ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 2,Mar. 2025,pp. 401－416
DOI:10. 16337/j. 1004-9037. 2025. 02. 010

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于融合语义信息的上下文感知图像修复

祖　奕，张孙杰，吴　鹏，马悦恒

（上海理工大学光电信息与计算机工程学院，上海 200093）

摘　要：近年来，生成对抗网络广泛应用于图像修复领域并取得了不错的效果。但目前的方法并没有考虑在高分辨率图像（512×512）中会产生模糊的结构以及纹理的问题，这些问题主要来源于缺乏有效特征信息。针对此问题，提出一种将图像特征与语义信息相结合的生成对抗网络。主要基于语义信息，提出一种上下文感知的图像修复模型，该模型自适应地将语义信息与图像特征融合，并且提出自适应卷积替代传统卷积，以及在解码器后增添一个多尺度上下文聚合模块捕捉远距离信息来进行上下文推理。在Places2、CelebA-HQ、Paris Street View和Openlogo数据集上进行实验，实验结果表明，在L_1损失、峰值信噪比（PSNR）和结构相似度（SSIM）上所提方法与现有方法对比均有所提升。

关键词：图像修复；语义信息；图像特征；多尺度上下文特征聚合

中图分类号：TP391.4　　**文献标志码：**A

Context-Aware Image Restoration Based on Fused Semantic Information

ZU Yi, ZHANG Sunjie, WU Peng, MA Yueheng

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

Abstract: In recent years, generative adversarial networks have been widely used in the field of image restoration and have achieved good results. However, current methods do not consider problems of blurred structures and textures in high-resolution images (512×512), which mainly come from the lack of effective feature information. To address this problem, this paper proposes a generative adversarial network that combines image features with semantic information. Based mainly on semantic information, a context-aware image restoration model is proposed, which adaptively fuses semantic information with image features, and adaptive convolution is proposed to replace the traditional convolution, as well as a multi-scale context aggregation module is added after the decoder to capture long-distance information for contextual inference. Experiments are conducted on Places2, CelebA-HQ, Paris Street View, and Openlogo datasets, whose results show that the proposed method improves in terms of L_1 loss, peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) in comparison with the existing methods.

Key words: image restoration; semantic information; image feature; multi-scale contextual feature aggregation

基金项目：国家自然科学基金(61603255);上海市晨光计划项目(18CG52)。
收稿日期：2024-05-11；**修订日期：**2024-08-01

引 言

图像修复(又称图像补全或图像孔洞填充)是在缺失区域合成替代内容的任务,使得修改在视觉上是真实的且语义上是正确的。这项技术不仅限于图像修复,还可扩展到图像/视频裁剪、旋转、拼接、重新定位、重新组合、压缩、超分辨率和协调等任务,是计算机视觉和人工智能应用中的一个基本问题。尽管这项技术有很大的好处,但图像修复存在很多挑战,比如在高分辨率图像且不规则掩码中恢复合理的内容以及清晰的纹理。早期的文献一般通过基于扩散或基于像素的算法解决该问题[1-3]。这些方法在处理固定背景中的窄孔时表现良好,但在语义表达和复杂场景中的缺失内容推理方面却显得不足。

随着深度特征学习[4]和对抗训练[5]的出现,图像修复[6-8]取得了重大进展。早期基于深度学习的图像修复主要采用传统卷积。然而,传统卷积在高分辨率图像修复中并不适用,因为空间共享的卷积滤波器将所有输入像素或特征视为相同的有效元素,在进行孔洞填充时,每一层的输入包括了孔洞外的有效像素/特征和掩码区域中的无效像素/特征。传统的卷积操作对所有有效、无效和混合像素/特征使用相同的滤波器,这可能在不规则掩码上测试时产生视觉伪影,例如颜色差异、模糊以及孔周围明显的边缘响应等。如今伴随着生成对抗网络(Generative adversarial network,GAN)研究逐渐流行,这些深度图像修复模型通常建立在生成对抗网络上[5]。通过重建损失和对抗损失的联合优化[9],一些模型取得了显著成果。尽管如此,现有的深度生成修复方法仍然面临不少挑战,这些挑战使得其与现实世界的应用距离较远。例如,模型在处理自然场景时可能无法准确恢复复杂的纹理细节,导致生成图像的质量不理想。此外,在医学影像或卫星图像等领域,这些模型的泛化性不足,可能无法适应不同类型的图像。此外,由于远距离的上下文特征难以捕捉,图片常常会出现扭曲的结构和模糊的纹理,难以在复杂场景中推断出合理的内容[9-10]。综上所述,当前的主要困难包括:在复杂的缺失区域难以生成高分辨率的图像以及修复图像时出现的伪影;模型的泛化性不足,导致其与现实世界的脱节;在远距离图像上下文中推断出合理内容的挑战。

(1) 传统卷积产生视觉伪影。为了解决这一限制,Liu等[11]提出了部分卷积,其中卷积仅以有效像素为条件进行掩码和归一化。接下来,采用基于固定规则的掩码更新步骤,以更新下一层的有效位置。部分卷积将所有输入位置分类为无效或有效,并将零或一掩码乘以所有层的输入,然而,这种卷积方法也存在局限性。首先,盲目地将所有位置分类为有效或无效,可能忽略一些重要的信息。其次,在扩展到用户引导的图像修复时,用户在掩模内提供的稀疏草图中的像素位置应该被视为有效还是无效? 如何正确更新下一层的掩码? 最后,对于部分卷积,无效像素会逐层逐渐消失。使用部分卷积不能提供当前位置是否在缺失区域中,无法提供正确的上下文信息。

(2) 模型的泛化性问题。与生成任务不同,图像修复具有其特定的目标,即在复杂的背景和缺失区域的情况下,生成自然的图像并保持较高的泛化性。尽管深度生成网络在图像修复方面取得了重大进展,但远未解决上述挑战。例如,近期的研究RFRNet[4]在编码器-解码器网络上进行特征推理,并在公共数据集上实现了最先进的性能。然而,对于具有不同缺失区域的面部图像,很难产生逼真的修复结果。此外,Guo等[12]已经注意到基于生成的修复方法存在的问题,并提出了JPGNet,通过使用图像级预测过滤来减轻模糊。图像级预测滤波通过像素的相邻像素重建像素,根据输入自适应地估计滤波核。因此,JPGNet能够恢复局部结构,同时保持良好的泛化性,从而帮助RFRNet实现显著的质量改进。然而,这些方法往往基于特定的退化假设[13],因此缺乏对其他退化[14-16]的推广能力。随着时间的推移,对不依赖特定退化假设的盲目恢复方法的需求逐渐增加[17-21]。在这一趋势下,通过更复杂的退化模型来近似合成真实世界的退化。

(3) 远距离的上下文推理。为了推断丢失区域的合理内容,深度图像修复模型利用来自远距离图像上下文的特征。例如,提出了上下文注意力模块,以通过逐块匹配来对孔洞区域和图像上下文之间的关系进行建模[22]。但由于重复图案的问题,逐块匹配经常导致扭曲的结构。另一种方式是串联多个膨胀卷积层来捕获远距离的上下文。文献[7]广泛地讨论了多个膨胀卷积层倾向于对网络预定义图案的特征进行编码,但忽略捕获上下文推理所感兴趣的图案。

