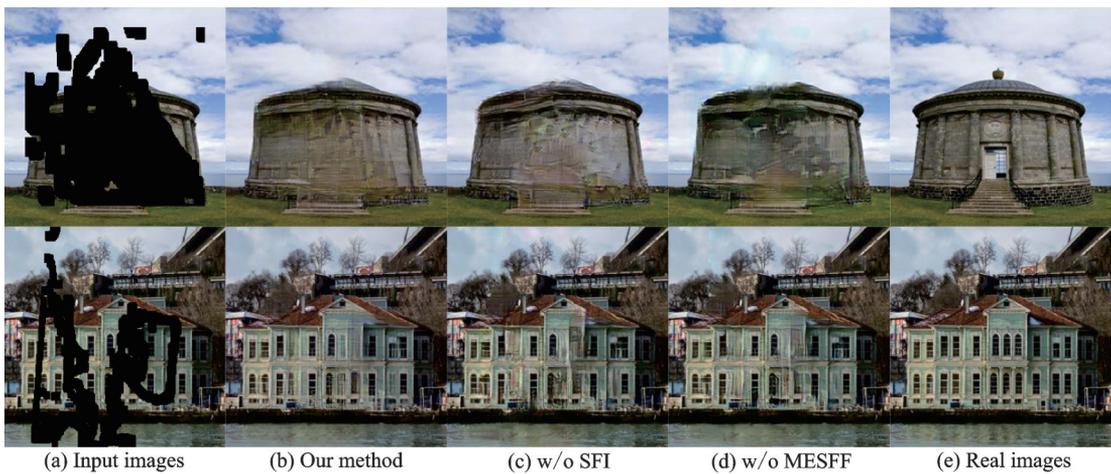
为了验证SFI和MESFF模块的有效性,本文在不同设置下进行3个实验,实验结果在Places2数据集上进行验证。(1)从原模型中去除多尺度外部空间特征融合模块,即w/o MESFF;(2)在原模型基础上去除空间特征感知推理块,即w/o SFI;(3)在原模型的基础上去除SFI模块和MESFF模块,并仅保有图像过滤模块进行训练,即w/o SFI & MESFF。结果如图6和表3所示。如图6所示,从主观角度来



(a) Input images (b) Edgeconnect (c) CTSDG (d) JPGNet (e) AOT-GAN (f) Our method (g) Real images

图 5 Places2和CelebA数据集以及实拍图像上不同方法的修复结果

Fig.5 Repair results of different methods on Places2 and CelebA datasets and real shooting images



(a) Input images (b) Our method (c) w/o SFI (d) w/o MESFF (e) Real images

图 6 Places2数据集上各消融算法对图像修复效果对比

Fig.6 Comparison of image inpainting effects of each ablation algorithm on Places2 dataset

表3 Places2数据集上各模块的消融实验

Table 3 Ablation experiments of each module on Places2 dataset

参数	方法	掩码比例		
		0%~20%	20%~40%	40%~60%
PSNR $\uparrow$	w/o MESFF	28.862	22.762	18.894
	w/o SFI	30.149	23.004	19.017
	w/o SFI & MESFF	27.489	21.663	17.773
	本文方法	<b>30.809</b>	<b>23.762</b>	<b>19.714</b>
SSIM $\uparrow$	w/o MESFF	0.940 6	0.818 8	0.659 6
	w/o SFI	0.943 3	0.821 3	0.661 3
	w/o SFI & MESFF	0.938 0	0.815 9	0.591 1
	本文方法	<b>0.947 5</b>	<b>0.835 5</b>	<b>0.685 0</b>
$L_1 \downarrow$	w/o MESFF	0.86	2.79	5.73
	w/o SFI	0.84	2.72	5.71
	w/o SFI & MESFF	0.89	2.90	6.04
	本文方法	<b>0.76</b>	<b>2.46</b>	<b>5.12</b>

看,不使用MESFF模块,模型对掩码周围的语义聚合能力减弱,产生严重的模糊效果;不适用SFI模块会使模型生成的语义难以和背景保持一致性,从而生成不和谐的结构和纹理。如表3所示,SFI和MESFF模块的联合使用,使模型在PSNR、SSIM和 $L_1$ 这3个指标上取得更好的得分。

