中图法分类号:TP18;TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2026)04-1029-15

论文引用格式:Ye X Y, Sui M C, Tan R J, Jiang D Q and Chen H H. 2026. Semantic correction-based image inpainting under semisupervised learning. Journal of Image and Graphics, 31(4):1029-1043(叶学义, 睢明聪, 谭瑞洁, 蒋德琦, 陈华华. 2026. 半监督学习下基于语义校正的图像修复. 中国图象图形学报, 31(4):1029-1043)[DOI:10.11834/jig.250305]

半监督学习下基于语义校正的图像修复

叶学义*,睢明聪,谭瑞洁,蒋德琦,陈华华

杭州电子科技大学通信工程学院,杭州　310018

摘　要:**目的**　针对现有语义引导图像修复方法因其单向性而存在的潜在错误累积问题,提出交互式图像修复框架,通过在图像修复与语义分割模型之间构建双向反馈与校正机制,提升修复质量。**方法**　构建"初始修复—半监督语义重校正—精细修复"三阶段框架,核心为"半监督语义重校正"模块:利用初始修复结果向语义分割模型反馈信息,结合跨图像语义一致性来校正语义分割结果;引入半监督学习机制,融合有标签和无标签数据进行语义分割模型的训练,减少对真实标签的依赖。**结果**　在公开的CelebA-HQ(CelebA-high quality)数据集和Cityscapes数据集上进行实验,并与现有先进方法进行比较。实验结果表明,该方法在学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)、峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity index measure, SSIM)指标上综合表现更优:在CelebA-HQ数据集上,相较于MDTG(mutual dual-task generator)算法,本文方法的LPIPS降低5.88%、PSNR提升0.52%、SSIM提升0.22%;在Cityscapes数据集上,相较于MDTG算法,本文方法的LPIPS降低6.15%、SSIM提升1.58%、PSNR提升0.70%。消融实验的结果进一步验证了校正机制的有效性。**结论**　该项工作成功地在修复和语义分割模型之间建立了交互式反馈机制,显著提高了图像修复的质量;尤其在处理复杂纹理和语义偏差时,表现出较好的修复效果;同时半监督学习策略的引入有效减少了对人工标注语义图的依赖。

关键词:图像修复;语义分割;半监督学习;跨图像语义一致性;生成对抗网络(GAN)

Semantic correction-based image inpainting under semisupervised learning

Ye Xueyi*, Sui Mingcong, Tan Ruijie, Jiang Deqi, Chen Huahua

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou　310018, China

Abstract: **Objective** The rapid advancement of deep learning techniques has resulted in significant progress in the field of image inpainting, particularly in utilizing semantic structures to guide the inpainting process. Semantic structures, such as semantic label maps, have become increasingly popular because of their ability to provide valuable contextual information about missing or damaged regions of images. These semantic structures help guide the inpainting algorithm to restore missing content in a way that aligns well with the overall context and semantic meaning of the image. Traditional inpainting models often rely on pixel-level information and texture-based techniques to reconstruct missing regions. However, they may struggle to maintain consistent semantic structures, particularly when dealing with complex images or large damaged areas. Recent advances have shown that combining semantic information with inpainting tasks can significantly improve the quality and realism of image restoration. In particular, semantic segmentation, which can be leveraged to guide the inpainting process, has become an essential tool for extracting meaningful context from an image. Although semantic segmentation

收稿日期:2025-07-28;**修回日期**:2025-10-21;**预印本日期**:2025-10-28

*通信作者:叶学义　xueyiye@hdu.edu.cn

基金项目:国家自然科学基金项目(U19B2016)

Supported by:National Natural Science Foundation of China(U19B2016)

models provide rich, semantic information that is crucial for accurate image inpainting, many existing methods remain unidirectional. These models often rely solely on pretrained semantic segmentation networks to provide guidance, without considering feedback from the inpainting model itself to improve or refine the segmentation. This limitation can prevent the model from fully correcting semantic errors, resulting in inadequately accurate inpainting outcomes. We propose a novel image inpainting method that introduces feedback correction within a semisupervised learning framework to overcome the aforementioned challenges. This approach allows for a bidirectional interaction between the inpainting and segmentation models, thereby enabling them to refine each other iteratively and ultimately enhance the final inpainting results. **Method** The proposed method follows a three-stage progressive image inpainting algorithm, which includes coarse inpainting, semisupervised semantic correction, and fine inpainting. Each of these stages plays a critical role in improving the quality and accuracy of the final image restoration, particularly in scenarios where semantic biases and texture errors are prevalent. In the first stage, coarse inpainting is performed. The goal of this module is to create an initial pixel-level reconstruction of the missing areas. At this stage, the algorithm focuses on restoring basic structures and textures within the damaged regions. Although the inpainting process may still introduce some semantic errors or distortions, it provides a solid foundation for further refinement. This coarse inpainting serves as the basis for the following stages, which aim to address semantic inconsistencies and improve texture details. The second stage involves semisupervised semantic correction. Here, we leverage a cross-image semantic consistency strategy to enhance the semantic understanding of the damaged regions and improve the accuracy of segmentation. This module employs semisupervised learning techniques to generate pseudo-labels for unlabeled images, which are then used to refine the semantic segmentation of the inpainted areas. The segmentation network can correct any semantic errors introduced during the initial inpainting stage by incorporating feedback from the inpainting model. This feedback loop helps ensure that the inpainted regions align closely with the rest of the image in terms of texture and semantic content. The third and final stage is fine inpainting. In this stage, the semantic segmentation model——now trained with the refined semantic labels——guides the inpainting model to restore the missing content with high precision. The improved semantic labels allow the inpainting model to generate accurate pixel-level details and textures, thereby ensuring that the final result is semantically consistent with the original image. This stage refines the image restoration process by incorporating the corrected segmentation and the inpainted content, thereby resulting in a realistic and seamless reconstruction. **Result** We conducted extensive experiments using two publicly available datasets, namely, CelebA-HQ and Cityscapes, to validate the effectiveness of the proposed method. The evaluation was based on several commonly used metrics, including learned perceptual image patch similarity (LPIPS), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). These metrics are designed to assess the perceptual quality, structural consistency, and overall fidelity of the inpainted images. The experimental results demonstrate that the proposed method outperforms existing inpainting techniques. On the CelebA-HQ dataset, our approach achieves a 5.88% reduction in LPIPS, a 0.52% increase in PSNR, and a significant improvement in SSIM, thereby indicating a notable enhancement in perceptual quality. The results on the Cityscapes dataset are even more impressive than those on the CelebA-HQ dataset, with the LPIPS decreasing by 6.15%, the SSIM increasing by 1.58%, and the PSNR values remaining superior to those of the other methods. These improvements highlight the effectiveness of the proposed feedback correction mechanism, which helps address semantic biases and texture errors in the inpainting process. Ablation studies were also conducted to confirm the contribution of the semantic correction mechanism further. Results show that the integration of semantic correction significantly enhances the inpainting quality, thereby validating the importance of the interaction between the inpainting and segmentation models. **Conclusion** This study provides new insights into the synergy between semantic segmentation and image inpainting. The proposed method addresses semantic inconsistencies and texture errors more effectively than previous approaches by fostering an interactive relationship between the two models. Experimental results demonstrate that the feedback correction mechanism, implemented within a semisupervised learning framework, significantly improves the overall performance of image inpainting tasks. This approach opens new possibilities for high-quality image restoration, particularly in applications where semantic consistency is crucial for achieving realistic and contextually accurate results.