本文的贡献总结如下:

(1) 引入自适应卷积学习所有层中每个空间位置和每个通道的动态特征选择机制,显著提高了不规则掩码和输入图像的颜色一致性及修复质量。

(2) 提出了一种新的感知图像修复模型。自适应地将学习到的语义先验和局部图像特征结合在一个统一的生成网络中,从而为模型提供较高的泛化性。

(3) 引进了一个使用多种膨胀率的多尺度上下文特征聚合模块,同时通过聚合多个变换结果来捕获丰富的感兴趣图案,从而为缺失区域提供更好的上下文推理。

1　相关工作

图像修复的目的是用合理的内容填充图像中缺失的区域[1]。现有的方法可分为两类:传统方法和基于深度学习的方法。

1.1　基于传统方法的图像修复

在深度学习流行之前,图像修复的方法通常分为两类,分别是基于扩散和基于像素填充的算法[1-3,23-24]。基于扩散的算法基本上是沿等照度线方向,将边界处的上下文像素传播到孔洞区域[1,23,25-26]。然而,这种方法通常会引入与扩散相关性的问题,因此可能无法有效填补较大的缺失区域。并且这些方法在复杂场景下效果不理想,因为没有考虑图像区域的语义信息,这种技术无法合成在已知图像上下文中不存在的相似图像块内容。

1.2　基于深度学习的图像修复

深度特征学习[27-29]和对抗训练[30-31]的出现使图像修复取得了重大进展。与传统的方法相比,深度图像修复模型能够为复杂场景合成更合理的内容。为了推断较大的缺失区域丢失的内容,深度图像修复模型利用远距离图像上下文的特征。Pathak等[6]提出的上下文编码器(Context encoders,CE)是基于上下文信息的无监督特征语义修复方法。CE是一种开创性的深度修复模型,在紧凑的潜在特征空间中采用全连接层进行全局图像上下文编码,在人脸图像、街景等方面显示出较好的结果。然而其忽略了图像局部和全局的关系。在其基础上,Iizuka等[32]引入上下文局部鉴别器以生成全局和局部语义一致的修复图像。Zeng等[33]进一步采用了一个名为注意力转移网络的非局部模块来填充特征金字塔中的缺失区域。此外,Li等[4]提出了基于图像卷积特征的受损图像的递归重建。

然而,最近的文献大多采用生成对抗网络(GAN)来提高质量[7-8,33]。具体来说,GAN的鉴别器任务是区分修复图像与真实图像,而生成器则被优化以生成逼真的图像,试图欺骗鉴别器。通过生成器与鉴别器之间的对抗博弈,模型能够生成更真实的纹理[11]。Yu等[7]继承了PatchGAN[34]的优点,因为PatchGAN在图像翻译方面取得了巨大的成功[35-36]。具体来说,PatchGAN的修复旨在区分真实图像的补丁和修复结果的补丁。此外,将频谱归一化应用于每个层的卷积,以稳定GAN的训练[37]。生成先验擅长于捕捉图像的内在结构,从而能够生成遵循自然图像分布的图像。GAN网络的出现突显了生成先验在图像修复中的重要性。不同的方法使用这些先验,包括GAN编码器[38-39],或使用GAN作为图像修复[40]的核心模块。除了GAN网络,其他生成模型也可作为先验[17,41-44]。本文的工作主要集中于从完

整的图像得到的生成先验，然而，这些基于生成先验的方法受到所使用生成模型的规模限制，给进一步提高其有效性带来了挑战。

2 本文方法

对于高分辨率图像中不规则形状的缺失区域，合理推断其内容并生成逼真的图像是图像修复的关键挑战。如图1所示，本文所提出的方法基于生成对抗网络。具体来说，通过语义和图像两个层面入手，通过引入自适应卷积，分别提取图像特征和语义先验特征，然后，通过动态空间调整标准化模块进行特征融合，并利用多尺度上下文聚合模块进行特征聚合，最后将结果送往鉴别器进行判别。训练过程中，通过重建损失、对抗损失、感知损失和风格损失的联合优化来训练。

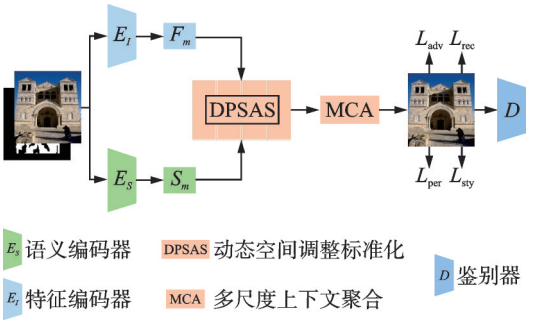


图1 本文方法整体框架

Fig.1 Overall framework of the proposed method

2.1 自适应卷积

特征自适应机制在视觉[45-46]、语言[47]、语义[48]及许多其他任务中得到了广泛的探索。例如，高速公路网络[21]利用特征来简化基于梯度的深度网络训练。SENet通过将每个通道与学习的S形值显式相乘来重新校准特征响应。WaveNets[7]通过采用特殊的特征机制来建模音频信号，从而获得更好的结果。

首先，详细解释在文献[7]中使用传统卷积不适合不规则掩码的图像修复任务的原因。对于传统卷积来说，所有空间位置(x,y)使用相同的滤波器进行输出。这种方法在图像分类和对象检测等任务中是有效的，因为输入图像的所有像素都是有效的，以滑动窗口的方式提取局部特征。然而，在图像修复中，输入包含有效像素和缺失区域中的无效像素。这会导致训练过程中的模糊性，并导致测试过程中的视觉伪影，如文献[7]所提到的颜色差异、模糊和明显的边缘响应等。

其次，针对该问题提出了部分卷积[7]，采用掩码和重新归一化步骤，使得卷积仅依赖于有效像素。

$$O_{x,y}=\begin{cases}\sum\sum Q\left(I\odot\dfrac{M}{\mathrm{sum}(M)}\right) & \mathrm{sum}(M)>0\\ 0 & \text{其他}\end{cases} \tag{1}$$

式中：M为对应的二进制掩码，$\odot$表示逐元素乘法，$O_{x,y}$表示输出，I表示输入，Q表示卷积滤波器。在每个部分卷积运算之后，需要掩码更新步骤来传播新的M，即

$$m'_{x,y}=\begin{cases}1 & \mathrm{sum}(M)>0\\ 0 & \text{其他}\end{cases} \tag{2}$$

式中：1表示位置(x,y)中的像素有效，0表示像素无效；其规则如下：当且仅当$\mathrm{sum}(M)>0$时$m'_{x,y}=1$。部分卷积[7]虽然提高了不规则掩码上的修复质量，但仍忽略了一些问题，例如不可学习的更新规则以及灵活性等等。换言之，部分卷积其更新规则比较死板，对不同的情况无法分辨。本质上部分卷积可被视为不可学习的单通道特征硬自适应卷积。