#### 4 结束语

本文探索了一种新的双分支交互式图像修复框架,该框架利用神经网络的特征先验和图像预测滤波来联合修复损坏的图像。使用多尺度外部空间特征融合模块来进行语义层面的初步填充,并作为先验条件来丰富语义。然后,在特征推理和图像滤波分支部分引入空间特征感知推理块,它能够过滤掉分散注意力的特征,同时捕捉信息丰富的远距离图像上下文进行推理。通过在包括人脸图像和自然场景在内的数据集上进行实验,从采用不规则掩码的实验结果上来看,该模型方法能够帮助图片恢复合理结构以及细节丰富的纹理,与现有的先进生成对抗网络修复方法相比,在3个关键度量上取得了较高的得分,证明了其在修复质量上的优越性。

#### 参考文献:

- [1] 张桂梅,李艳兵.结合纹理结构的分数阶TV模型的图像修复[J].中国图象图形学报,2019,24(5):700-713.  
ZHANG Guimei, LI Yanbing. Image inpainting based on fractional-order TV model with combined texture structures[J]. Journal of Image and Graphics, 2019, 24(5): 700-713.
- [2] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Transactions on Graphics, 2009, 28(3): 24.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [5] NAZERI K, NG E, JOSEPH T, et al. Edgeconnect: Structure guided image inpainting using edge prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul, Korea (South): IEEE, 2019.
- [6] SONG Y, YANG C, SHEN Y, et al. SPG-Net: Segmentation prediction and guidance network for image inpainting[EB/OL]. (2018-08-06). <http://arxiv.org/abs/1805.03356v2>.

- [7] GUO Q, LI X, XU J F, et al. JPGNet: Joint predictive filtering and generative network for image inpainting[C]//Proceedings of the 29th ACM International Conference on Multimedia. [S.l.]: ACM, 2021: 386-394.
- [8] 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. [S.l.]: IEEE, 2021: 4692-4701.
- [9] LAHIRI A, JAIN A K, AGRAWAL S, et al. Prior guided Gan based semantic inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 13696-13705.
- [10] BAKO S, VOGELS T, MCWILLIAMS B, et al. Kernel-predicting convolutional networks for denoising Monte Carlo renderings[J]. ACM Transactions on Graphics, 2017, 36(4): 97.
- [11] GUO Q, SUN J, XU J F, et al. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining [C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2021: 1487-1495.
- [12] FU L, ZHOU C, GUO Q, et al. Auto-exposure fusion for single-image shadow removal[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 10571-10580.
- [13] WANG N, LI J, ZHANG L, et al. MUSICAL: Multi-scale image contextual attention learning for inpainting[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China: [s. n.], 2019: 3748-3754.
- [14] LIU H, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 4170-4179.
- [15] 唐小煜, 熊浩良, 黄锐珊, 等. 基于改进的U-NET和YOLOv5的绝缘子掩模获取与缺陷检测[J]. 数据采集与处理, 2021, 36(5): 1041-1049.  
TANG Xiaoyu, XIONG Haoliang, HUANG Ruishan, et al. Insulator mask acquisition and defect detection based on improved U-Net and YOLOv5[J]. Journal of Data Acquisition and Processing, 2021, 36(5): 1041-1049..
- [16] 孙琪, 翟锐, 左方, 等. 基于部分卷积和多尺度特征融合的人脸图像修复模型[J]. 计算机工程与科学, 2023, 45(2): 304-312.  
SUN Qi, ZHAI Rui, ZUO Fang, et al. Face image restoration model based on partial convolution and multi-scale feature fusion [J]. Computer Engineering and Science, 2023, 45(2): 304-312.
- [17] 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. [S.l.]: IEEE, 2021: 14134-14143.
- [18] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 29(7): 3266-3280.
- [19] ZHOU B, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464.
- [20] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2015: 3730-3738.

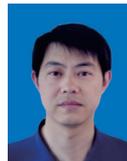
#### 作者简介:



吴鹏(1995-),男,硕士研究生,研究方向:图像修复与图像处理,E-mail:212240440@st.usst.edu.cn。



张孙杰(1988-),通信作者,男,副教授,研究方向:智能图像处理,模糊控制与滤波,E-mail:zhang\_sunjie@126.com。



王永雄(1970-),男,教授,博士生导师,研究方向:机器视觉和智能机器人等。



陈远峰(2000-),男,本科,研究方向:图像修复与图像处理。



覃海旺(2001-),男,本科,研究方向:图像修复与图像处理。