Key words: image inpainting; semantic segmentation; semisupervised learning; cross-image semantic consistency; generative adversarial network (GAN)

0 引　言

随着深度学习的持续进步，图像修复技术取得了显著的发展（Xiang等，2023）。其核心目标是填补图像中的缺失区域，使得修复后的图像在视觉上既真实又具有合理的语义内容。这项技术在多个应用场景中表现出色，包括对象去除、视觉照片编辑以及损坏图像的修复。近年来研究趋向于利用语义结构（如语义分割）作为指导来完成图像修复任务（Song等，2018；Xiang等，2023）。通过提供受损区域的语义结构信息，为推断缺失图像纹理提供先验知识，能够有效保证修复内容与已知语义结构的一致性。

现有语义引导的图像修复方法主要分为两类。单次引导方法通常使用级联式架构，即先预测完整的语义图，再指导纹理修复。Xiong等人（2019）提出前景感知修复模型，借助修复前景对象轮廓指导整体修复。杨红菊等人（2022）整合语义信息和边缘信息以指导修复过程。虽然单次引导方法可以提供有效的语义信息指导，但由于缺乏反馈机制，第一阶段生成的错误语义信息将直接传递至修复阶段，导致错误累积效应。

为缓解此问题，渐进式引导方法借助分层语义预测实现动态校正，建立语义预测与纹理修复的迭代优化机制。典型工作包括：Liao等人（2020）提出了一种单阶段语义引导图像修复模型，该模型利用逐步生成更高分辨率语义图指导图像修复。考虑到多阶段实现的方法非常耗时，Yu等人（2022）开发了一个端到端的多模态引导修复网络，包括一个修复分支和两个辅助分支，辅助分支分别用于语义分割和边缘纹理。Zhang等人（2023b）提出语义金字塔网络，借助多尺度语义先验的提取逐步改进低级视觉表示。尽管这些方法取得了一定进步，但其本质上仍依赖于从纹理信息到语义信息的单向推断。近期的MDTG（mutual dual-task generator）算法（Zhang等，2024）虽然引入了修复与分割的交互思想，但其模型内部的循环依赖性，在缺乏显式纠错机制的情况下，反而可能导致初始的语义偏差在迭代中被持续放大，影响修复结果的合理性。

在深入分析问题产生的原因后，本研究提出了一种基于半监督学习的语义重校正图像修复方案。该方法不再将语义信息视为一个固定的、单向的先验，而是构建了一个动态的、双向的交互框架。该框架的核心思想是：利用初始修复结果对一个并行的语义分割网络进行“检验”和“校正”，再利用校正后的高精度语义信息反过来指导最终的精细化修复。为此，本研究创新性地引入了“跨图像语义一致性”策略，并在半监督学习框架下，高效利用未标记数据进行语义纠错，从而显著降低了对昂贵人工标注的依赖。

本文的主要贡献可概括为以下3点：1）提出了一种新颖的结合语义分割模型和图像修复模型的交叉验证框架，专注于图像修复任务。实现了两个任务的协同进化与相互促进，从而缓解初始修复中的误差累积。2）引入了基于跨图像语义一致性的半监督语义重校正策略。该策略能够有效利用未标记数据，对初始修复结果中潜在的语义偏差进行识别和校正，在显著提升语义引导准确性的同时，大幅降低了对像素级人工标注的需求。3）在多个基准数据集上验证了方法的优越性。大量定性和定量实验表明，本文算法通过语义分割模型与图像修复模型的交互作用，显著提高了图像修复的准确性和修复效果。

1 方　法

为解决引言中提出的“单向引导”与“错误累积”问题，本文设计的算法总体框架如图1所示。该框架遵循“初始修复→语义校正→精细修复”的三阶段渐进式范式，其核心在于通过引入主动的语义校正环节，构建了一个反馈闭环，从而阻断错误的传播与累积。各阶段的功能与逻辑关系如下：

初始修复阶段对缺损区域进行初步的内容填充。其目标并非追求完美的纹理细节，而是生成一个结构相对完整但可能存在语义偏差的基准图像，为后续的语义分析与校正提供必要的基底。该过程始于一幅带有掩码$\boldsymbol{M}$的受损图像$\boldsymbol{I}_{\text{in}}$，接着采用一个初始生成器$\boldsymbol{G}_c$，它以受损图像$\boldsymbol{I}_{\text{in}}$作为输入，输出一幅经过粗略修复的图像$\boldsymbol{I}_g$，而最终的初始修复图像$\boldsymbol{I}_{\text{inc}}$则由生成的掩蔽区域的内容与原始图像中未掩蔽区域的已知像素进行整合得到。该阶段，当修复区域的纹理和细节较为复杂时，算法依赖模型进行粗略修复，虽然能恢复图像的基本结构，但容易出现一些语义错误或细节失真。

为了纠正初始修复产生的语义错误，本研究提出半监督语义重校正，此为本文的核心创新模块。该模块使用标签图像和无标签图像，借助半监督学习框架对初始修复图像$\boldsymbol{I}_{\text{inc}}$进行语义分割训练。在该过程中，语义分割模型$S$是对经过初始修复阶段修复后的图像进行语义标注，接着利用半监督学习与跨图像语义一致性策略，识别并修正初始修复图像中的语义偏差，从而得到较为准确的语义图$\boldsymbol{I}_s$。这是打破错误累积、实现反馈校正的关键环节。在图1中，半监督语义重校正部分相对简略，在后续内容中将对其进行详细介绍。