而本文提出一种可自适应的卷积以及新的掩码更新规则，如图2所示。代替部分卷积死板的掩码更新规则，自适应卷积自动从数据中学习新掩码，其公式为

$$\mathrm{Adapting}_{x,y}=\sum_i\sum_j Q_A\cdot I \tag{3}$$

$$\mathrm{Feature}_{x,y}=\sum_i\sum_j Q_B\cdot I \tag{4}$$

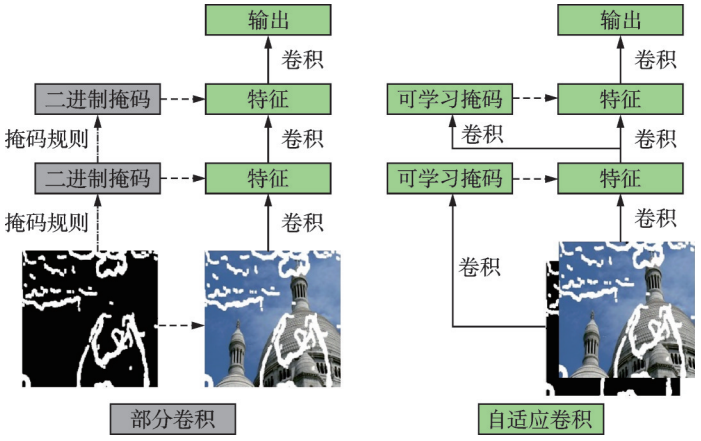


图2　自适应卷积与部分卷积掩码更新规则的区别

Fig.2　Difference between adaptive convolution and partial convolution mask update rules

$$Z_{y,x}=\phi(\text{Feature}_{x,y})\odot\sigma(\text{Adapting}_{x,y}) \tag{5}$$

式中:σ为Sigmoid函数,因此输出的自适应变量在0和1之间;ϕ可以是任何激活函数;Q_A和Q_B为两个不同的卷积滤波器;$Z_{y,x}$表示输出,I表示输入。

所提出的自适应卷积是一种通过学习每个通道和每个空间位置的动态特征选择机制来改进图像处理的卷积神经网络结构。这种卷积机制允许网络在处理图像时,基于输入的背景、掩模、草图以及某些通道中的语义分割信息,动态选择特征。

具体来说,自适应卷积在中间层引入了可视化的自适应变量。在深层中,自适应卷积也可学习在单独的通道中突出显示掩码区域和草图信息,以更好地生成修复结果,这意味着即使在网络深层,模型仍然能够区分不同通道中的语义信息,并根据需要突出显示或弱化某些特征。这样的动态特征选择机制有助于生成更好的修复结果,使网络能够更灵活地适应不同类型和内容的图像修复任务。

总的来说,本文提出的自适应卷积使模型能够更好地利用输入信息,并动态调整特征选择及其灵活的更新规则,从而提高图像修复的效果。

2.2　特征提取和融合

在本文的整个流程中,给定损坏的图像I_m和相应的掩码$M_K\in\{0,1\}$,其中1表示被掩码像素的位置,0表示无掩码像素的位置。首先使用两个不同的编码器E_I和E_S来学习图像特征F_m和语义先验S_m。这两个编码器采用2.1节提到的自适应卷积和残差块的结合组成,并且不共享任何参数。首先是图像编码器E_I,从可见内容中提取图像基础特征,利用自适应卷积构成的编码器,从损坏的图像中提取出图像特征F_m,即

$$F_m=E_I(I_m,M_K) \tag{6}$$

对于语义先验编码器E_S,其目标是学习缺失区域的完整语义先验,通常包含多个不同对象的全局上下文,并更加注重模型对缺失图像新风格的理解。E_S同样利用缺失图像和掩码作为输入来生成语义先验,即

$$S_m=E_S(I_m,M_K) \tag{7}$$

由于图像特征F_m和学习的语义先验S_m关注图像内容的不同方面,直接特征融合不仅会干扰可见区域的局部纹理,而且还会影响编码器的学习过程,为此,提出动态空间调整标准化模块将语义先验和图像特征完美融合,减少了不必要的影响。模型如图3所示。

在本文模型中，动态空间调整标准化(Dynamic programmatic space adjustment standardization, DPSAS)中的γ、β为从语义先验中学习的仿射变换参数。DPSAS模块利用实例归一化对图像特征F_m进行归一化，然后从S_m中提取两组不同的参数，对图像特征F_m进行仿射变换得到F'_m。至此，图像特征和语义先验特征融合在一起。

$$[\gamma,\beta]=\mathrm{DPSAS}(S_m) \tag{8}$$

$$F'_m=\gamma\cdot\mathrm{IN}(F_m)+\beta \tag{9}$$

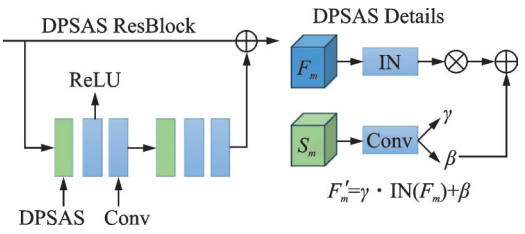


图3 DPSAS结构和细节

Fig.3 Structure and details of DPSAS

2.3 多尺度上下文聚合

为了能更有效地获取远距离的上下文以及识别出能够填补缺失区域的现有区域，设计了多尺度上下文特征聚合模块，如图4所示。

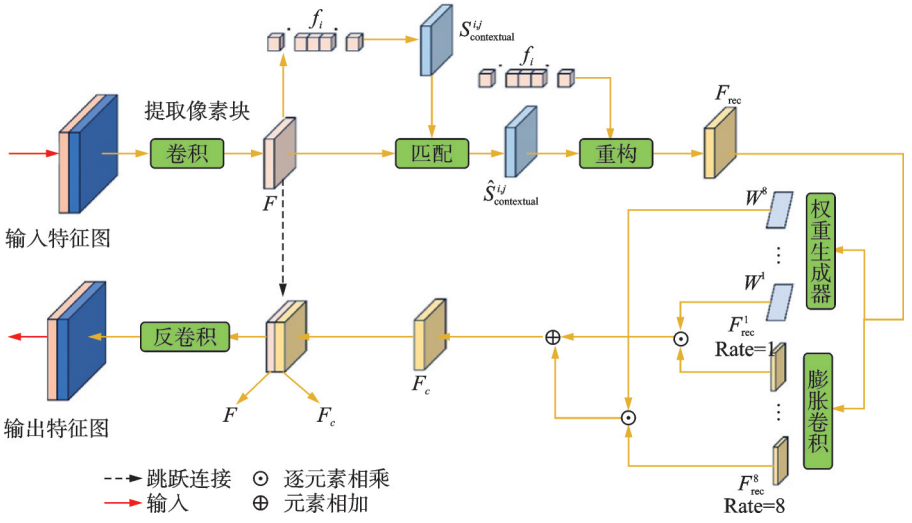


图4 多尺度上下文特征聚合模块

Fig.4 Multi-scale contextual feature aggregation module

该模块不仅能够强化图像局部特征之间的关联，还能保持整体的连贯性。受文献[49]的启发，不同于其固定尺度的补丁填充方法，本文采用各个尺度特征的聚合，以编码丰富的语义特征，从而很好地平衡准确性和复杂性，以应对尺度变化。