在最后的精修复阶段，目标是合成具有高度视觉保真度的最终修复图像$\boldsymbol{I}_{\text{out}}$。精修复生成器$G_S$以初始修复图像$\boldsymbol{I}_{\text{inc}}$和经过校正的语义图$\boldsymbol{I}_s$为双重输入。通过利用纠正后的语义图提供精确的语义引导，可以明显提高修复效果，并消除初始修复阶段潜在的语义偏差，使其更符合真实场景的要求。

此外，整个框架的训练由对应的判别器D进行对抗性监督（Goodfellow等，2014），以确保生成图像的真实性。通过这种从粗到细的递进式设计，该方法能够显著提升修复结果的结构完整性、语义合理性与视觉真实感。

1.1 初始修复

本阶段的目标是为缺失区域生成初步的像素内容，为后续的语义校正提供基础。给定原始图像$\boldsymbol{I}_o$和受损掩码$\boldsymbol{M}$（1表示该区域的像素信息已损坏，而0表示未受损区域），受损图像可表示为$\boldsymbol{I}_{\text{in}} = \boldsymbol{I}_o \odot (1 - \boldsymbol{M}) + \boldsymbol{M}$，其中，$\odot$代表矩阵逐个元素相乘。修复过程使用编码器—解码器结构来实现（Pathak等，2016），$\boldsymbol{I}_g = G_c(\boldsymbol{I}_{\text{in}}, \boldsymbol{M})$是初步修复图像。本阶段旨在利用上述信息生成粗略的修复图像$\boldsymbol{I}_{\text{inc}}$，使其在结构上尽可能完整。

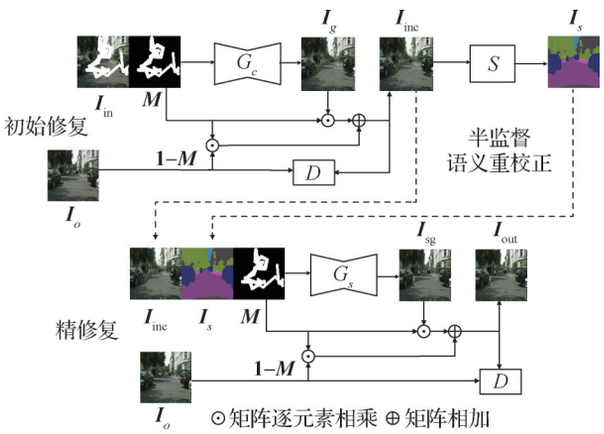

图1 整体架构

Fig. 1 The overall architecture

为实现这一目标，该模块借鉴叶学义等人（2023）的设计，采用经典的U-Net结构作为初始修复网络的生成器，如图2所示。该网络使用跳跃连接将编码器中提取的高层次特征信息有效传递到对应的解码器层，实现了特征图的逐层融合。这种特征传递机制不仅有助于修复网络中缺失的区域，而且保证了图像整体结构的完整性与连贯性，提供了

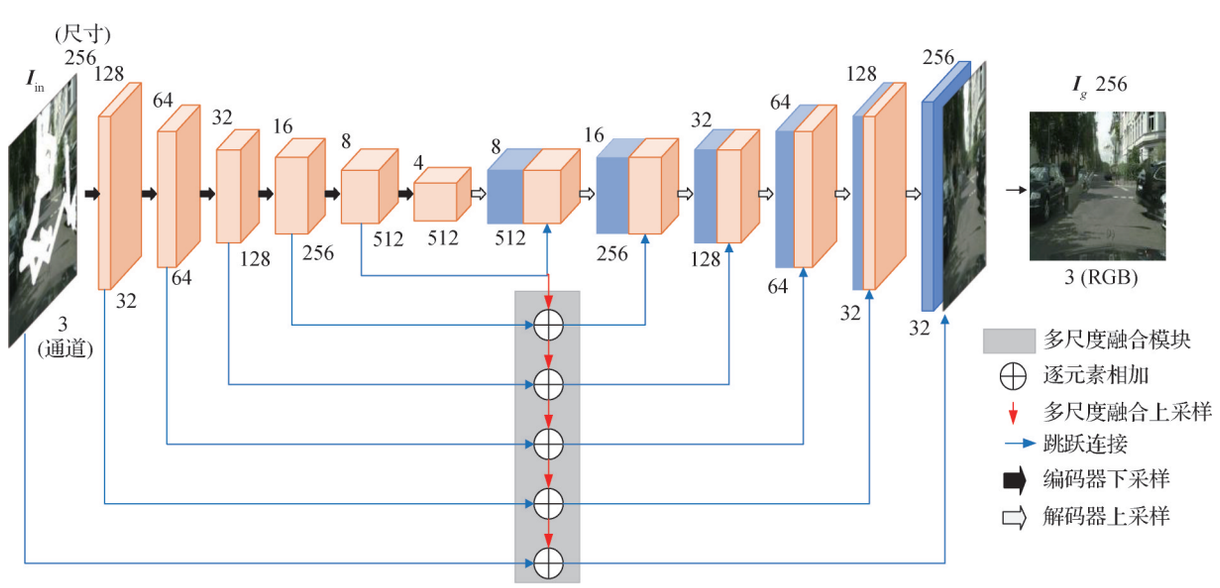

图2 初始修复模块的生成器

Fig. 2 Generator of the initial reconstruction module

丰富的视觉信息。

如图 3 所示，鉴别器基于 Patch GAN（patch generative adversarial network）（Isola 等，2017）设计。它不仅关注图像的全局信息并在局部图像块上评估真实性，能有效提升生成纹理的细节质量。

损失函数的设计结合了全局损失和对抗损失，二者共同作用于鉴别器的训练过程。总损失函数计算为

$$L_{G_C} = \lambda_{\text{adv}}^{C} L_{\text{adv}}^{C} + \lambda_{L1}^{C} L_{L1}^{C} \tag{1}$$

式中，L_{adv}^{C} 为鉴别器的对抗损失，L_{L1}^{C} 为全局损失，λ_{adv}^{C} 和 λ_{L1}^{C} 是超参数，借助实验经验设置。此处采用 L_1 距离度量修复图像与真实图像在像素层面的差异，其定义为

$$L_{L1}^{C} = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} \left| (\boldsymbol{I}_o)_{c,h,w} - (\boldsymbol{I}_{\text{inc}})_{c,h,w} \right| \tag{2}$$

式中，C、H、W 分别代表图像的通道数、高度和宽度。该表达式计算了两幅图像在每个像素点上差值的绝对值，并求取其平均值，即平均绝对误差（mean absolute error，MAE）。