具体地说，对于特征图F，提取其3×3像素区域，并计算共余弦相似度

$$S^{i,j}_{\mathrm{contextual}}=\left\langle\frac{f_i}{\|f_i\|},\frac{f_j}{\|f_j\|}\right\rangle \tag{10}$$

式中：f_i与f_j分别代表特征图中的第i与第j区块。接着，通过对相似度应用Softmax函数便能为每个区块计算得到其注意力评分

$$\hat{S}^{i,j}_{\mathrm{contextual}}=\frac{\exp(S^{i,j}_{\mathrm{contextual}})}{\sum_{j=1}^{N}\exp(S^{i,j}_{\mathrm{contextual}})} \tag{11}$$

接下来，基于注意力图重新使用提取的块来重建特征图

$$\tilde{f}_i=\sum_{j=1}^{N}f_j\cdot\hat{S}^{i,j}_{\text{contextual}} \tag{12}$$

本文采取了一种创新的方法来重构特征图F_{rec},其中$\tilde{f}_i$代表其第i个分割区域。此过程通过卷积、逐通道Softmax以及反卷积操作依次实现。重点在于利用4组具有不同膨胀率的膨胀卷积层,这样做的目的是为了有效捕捉到不同尺度下的语义信息,以优化重构质量。

$$F^{k}_{\text{rec}}=\text{Conv}_k(F_{\text{rec}}) \tag{13}$$

式中:$\text{Conv}_k(\cdot)$表示具有k膨胀率的膨胀卷积层,$k\in\{1,2,4,8\}$。

为了更好地聚合多尺度语义特征,进一步设计了一个像素级权重图生成器G_w,其目的是预测像素级权重图。G_w由两个卷积层组成,内核大小分别为3和1,每个卷积层后面都有ReLU非线性激活,G_w的输出通道数设置为4。

$$W=\text{Softmax}(G_w(F_{\text{rec}})) \tag{14}$$

$$W^1,W^2,W^4,W^8=\text{Slice}(W) \tag{15}$$

因为G_w生成了一个包含4个通道的特征图,且每个通道可能对应于像素级权重图的不同方面或特性,因此,采用通道维度进行Softmax处理,对这些权重进行规范化,以得到一个综合的、在每个像素位置上都有意义的权重值。之后对其进行切片操作Slice,相当于提取了每一个通道上的特征并分别命名为W^1、W^2、W^4和W^8。

最后,通过加权聚合多尺度语义特征,生成细化特征图F_c,即

$$F_c=(F^1_{\text{rec}}\odot W^1)\oplus(F^2_{\text{rec}}\odot W^2)\oplus(F^4_{\text{rec}}\odot W^4)\oplus(F^8_{\text{rec}}\odot W^8) \tag{16}$$

2.4　损失函数和鉴别器

2.4.1　鉴别器

本文继承了PatchGAN[34]的鉴别器,因为其在图像修复领域取得了显著成功。具体来说卷积神经网络由几层标准卷积组成,每一层将特征映射的维度压缩至原先的一半。此预测模型接受经过修复或真实数据处理后的图像输入,并输出预测图。

2.4.2　损失函数

图像修复的目标是精确再现每个像素点并实现视觉上的逼真度。为此,选择了4个优化目标,即重构损失、感知损失[50]、风格损失[51]和GAN的对抗性损失。根据深度模型[52]现有的方法,首先,本文的目标是最小化L_1距离,以确保像素级的重建精度

$$L_{\text{rec}}=\|x-G(x\odot(1-m),m)\|_1 \tag{17}$$

鉴于感知损失[50]和风格损失[51]的有效性已经得到了广泛的验证[11],因此将这些因素纳入考虑,以提升感知重建的精确度。具体来说,感知损失的目标是减少修复图像与真实图像在激活图上的L_1距离,即

$$L_{\text{per}}=\sum_i\frac{\|\boldsymbol{\phi}_i(x)-\boldsymbol{\phi}_i(z)\|_1}{N_i} \tag{18}$$

式中:ϕ_i为源于预训练的网络第i层的激活图(VGG 19[53]),N_i为在ϕ_i中的元素的数量。类似地,风格损失L_{sty}被定义为真实值和修复结果其深度特征的Gram矩阵之间的L_1距离,即

$$L_{\text{sty}}=\mathbb{E}_i[\left\|\boldsymbol{\phi}_i(x)^{\text{T}}\boldsymbol{\phi}_i(x)-\boldsymbol{\phi}_i(z)^{\text{T}}\boldsymbol{\phi}_i(z)\right\|_1] \tag{19}$$

GAN网络的对抗性损失为

$$L^D_{\text{adv}}=\mathbb{E}_{z\sim P_z}\left[(D(z)-\sigma(m))^2\right]+\mathbb{E}_{x\sim P_{\text{data}}}\left[(D(x)-1)^2\right] \tag{20}$$

$$L_{\text{adv}}^{G}=\mathbb{E}_{z\sim P_z}\left[\left(D(z)-1\right)^2\right] \tag{21}$$

整个网络是通过这4个目标的联合优化来训练的,得出总体优化目标如下

$$L=\lambda_{\text{adv}}L_{\text{adv}}^{G}+\lambda_{\text{rec}}L_{\text{rec}}+\lambda_{\text{per}}L_{\text{per}}+\lambda_{\text{sty}}L_{\text{sty}} \tag{22}$$

根据多次实验以及综合性能得出的经验选择$\lambda_{\text{adv}}=0.01$,$\lambda_{\text{rec}}=1$,$\lambda_{\text{per}}=0.1$和$\lambda_{\text{sty}}=250$进行训练。

2.5 实施细节

模型的输入是一个形状为(N,3,512,512)的四维张量,其中N表示batchsize大小,3表示图像的RGB通道数,两个512分别表示图像的高度和宽度。接下来,输入经过图像编码器,主要使用自适应卷积替代传统卷积,主要使用3层自适应卷积(每层细节可参考2.1节的结构)进行下采样以及采集信息,通过在DPSAS(通过2个3×3卷积层和池化层点乘)融合语义信息,此时的维度变为(N,128,128,128),进行下一步输入到多尺度上下文聚合(Multiscale contextual aggregation, MCA)(主要由多个膨胀卷积组合而成)模块,最后模型的输出是一个形状与输入相同的四维张量,即(N,3,512,512)表示生成的图像。

对于模型训练,随机采样8个图像,并在每个小批量中随机创建相应的掩码。本文使用e^{-4}的固定学习率来训练学习器和生成器,使用$\beta_1=0$和$\beta_2=0.9$的Adam优化器进行训练。在实践中,使用在ImageNet[54]上预训练的VGG19[53]作为预训练网络,用于计算风格损失和感知损失。