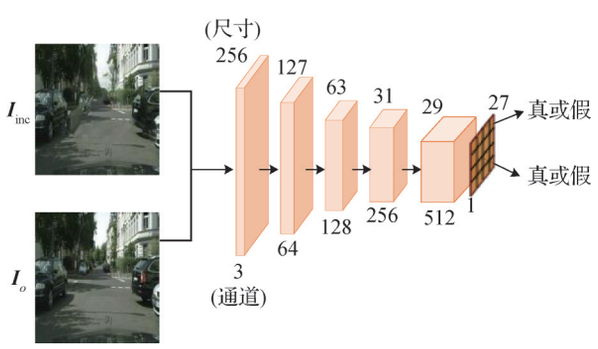

图 3　初始修复模块鉴别器网络结构

Fig. 3　Discriminator network structure of the initial reconstruction module

为提升生成图像的真实感，在此采用带有 Hinge 函数的对抗训练策略，定义为

$$L_{\text{adv}}^{C} = -E_{I_{\text{inc}} - P_{d(I_{\text{inc}})}} \left[D(\boldsymbol{I}_{\text{inc}}) \right] \tag{3}$$

$$L_D = E_{I_o - P_{d(I_o)}} \left[ReLU\left(1 - D(\boldsymbol{I}_o)\right) \right] + E_{I_{\text{inc}} - P_{d(I_{\text{inc}})}} \left[ReLU\left(1 + D(\boldsymbol{I}_{\text{inc}})\right) \right] \tag{4}$$

式中，D 为鉴别器，L_D 为鉴别器的总损失，E 表示期望，$P_{d(I_o)}$ 为真实图像样本分布，$P_{d(I_{\text{inc}})}$ 为修复图像样本分布，$\boldsymbol{I}_{\text{inc}} - P_{d(I_{\text{inc}})}$ 表示 $\boldsymbol{I}_{\text{inc}}$ 从 $P_{d(I_{\text{inc}})}$ 中得到，$\boldsymbol{I}_o - P_{d(I_o)}$ 同理。

1.2　半监督语义重校正

初始修复阶段生成的粗略图像不可避免地存在语义偏差，若直接用于引导精修复，将导致“错误累积”。为解决此问题，作为本框架核心的半监督语义重校正模块，其目标并非进行通用场景的语义分割，而是专门训练一个能够识别并纠正初始修复图像中潜在语义错误的修复图像分割模型。

本研究创新性地引入了基于跨图像语义一致性（Wu 等，2023）的半监督语义重校正框架。在语义分割中，跨图像语义一致性思路是使用标注图像中可靠且准确的语义信息纠正未标注图像的伪标签。而在图像修复任务中，对于那些初始修复模糊的图像，借助初始修复图像和真实图像（包括真实标签和伪标签）之间的语义关系，目标是改善初始修复图像的语义分割质量，校正粗修复过程中产生的语义错误。

下面以城市街景作为示例来展示图像修复任务中跨图像语义一致性的概念，如图 4 所示，图中红色与黄色矩形框标示待对比的相应区域。该流程始于存在语义偏差的初始修复图像（图 4(a)）。基于它生成的初始语义图（图 4(b)）中包含了明显的分类错误。例如，如红色标注框所示，人行道（粉色）边界区域存在误判；黄色标注框内，本应属于“汽车”（蓝色）的区域被错误地划分为“树木”（绿色）。为校正这些错误，本文策略的核心是引导模型进行跨图像比较。具体而言，模型会从数据集中检索包含相关类别的、具有准确语义标签的参考图像，如有标签图像（图 4(c)）及其对应的语义标签图（图 4(d)）。通过计算未标记的初始语义图（图 4(b)）与可靠的参考标签图（图 4(d)）之间的语义一致性关系，模型能

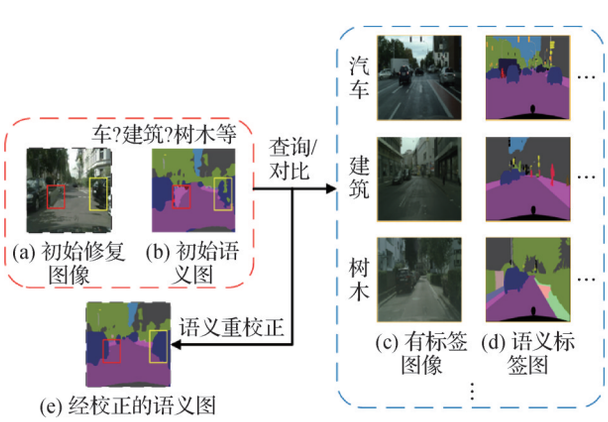

图 4　跨图像语义一致性概念展示图

Fig. 4　Conceptual diagram of cross-image semantic consistency ((a) initial inpainted image; (b) initial semantic map with errors; (c) labeled reference images; (d) semantic label maps; (e) corrected semantic map)

够识别并定位初始语义图(图4(b))中的不一致区域。最终,如经校正的语义图(图4(e))所示,上述一致性比较的结果被用于引导语义图的更新,红色与黄色标注框内的语义错误均得到了有效纠正。这表明使用相对可靠且准确的语义信息纠正初始修复过程中存在的语义错误是可行的。

如何选择可靠的标签也是一个关键问题,图像选择过程可以看做排序问题,根据图像之间的相似性对数据库中的图像进行排序。传统的图像检索任务侧重于查找与目标图像相似的图像,而在本文中是使用跨图像语义一致性的方法来实现的。这种图像选择方法(Wu等,2023)简要介绍如下:首先,为每个语义类别构建类别锚向量,作为该类别的特征基准;然后,通过计算任意图像与这些锚向量的匹配距离,可以衡量该图像与各类别语义中心的整体一致性程度,距离越小表示一致性越高,可靠性越强。

在半监督语义分割任务中,将未标注图像划分为可靠集和不可靠集,使用分阶段选择性再训练策略。该策略通过优先在可靠数据上训练模型,并利用更新后的模型迭代优化对不可靠数据的伪标签预测,从而稳步提升分割模型的整体性能。其确保了模型在迭代初期仅从高质量的伪标签中学习,从而有效避免了因初始预测不准而可能导致的负反馈与错误累积问题,且无需任何人工干预。本研究所提出半监督学习语义分割的实现流程如图5所示,具体流程如下:

1)有监督初始训练。

步骤1:使用带有真实语义标签的图像集,训练语义分割模型。

步骤2:对带有真实语义标签的图像集施加随机掩码,使用初始修复网络生成初始修复结果,并训练初始修复图像分割模型。

2)伪标签生成与筛选。

步骤3:对未标注图像集,利用语义分割模型生成伪标签。借助跨图像语义一致性纠正伪标签并筛选合格的伪标签子集,具体如下:

对于待校正的无标签图像,从有标签图像中筛选出与之语义一致性最高的图像作为参考集,并构建类别支持向量。通过计算无标签图像像素特征与支持向量的余弦相似度,对初始伪标签进行像素级校正。

继而将一致性度量应用于所有校正后的无标签图像,根据匹配距离排序,将一致性高的图像划分为可靠集,其余划分为不可靠集,用于后续的分阶段训练。

3)半监督迭代优化。

步骤4:将带有真实语义标签的图像与合格伪标签图像共同用于重新训练语义分割模型与修复图像分割模型,得到优化后的语义分割模型与初始修复图像分割模型。

步骤5:对未通过验证的图像,使用优化后的语义分割模型重新生成伪标签,将带有真实语义标签的图像与所有伪标签图像共同用于重新训练初始修复图像分割模型。

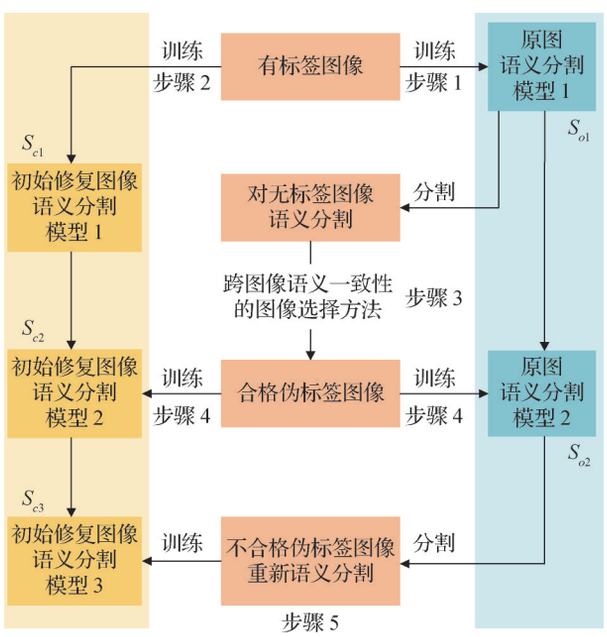

图5 半监督学习语义分割训练流程

Fig. 5 Semisupervised learning semantic segmentation training process

值得说明的是,Wu等人(2023)通过利用标注图像为未标注图像生成高质量的伪标签,有效提升了语义分割的精度。本文将该策略应用于图像修复任务中,旨在训练一个面向修复图像的语义分割模型。然而,Wu等人(2023)的方法仅着眼于通用场景下的语义分割精度提升,而本文的核心创新在于构建了一个面向图像修复任务的交互式反馈框架。具体而言,本文是将修复模型的输出作为分割模型的优化目标,使分割模型能够主动适应修复结果中存在的特殊伪影和错误。这种针对性的适配过程,使得分割模型获得的校正能力能够反向引导精修复阶段,

从而形成一个“修复—校正—再修复”的优化闭环。因此，本文的主要贡献在于创新性地将先进的半监督学习机制融入图像修复流程，并设计了双向的任务协同机制，以解决因单向引导而导致的错误累积问题。

在半监督学习框架中，真实标签的比例是影响模型性能的关键超参数。为了平衡标注成本与修复效果，本工作通过对比实验评估了模型在使用 1/8、1/4 和 1/2 三种不同标注数据比例时的性能。

从表 1 可以看出，随着真实标签比例从 1/8 增加到 1/2，各项修复指标均有显著提升。当比例达到 1/2 时，模型性能趋于稳定，在保证较高修复质量的同时，也体现了半监督学习在减少标注依赖方面的价值。因此，在后续所有对比实验中均采用此配置。

该算法通过结合图像修复与语义分割任务，利用跨图像语义一致性策略，在半监督学习框架下提升了对初始修复图像的语义分割精度。该框架不仅能有效利用未标记数据优化伪标签，更核心的是，它为图像修复中存在的语义偏差问题提供了一种新颖的解决方案。

表 1　在 Cityscapes 数据集上不同比例真实标签图像对比

Table 1　Comparison of real label images at different ratios on the Cityscapes dataset

缺损率	PSNR/dB ↑			SSIM ↑			LPIPS ↓		
	1/8	1/4	1/2	1/8	1/4	1/2	1/8	1/4	1/2
(0.0, 0.2]	30.23	31.58	**32.03**	0.936	0.948	**0.964**	0.068	0.039	**0.022**
(0.2, 0.4]	23.97	25.78	**26.95**	0.862	0.878	**0.899**	0.102	0.086	**0.063**
(0.4, 0.6]	21.48	22.96	**23.31**	0.781	0.813	**0.832**	0.147	0.112	**0.098**
平均	25.23	26.77	**27.43**	0.860	0.880	**0.898**	0.106	0.079	**0.061**

注：加粗字体表示在各缺损率区间使用不同标注数据比例时各性能指标的最优结果。↑表示值越高越好，↓表示值越低越好。

1.3　精修复

借助半监督学习语义校正训练的语义分割模型，可以为初始修复图像生成准确的语义标签，进一步将其作为强先验信息，引导最终的精修复阶段。为有效融合此先验，精修复模块扩展了初始修复模块的生成器网络，在各层的跳跃连接中嵌入了 SPADE（spatially-adaptive denormalization）模块（Park 等，2019），具体结构如图 6 所示。该模块能将校正后语义图中蕴含的空间语义信息，以仿射变换参数（缩放因子 γ 和偏移量 β）的形式直接注入到网络的每个归一化层中，使得网络能够根据每个像素的语义类别，进行精细化的、空间可变的特征调制，确保生成图像的细节更加符合输入的语义标签。

精修复生成器网络结构如图 7 所示，将 SPADE 模块嵌入跳跃连接中，可以保证从不同层次提取的特征保持一致的语义信息。鉴别器网络结构与初始修复模块（1.1 节）中描述的马尔可夫判别器（PatchGAN）相同。