通过自适应卷积替换传统卷积去除伪影问题,通过DPSAS提高模型泛化能力,最后通过MAC模块捕获更多的上下文信息,以提供更好的上下文推理。

3 实验部分

3.1 数据集

本文实验部分对所提方法在以下4个数据集上进行评估。

(1) Places 2挑战数据集[55]。Places 2数据集包含超过800万张图像,这些图像来自365种不同的场景,从中随机挑选5万张,其中3万张用来训练,1万张用来测试,1万张用来验证。

(2) CelebA-HQ[56]。CelebA是由香港科技大学汤晓鸥团队收集的人脸图像数据集(总共20万张图片),包括了人脸特征点(landmark)、人脸属性(attribute)等信息。CelebA-HQ是对CelebA数据集的一个升级版本,包含了从CelebA中挑选出的3万张高分辨率的人脸图像,其中抽取2万张用来训练,5千张用来测试,5千张用来验证。

(3) Paris街景。该数据集主要关注城市中的建筑物,共包含14 900个训练的数据集以及从巴黎街景中收集的1 000张测试图片。

(4) QMUL-OpenLogo。包含来自352个徽标类的27 083张图像,每张图像都由细粒度的徽标边界框注释。随机挑选了2万张用来训练,5千张用来测试。

这些数据集使得本文方法能够在自然场景和人脸场景上进行评估。具体来说,所有图像都被调整大小并裁剪为512×512,用于遵循标准高分辨率设置的训练和测试[36,57]。在实验中采用不规则掩码数据集进行训练和测试,使用不规则掩码数据集[11],该数据集已在许多作品中使用[55],以生成损坏的图像。掩码图像被分为3类(0%~20%、20%~40%和40%~60%)。

3.2 评价指标

本文列出了用于定量比较的所有客观指标来衡量图像修复的质量,以及使用的原因:

(1) L_1范数损失亦称平均绝对误差(Mean absolute error,MAE)作为经典回归损失函数被广泛采用于像素级重建任务中[7]。其数学本质在于计算预测图像与真实图像对应像素间绝对误差的均值,即通

过逐像素差异的绝对值之和进行归一化处理。图像修复领域采用L_1损失的意义主要是其对异常值不敏感,避免少数极端误差影响整体评价,直接度量生成图像与目标图像的像素级别差异,易于理解,对较小差异更加敏感,有助于生成更平滑的图像,避免过度惩罚。

(2) 峰值信噪比(Peak signal-to-noise ratio, PSNR)是许多修复方法使用的经典图像质量评估之一[7]。

(3) 结构相似性指数(Structural similarity, SSIM)[58]从亮度、对比度和结构方面将修复结果与原始图像进行比较。

3.3　对照组

本文实验中比较了6种在图像修复领域比较前沿的方法,即CA[7]、PConv[11]、GateConv[49]、EdgeConnect[8]、CoModGAN[59]和AOT-GAN[60]。

3.4　比较结果

本文实验定量比较了本文方法与4个公共数据集上的5种最经典的修复方法,结果分别如表1、2所示。观察可得,与其他竞争方法相比,本文方法在所有数据集和掩码比上都获得了更好的PSNR、SSIM和L_1得分。与AOT-GAN相比,在Places 2数据集上0%～20%的掩码占比下实现了11.37%的相对较

表1　在Places 2和CelebA-HQ数据集上的定量实验对比结果

Table 1　Comparison of quantificative experimental resuits on Places 2 and CelebA-HQ datasets

评价指标	对比方法	Place 2			CelebA-HQ		
		掩码占比 0%～20%	掩码占比 20%～40%	掩码占比 40%～60%	掩码占比 0%～20%	掩码占比 20%～40%	掩码占比 40%～60%
PSNR/dB ↑	CA	24.964	22.762	17.215	25.760	23.683	19.223
	PConv	27.901	25.316	18.919	28.791	26.340	21.126
	EdgeConnect	27.630	24.751	18.374	28.511	25.752	20.517
	GateConv	26.109	23.389	17.061	26.942	24.334	19.051
	CoModGAN	26.848	23.715	17.693	27.704	24.674	19.757
	AOT-GAN	28.459	25.494	19.109	29.367	26.525	21.338
	Ours	31.696	26.695	19.546	33.052	28.562	22.389
SSIM ↑	CA	0.854 7	0.795 2	0.689 3	0.869 7	0.803 1	0.723 4
	PConv	0.875 8	0.828 0	0.718 7	0.891 2	0.836 2	0.754 3
	EdgeConnect	0.878 4	0.824 0	0.700 6	0.893 8	0.832 1	0.735 3
	GateConv	0.816 9	0.803 6	0.664 6	0.831 2	0.811 5	0.697 5
	CoModGAN	0.877 6	0.820 8	0.698 6	0.892 9	0.828 9	0.733 2
	AOT-GAN	0.886 4	0.843 0	0.723 8	0.901 9	0.851 3	0.759 6
	Ours	0.927 5	0.876 9	0.748 2	0.941 1	0.893 1	0.803 2
$L_1/10^{-2}$ ↓	CA	5.25	5.86	10.15	4.68	5.63	10.05
	PConv	3.24	3.94	7.63	2.90	3.72	7.66
	EdgeConnect	3.25	3.98	7.74	2.89	3.76	7.73
	GateConv	3.92	4.61	8.09	3.50	4.43	8.01
	CoModGAN	3.32	4.27	8.38	2.95	4.10	8.29
	AOT-GAN	3.02	3.62	7.10	2.69	3.48	7.03
	Ours	2.22	2.88	6.93	1.47	2.22	5.22

表 2 在 Paris 街景和 OpenLogo 上的定量实验比对结果

Table 2 Comparison of quantitative experimental results on Paris street view and OpenLogo

评价指标	对比方法	OpenLogo			Paris
		掩码占比 0%～20%	掩码占比 20%～40%	掩码占比 40%～60%	随机掩码
PSNR/dB ↑	CA	20.052	18.579	13.987	21.628
	PConv	22.411	20.664	15.372	24.055
	EdgeConnect	22.194	20.202	14.928	23.292
	GateConv	20.972	19.090	13.862	22.228
	CoModGAN	21.566	19.357	14.375	22.534
	AOT-GAN	22.860	20.809	15.526	24.224
	Ours	27.227	23.616	16.165	27.313
SSIM ↑	CA	0.740 9	0.683 8	0.575 1	0.744 1
	PConv	0.759 2	0.712 0	0.299 7	0.774 8
	EdgeConnect	0.761 5	0.708 6	0.584 6	0.771 0
	GateConv	0.708 2	0.691 0	0.554 5	0.751 9
	CoModGAN	0.760 7	0.705 8	0.582 9	0.768 0
	AOT-GAN	0.768 4	0.724 9	0.603 9	0.788 8
	Ours	0.875 0	0.826 3	0.673 8	0.864 8
$L_1/10^{-2}$ ↓	CA	11.44	12.91	19.84	10.75
	PConv	7.37	8.55	15.12	7.10
	EdgeConnect	7.36	8.60	15.26	7.17
	GateConv	8.90	10.16	15.82	8.46
	CoModGAN	7.53	9.40	16.37	7.83
	AOT-GAN	6.85	7.97	13.88	6.64
	Ours	4.71	5.40	11.35	4.96