在精修复阶段，先借助半监督语义重校正模型，由初始修复图像 $\boldsymbol{I}_{\text{inc}}$ 获得校正后的语义图 $\boldsymbol{I}_s$。生成器 G_s 在输入 $\boldsymbol{I}_{\text{inc}}$、$\boldsymbol{S}$ 和 $\boldsymbol{M}$ 后输出预测图像 $\boldsymbol{I}_{\text{sg}} = G_s(\boldsymbol{I}_{\text{inc}}, \boldsymbol{I}_s, \boldsymbol{M})$。

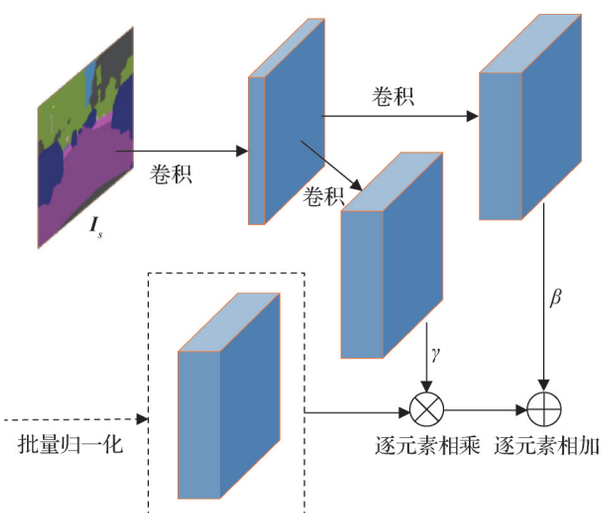

图 6　SPADE 模块示意图

Fig. 6　Schematic diagram of the SPADE module

保留 $\boldsymbol{I}_o$ 未缺损区域得到最终修复图 $\boldsymbol{I}_{\text{out}} = \boldsymbol{I}_{\text{sg}} \odot \boldsymbol{M} + \boldsymbol{I}_o \odot (\boldsymbol{1} - \boldsymbol{M})$。鉴别器 D 以 $\boldsymbol{I}_{\text{out}}$ 和 $\boldsymbol{I}_o$ 为输入进行判别。为了使二者尽可能相似，使用联合损失，其结合了生成对抗损失 L^S_{adv}、金字塔损失 L^S_{py}、感知损失 L^S_{pe}、风格损失 L^S_{sy} 和全局损失 L^S_{L1}，以从多个维度全面优化修复质量，具体为

$$L_{G_s} = \lambda^S_{\text{adv}} L^S_{\text{adv}} + \lambda^S_{\text{py}} L^S_{\text{py}} + \lambda^S_{\text{pe}} L^S_{\text{pe}} + \lambda^S_{\text{sy}} L^S_{\text{sy}} + \lambda^S_{L1} L^S_{L1} \quad (5)$$

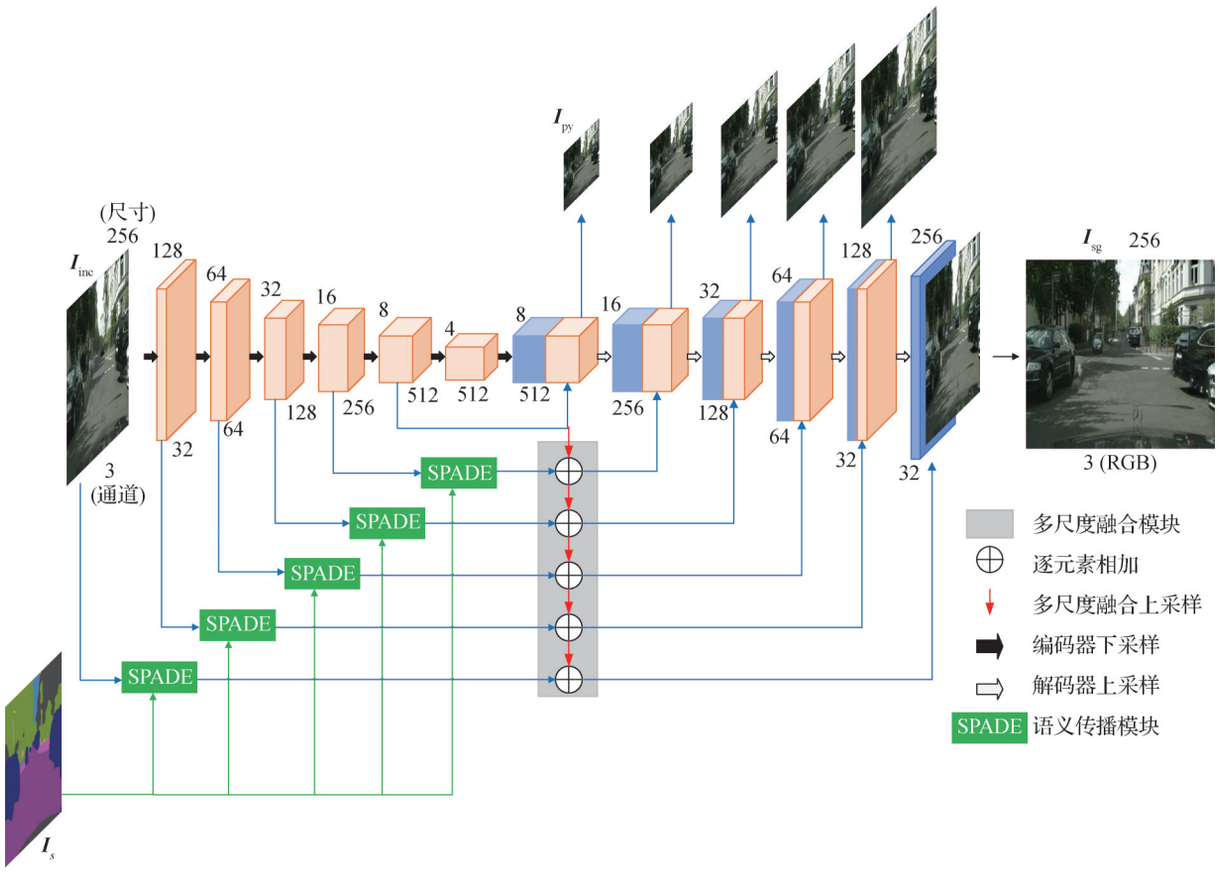

图7　精修复生成器

Fig. 7　Generator of fine inpainting

式中，λ^S_{adv}、λ^S_{py}、λ^S_{pe}、λ^S_{sy}和λ^S_{L1}是超参数。对抗损失L^S_{adv}和全局损失L^S_{L1}与初始修复阶段的定义（式(2)(3)）类似，分别用于保证像素级别的保真度和提升生成图像的真实感。