高的 PNSR。实验结果表明，该方法在高分辨率情况下恢复方面具有明显的优势。在 CelebaA-HQ 数据集上，在 0%～20% 掩码比率下获得了比 AOT-GAN 相对低 54.65% 的 L_1 损失以及相对高 4.34% 的 SSIM。结果表明本文方法在感知恢复方面已取得较大的进展。在不同的数据集和掩码比率上的一致优势表明具有很高的泛化能力。

定性比较 4 个数据集（CelebA、Place 2、Paris 和 OpenLogo）上提供的 4 个案例的可视化结果（见图 5），可发现，一方面本文方法生成的图像更自然、更逼真；另一方面，其他方法引入了许多伪像，如结构扭曲和大片缺失区域的模糊。例如，对于 CA、PConv、EdgeConnect、GateConv 和 AOT-GAN 中的局部结构，都未能恢复细节结构。相比之下，本文方法可准确地完成所有细节。

3.4.1 结果分析和讨论

由实验结果可以发现通过本文的网络结构改进之后，对于高分辨率下的图像修复效果有了显著提升。针对伪影问题，通过自适应卷积来代替传统卷积。针对泛化性问题，采用将语义信息和图像特征进行了融合来提升泛化性。最后，对于远距离上下文信息，通过多尺度聚合上下文聚合模块改善捕捉信息能力，使获取上下文信息更加快捷。在 3 个评价指标上和其他方法相比都有不同的提升。然而，本文方法依旧存在一些局限性，虽然提升了泛化性，但对于特定领域或特定类型的图像时，对泛化性的提升可能有限。同时在实际应用中，不同的环境和设备，其性能可能受到影响。未来的研究可针对这些方面进行进一步优化和改进，以提升方法的实用性和普适性。

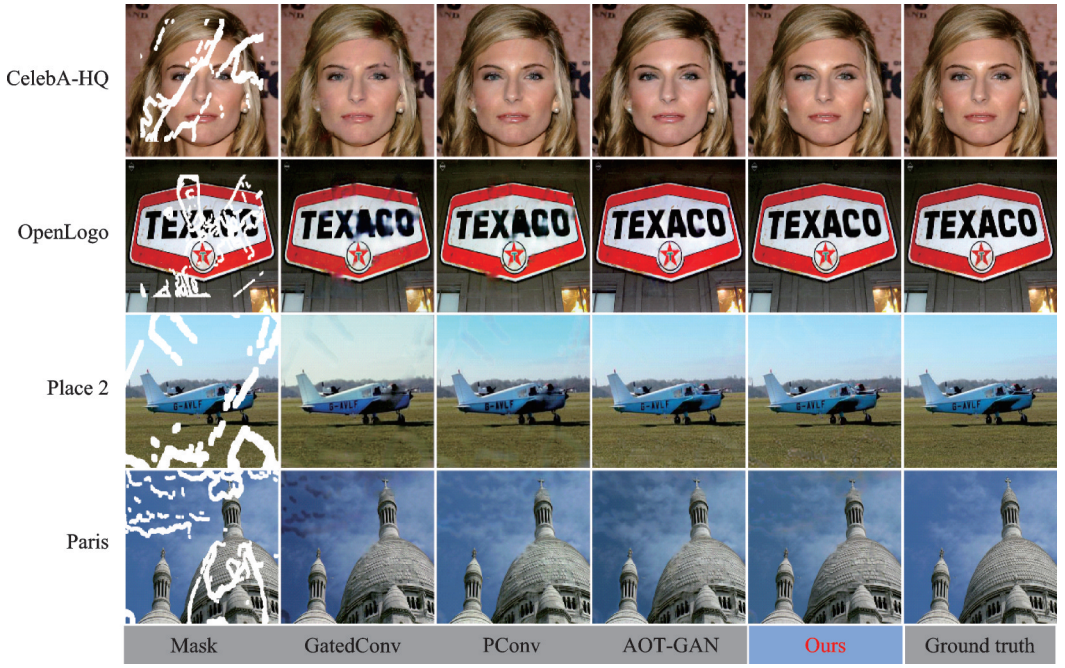


图5　在4个数据集上不同方法的定性比较

Fig.5　Qualitative comparisons of different methods on four datasets

3.4.2　成本分析

本实验在单GPU下进行,训练3万次,batchsize为4,学习率为0.000 1。训练一次完整的实验需要5 d。在进行本文方法实验时,按照实验的配置无论是时间成本还是效果都优于其他方法。表3所示均为在显卡3060下进行实验的结果。

表3　其他方法成本的直观比较

Table 3　Visual comparison of costs with alternative methods

方法	时间成本/h	效果(SSIM)↑
CA	80	0.760 9
PConv	120	0.779 2
CoModGAN	190	0.810 7
AOT-GAN	180	0.838 4
Ours	120	0.875 0

3.5　消融实验

本文进行了6组消融实验(均在CelebA-HQ数据集下)以验证本网络的3个组成部分的有效性,即自适应卷积、特征提取和融合和多尺度上下文特征聚合模块。如表4所示,每组消融实验均基于之前的组件进行。

表4　本文各模块消融实验

Table 4　Ablation experiments on each of modules proposed in this paper

模型	多尺度上下文聚合模块	自适应卷积	特征提取融合	$L_1/10^{-2}$↓	PSNR/dB↑	SSIM↑
1				4.62	27.714	0.868 0
2	√			3.20	29.281	0.881 0
3		√		2.54	29.724	0.878 0
4			√	2.18	32.854	0.947 2
5	√	√		1.87	33.192	0.916 1
6	√	√	√	1.46	35.674	0.963 5

3.5.1 多尺度上下文聚合模块

为了验证上下文聚合模块MCA能提高局部特征之间的相关性和整体图像的一致性。本文使用相同网络深度模型作为基准进行比较,删除了自适应卷积和感受野与先验信息相结合的滤波,进行了广泛的比较多尺度上下文聚合模块的设计,使用不同数量的分支和膨胀率。

表5中的定量结果表明具有更多样化扩张率的更大数量的分支能够实现更大的改进,且可在很大程度上丰富MCA模块,尤其是对于高分辨率图像的修复。

表5 不同膨胀率的定量比较

Table 5 Quantitative comparison of different expansion rates

分支	膨胀率	$L_1/10^{-2}$ ↓	PSNR/dB ↑	SSIM ↑
1	1	3.56	28.23	0.853
1	2	3.50	28.39	0.856
2	1,2	3.49	28.30	0.858
2	2,8	3.28	29.01	0.877
3	1,2,8	3.28	29.10	0.878
4	1,2,4,8	**3.20**	**29.28**	**0.881**

定性结果根据表5中的定量比较,本文在最终模型中使用4个分支,每个分支的扩张率分别为1、2、4和8。将此设置与膨胀率为2的单分支模块(即EdgeConnect[54]中使用的残差块)进行对比试验。定性结果如图6所示。通过引入多尺度上下文聚合模块,提高了局部特征与整体图像之间一致性的相关性。如图6所示,可将单分支且膨胀率为1的设定视为原始情况,可看出没有MCA的模型渲染的图像质量较低,并且纹理填充对结构噪声很敏感。表5中的量化结果也验证了其必要性.拥有MCA模块的3个评价指标相较其他分支都有显著提升。

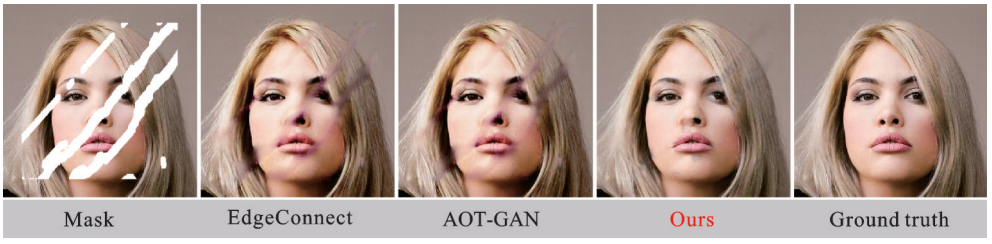


图6 与不同膨胀率的方法对比

Fig.6 Comparison of methods with different dilation rates

3.5.2 自适应卷积

(1) 定量结果

本文提出了用于图像修复网络的自适应卷积代替部分卷积掩码更新规则。为了验证自适应卷积的有效性,将其与几个对应物进行比较,即传统卷积和部分卷积。具体来说通过替换编码器中的卷积来比较这些方法。

(2) 定性结果

图7中的视觉比较表明,利用自适应卷积,模型能够生成比其他卷积更逼真的面部纹理。这是因为

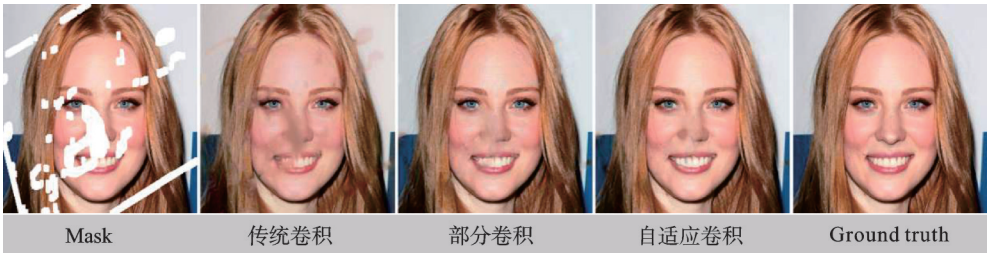


图7 传统卷积、部分卷积及自适应卷积的定性实验结果对比

Fig.7 Comparison of qualitative experimental results of conventional convolution, partial convolution and adaptive convolution

所提出的自适应卷积提供了输入特征和学习的残余特征之间的可学习的空间变化连接,同时保留孔外有效像素的低级别细节。

通过使用自适应卷积用于在每个编码器中聚合输入特征和学习的残差特征。具体来说,自适应卷积概括了部分卷积。表6中的定量比较显示,本文的自适应卷积在所有指标方面都优于其他连接。图7中的视觉比较表明,利用自适应卷积连接,模型能够生成比其他残差融合方法更逼真的面部纹理。这是因为所提自适应卷积在输入特征和学习的残差特征之间提供了可学习的空间变化连接。这种连接能够更新缺失区域内的特征,同时保留孔外有效像素的低级别细节,相比传统卷积和部分卷积都有提升。

表6　不同卷积定量实验结果对比

Table 6　Comparison of quantitative experimental results of different convolutions

方法	$L_1/10^{-2}\downarrow$	PSNR/dB↑	SSIM↑
传统卷积	2.90	28.857	0.866 1
部分卷积	2.80	29.100	0.867 5
自适应卷积	**2.61**	**29.460**	**0.870 9**

3.5.3　特征提取和融合

(1) 定量结果

为了验证结合的有效性,考虑了3种变体:只有图像特征信息、语义信息和图像特征直接融合以及本文结构。从表6中可观察到,通过本文提出的结合方式,CelebA-HQ数据集上和3个缺失大小的所有指标上都比其他方法得到更好的分数。

(2) 定性结果

实验结果表明,普通的图像特征能有效地完成小的缺失区域,而在处理大面积方面特别是高分辨率效果较差。图8比较了其与本文结构的实验效果,可以看出普通图像特征明显失去了很多细节,而本文结构产生了丰富的细节与自然结构。

图8　只有图像特征、直接融合以及本文所提结构的结果对比

Fig.8　Result comparison of structures with only image features and with direct fusion, and the proposed structure

不使用DPSAS模块,而是直接将学习的语义先验和来自图像编码器的特征串联起来形成残差块的输入,这是以前的两阶段生成框架通常使用的。表7中的结果显示,连接的直接操作会导致所有3个评估指标的减少。而本文使用DPSAS让语义信息通过特定的方式与图像特征进行融合,减少了不必要的损失,可以有效提升评估指标以及性能。

表7　不同方式定量实验结果对比

Table 7　Comparison of quantitative experimental results of different modalities

方式	$L_1/10^{-2}\downarrow$	PSNR/dB↑	SSIM↑
只有图像特征	4.82	27.023	0.929 8
直接融合	1.91	34.077	0.958 3
本文结构	**1.80**	**34.319**	**0.962 7**

4　结束语

本文提出了一种可融合语义信息的图像修复网络,解决了在高分辨率图像修复中遇到的伪影、泛化性和上下文理解问题。首先,在编码器中使用自适应卷积替换传统卷积来过滤过多的无效信息;其

次,提出了一种可特征融合的方法将图像特征和语义特征融合在一起来增强泛化性;最后,通过一个可聚合上下文特征的模块来更好地聚合上下文信息,并且验证了其有效性。相较于其他方法,本文方法更显著的优势在于其优秀的泛化能力以及高分辨率下依旧可进行无误的图像修复。在广泛使用的公共数据集上验证和训练模型,这些数据集可能涵盖了部分真实世界场景,本文方法在4个公共数据集上的性能优于很多经典的方法。

虽然本文提出的方法通过融合其语义信息在跨不同架构的不同恢复网络的性能方面表现出了显著的效果,但改进的程度似乎在不同的实验中有所不同。一些实验显示了明显的增强,而另一些实验则没有。这种差异与目标网络的容量和目标任务的复杂性相关。

未来将深入研究更有效的方法,专门帮助目标恢复任务。目标是使用一个量身定制的蒸馏框架来推导出精细化的恢复特征先验,最终超越现有的上限。

参考文献:

[1] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. [S.l.]: ACM, 2000: 417-424.

[2] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Transactions on Graphice, 2009, 28(3): 24.

[3] CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[4] LI J, WANG N, ZHANG L, et al. Recurrent feature reasoning for image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 7760-7768.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27.