金字塔损失L^S_{py}在解码器的多个尺度上计算L_1距离，确保了由粗到精的层次化修复质量，有助于生成细节更丰富的图像，具体为

$$L^S_{\text{py}} = \sum_l \left\| \boldsymbol{I}^l_o - \boldsymbol{I}^l_{\text{py}} \right\|_1 \tag{6}$$

式中，$\boldsymbol{I}^l_{\text{py}}$为$\boldsymbol{I}_{\text{py}}$中的各个尺度的图像修复结果，$\boldsymbol{I}^l_o$由$\boldsymbol{I}_o$下采样至与$\boldsymbol{I}^l_{\text{py}}$具有相同尺度大小得到。

为了量化修复结果与真实结果之间的感知差异，精修复使用了感知损失，定义为

$$L^S_{\text{pe}} = E\left[\sum_i \frac{1}{N_i}\left\| \boldsymbol{\Phi}_i(\boldsymbol{I}_o) - \boldsymbol{\Phi}_i(\boldsymbol{I}_{\text{out}}) \right\|\right] \tag{7}$$

式中，$\boldsymbol{\Phi}_i(\boldsymbol{I}_o)$是图像$\boldsymbol{I}_o$经过VGG-19（Visual Geometry Group network-19 layers）网络后第i层的输出特征图，$\boldsymbol{\Phi}_i(\boldsymbol{I}_{\text{out}})$同理。借助输出特征图可以计算感知损失和风格损失（Johnson等，2016）。当特征图大小为$C_j \times H_j \times W_j$时，风格损失借助衡量特征图之间协方差的差异进行计算，定义为

$$L^S_{\text{sy}} = E_j\left[\left\| \boldsymbol{G}^{\boldsymbol{\Phi}}_j(\boldsymbol{I}_o) - \boldsymbol{G}^{\boldsymbol{\Phi}}_j(\boldsymbol{I}_{\text{out}}) \right\|\right] \tag{8}$$

式中，$\boldsymbol{G}^{\boldsymbol{\Phi}}_j(\boldsymbol{I}_o)$是从特征图$\boldsymbol{\Phi}_i(\boldsymbol{I}_o)$构建的$C_j \times C_j$格拉姆矩阵，$\boldsymbol{G}^{\boldsymbol{\Phi}}_j(\boldsymbol{I}_{\text{out}})$同理。应用风格损失，能有效地减轻转置卷积引入的不良视觉伪影，确保了修复结果更加美观和真实。

2　实　验

2.1　数据与实验细节

本实验在两个广泛使用的公开数据集上进行：CelebA-HQ（CelebA-high quality）人脸数据集和Cityscapes城市景观数据集。CelebA-HQ人脸数据集有29 000幅训练图像和1 000幅测试图像，该数据集原始的19个经精细注释的语义类别被合并为15个（合并了左右对称部分）。Cityscapes数据集包含20个类别的城市街道场景。为了构建一个规模

更大、多样性更强的训练环境以充分训练深度网络模型，在此遵循了Zhang等人(2024)的通用做法，使用来自原始训练集的2 975幅图像和来自原始测试集的1 525幅图像构建训练数据集，来自原始验证集的500幅图像作为测试数据集。所有实验中的图像均统一缩放至256×256像素。二进制掩码设置图像的损坏区域，参考Yu等人(2019)研究训练数据集创建不规则掩码。掩码到图像的比例分布为0%～20%、20%～40%和40%～60%。

为了准确评估本文提出的图像修复算法的性能，本工作使用定性和定量评价方法，评价采用了3个广泛认可的评价指标：峰值信噪比(peak signal-to-noise ratio，PSNR)、结构相似性(structural similarity index，SSIM)(Wang等，2004)和学习感知图像块相似度(learned perceptual image patch similarity，LPIPS)(Zhang等，2018)。

所有模型均基于深度学习框架PyTorch实现，并使用NVIDIA 3090Ti显卡进行训练与测试。实验将批处理大小设置为24，并使用Adam(adaptive moment estimation optimizer)优化器对目标函数进行优化。

实验使用了多损失函数的训练策略，并基于预设的基准权重对这些损失函数的权重进行了调整。在训练过程中初始修复损失函数权重为$\lambda_{\text{adv}}^{C}=1$和$\lambda_{L1}^{C}=10$；半监督学习语义分割训练参考Wu等人(2023)的研究；精修复模块损失函数权重为$\lambda_{\text{py}}^{S}=\lambda_{L1}^{S}=1$、$\lambda_{\text{adv}}^{S}=\lambda_{\text{pe}}^{S}=0.1$和$\lambda_{\text{sy}}^{S}=250$。

2.2 定性对比

对于CelebA-HQ数据集，实验将本文算法与RFR(recurrent feature reasoning)(Li等，2020)、CTSDG(conditional texture and structure dual generation)(Guo等，2021)、MMT(multi-modality technique)(Yu等，2022)和LG-Net(local and global network)(Quan等，2022)4种图像修复方法进行定性比较。

图8展示了各方法在CelebA-HQ数据集上的视觉比较结果。对实验图像分析后发现，与未能恢复眼镜框架的CTSDG、LG-Net等方法相比，本文方法成功重构了合理的几何形状，这直接得益于所提出的语义校正模块对“眼镜”这一语义先验的精准维持。值得注意的是，尽管同为语义引导的MMT方法也尝试修复了眼镜，但引入了明显的黑色伪影，而本文方法的结果则干净、自然。在第2行中，LG-Net生成的眉毛不符合基本纹理，而MMT算法则再次出现黑色噪声问题。CTSDG和RFR算法在该图像上的修复效果也整体不佳。在处理人脸与背景、衣物的交界处时(如第3、4行)，其他方法都普遍出现语义