[6] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition. [S.l.]: IEEE, 2016: 2536-2544.

[7] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 5505-5514.

[8] NAZERI K, NG E, JOSEPH T, et al. Edgeconnect: Generative image inpainting with adversarial edge learning[J]. arXiv preprint arXiv:1901.00212, 2019.

[9] YANG C, LU X, LIN Z, et al. High-resolution image inpainting using multi-scale neuralpatch synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 6721-6729.

[10] WANG Y, TAO X, QI X, et al. Image inpainting via generative multi-column convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. [S.l.]: ACM, 2018, 31.

[11] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCV, 2018: 85-100.

[12] GUO Q, LI X, JUEFEI-XU 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.

[13] LIU A, LIU Y, GU J, et al. Blind image super-resolution: A survey and beyond[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5461-5480.

[14] GU J, MA X, KONG X, et al. Networks are slacking off: Understanding generalization problem in image deraining[J]. Advances in Neural Information Processing Systems, 2023, 36: 28565-28584.

[15] LIU Y, LIU A, GU J, et al. Discovering distinctive "semantics" in super-resolution networks[J]. arXiv preprint arXiv: 2108.00406, 2021.

[16] ZHANG R, GU J, CHEN H, et al. Crafting training degradation distribution for the accuracy-generalization trade-off in real-world super-resolution[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2023: 41078-41091.

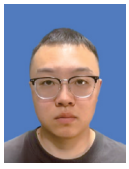
[17] CHEN C, SHI X, QIN Y, et al. Real-world blind super-resolution via feature matching with implicit high-resolution priors [C]//Proceedings of the 30th ACM International Conference on Multimedia. [S.l.]: ACM, 2022: 1329-1338.

[18] HUI Z, LI J, WANG X, et al. Learning the non-differentiable optimization for blind super-resolution[C]//Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 2093-2102.
[19] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 1833-1844.
[20] LIANG J, ZHANG K, GU S, et al. Flow-based kernel prior with application to blind super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 10601-10610.
[21] LIANG J, ZENG H, ZHANG L. Efficient and degradation-adaptive network for real-world image super-resolution[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2022: 574-591.
[22] MILDENHALL B, BARRON J T, CHEN J, et al. Burst denoising with kernel prediction networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 2502-2510.
[23] BERTALMIO M, VESE L, SAPIRO G, et al. Simultaneous structure and texture image inpainting[J]. IEEE Transactions on Image Processing, 2003, 12(8): 882-889.
[24] SUN J, YUAN L, JIA J, et al. Image completion with structure propagation[M]. [S.l.]: ACM, 2005: 861-868.
[25] BALLESTER C, CASELLES V, VERDERA J, et al. A variational model for filling-in gray level and color images[C]//Proceedings Eighth IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2001, 1: 10-16.
[26] BERTALMIO M, BERTOZZI A L, SAPIRO G. Navier-Stokes, fluid dynamics, and image and video inpainting[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2001: 1-8.
[27] ZHENG H, YANG H, FU J, et al. Learning conditional knowledge distillation for degraded-reference image quality assessment[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 10242-10251.
[28] 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. [S.l.]: IEEE, 2016: 770-778.
[29] XUE H, LIU B, YANG H, et al. Learning fine-grained motion embedding for landscape animation[C]//Proceedings of the 29th ACM International Conference on Multimedia. [S.l.]: ACM, 2021: 291-299.
[30] YANG F, YANG H, FU J, et al. Learning texture transformer network for image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 5791-5800.
[31] ZENG Y, YANG H, CHAO H, et al. Improving visual quality of image synthesis by a token-based generator with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 21125-21137.
[32] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics (ToG), 2017, 36(4): 1-14.
[33] ZENG Y, FU J, CHAO H, et al. Learning pyramid-context encoder network for high-quality image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 1486-1494.
[34] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 1125-1134.
[35] XIONG W, YU J, LIN Z, et al. Foreground-aware image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 5840-5848.
[36] REN Y, YU X, ZHANG R, et al. Structureflow: Image inpainting via structure-aware appearance flow[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 181-190.
[37] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[J]. arXiv preprint arXiv:1802.05957, 2018.
[38] CHAN K C K, WANG X, XU X, et al. GLEAN: Generative latent bank for large-factor image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 14245-14254.
[39] ZHU J, ZHAO D, ZHANG B, et al. Disentangled inference for GANs with latently invertible autoencoder[J]. International Journal of Computer Vision, 2022, 130(5): 1259-1276.
[40] WANG X, LI Y, ZHANG H, et al. Towards real-world blind face restoration with generative facial prior[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 9168-9178.
[41] JO Y, YANG S, KIM S J. SRFlow-DA: Super-resolution using normalizing flow with deep convolutional block[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 364-372.
[42] ZHANG Y, SHI X, LI D, et al. A unified conditional framework for diffusion-based image restoration[C]//Proceedings of Advances in Neural Information Processing Systems. [S.l.]: ACM, 2024, 36.

[43] ZHAO Y, SU Y C, CHU C T, et al. Rethinking deep face restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 7652-7661.
[44] ZHOU S, CHAN K, LI C, et al. Towards robust blind face restoration with codebook lookup transformer[J]. Advances in Neural Information Processing Systems, 2022, 35: 30599-30611.
[45] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks[J]. arXiv preprint arXiv:1505.00387, 2015.
[46] WANG H, WANG Y, ZHANG Q, et al. Gated convolutional neural network for semantic segmentation in high-resolution images[J]. Remote Sensing, 2017, 9(5): 446.
[47] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2017: 933-941.
[48] OORD A, DIELEMAN S, ZEN H, et al. WaveNet: A generative model for raw audio[J]. arXiv preprint arXiv:1609.03499, 2016.
[49] 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. [S.l.]: IEEE, 2019: 4471-4480.
[50] 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. [S.l.]: IEEE, 2016: 2414-2423.
[51] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]//Proceedings of Computer Vision—ECCV 2016: 14th European Conference. Amsterdam: Springer, 2016: 694-711.
[52] YAN Z, LI X, LI M, et al. Shift-Net: Image inpainting via deep feature rearrangement[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCV, 2018: 1-17.
[53] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
[54] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2009: 248-255.
[55] 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.
[56] 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.
[57] SONG Y, YANG C, LIN Z, et al. Contextual-based image inpainting: Infer, match, and translate[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCV, 2018: 3-19.
[58] YAN W Q, WANG J, KANKANHALLI M S. Automatic video logo detection and removal[J]. Multimedia Systems, 2005, 10: 379-391.
[59] ZHAO S, CUI J, SHENG Y, et al. Large scale image completion via co-modulated generative adversarial networks[J]. arXiv preprint arXiv:2103.10428, 2021.
[60] 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.

作者简介:



祖奕(2000-),男,硕士研究生,研究方向:图像修复与图像处理,E-mail: 212240440@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。

吴鹏(1995-),男,硕士研究生,研究方向:图像修复与图像处理。

马悦恒(2002-),男,硕士研究生,研究方向:计算机视觉。

(编辑:夏道家)