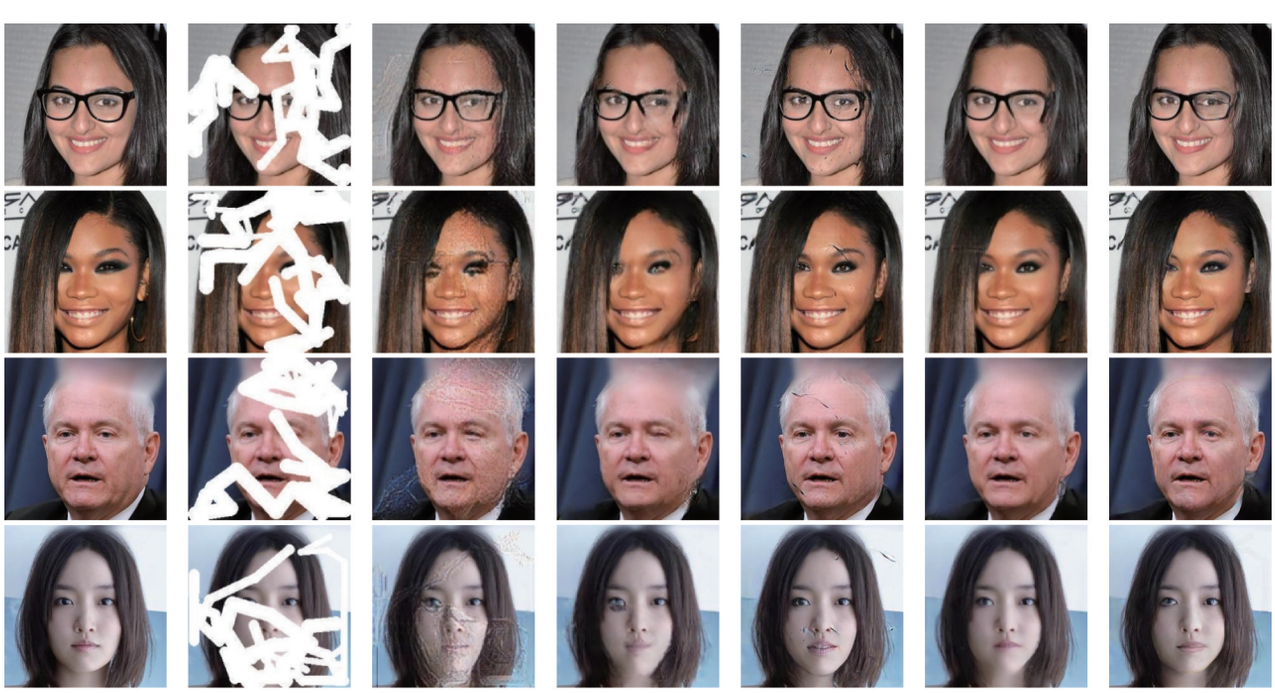

图8 CelebA-HQ数据集上定性对比

Fig. 8 Qualitative comparison on the CelebA-HQ dataset

模糊和结构混淆的问题，这表明它们在缺乏强语义约束时，容易产生纹理融合错误。本文方法则凭借精准的语义指导，生成了清晰的边界和合理的结构。

Cityscapes数据集包含的街景图像在结构上复杂且具有显著的多样性，这使得图像修复工作面临较大挑战。将本文方法与RFR、CTSDG、MMT和LG-Net这4种方法进行定性比较。

图9展示了该算法与其他图像修复算法在Cityscapes数据集上的视觉比较结果。在处理远景建筑（第1、4行）和物体边界（第3行的墙壁与道路）等挑战性场景时，本文方法的优势更为突出。其他算法的修复结果普遍存在结构扭曲、边界模糊和细节丢失的问题。例如，在第1行中，只有本文方法清晰地恢复了远方建筑的轮廓。这充分证明，在语义信息复杂多变的场景下，本文方法提出的“初始修复—语义校正—精修复”框架，能够有效识别并纠正初始阶段的语义偏差，从而避免错误累积，生成在结构和内容上都高度合理的修复结果。

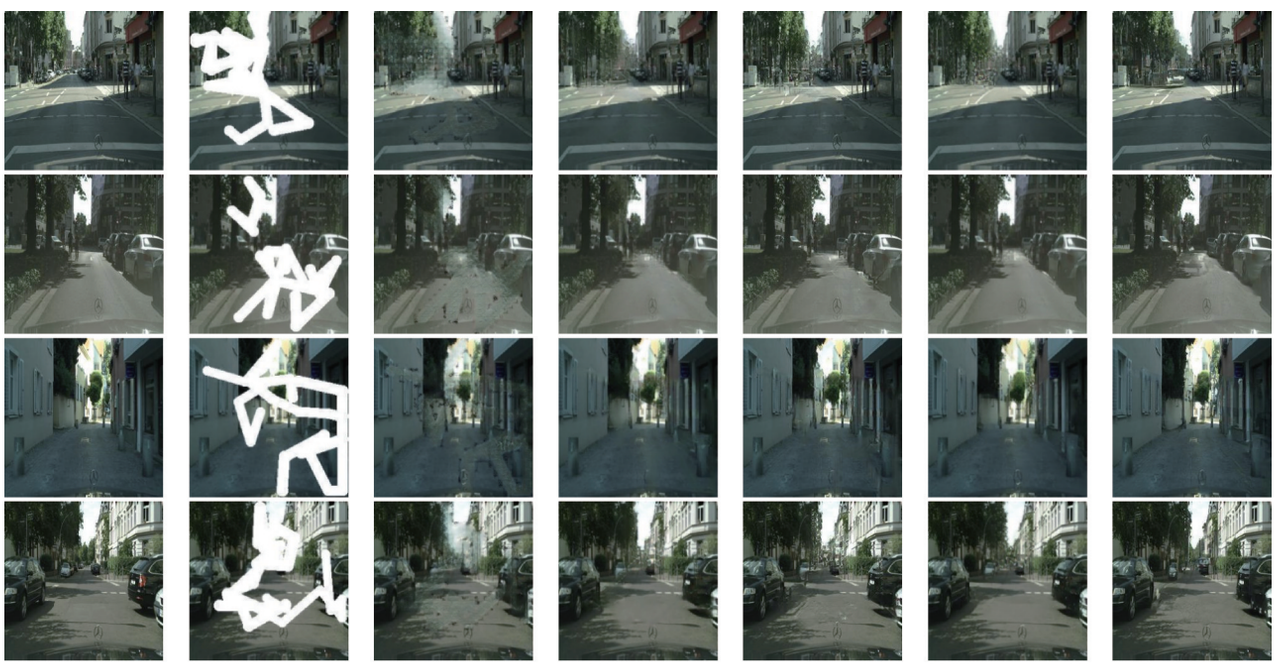

图9　Cityscapes数据集上定性对比

Fig. 9　Qualitative comparison on the Cityscapes dataset

2.3　定量评价

为了客观评价本文算法的图像修复效果，本文选取了不规则缺损掩码图像（Liu等，2018），将其与原始图像合成缺损图像作为实验样本，通过定量评估比较修复效果。表2列出了不同方法的修复评估结果，并重点与近年先进的SOTA（state-of-the-art）方法，如W-Net（W-shaped network）（Zhang等，2023a）和MDTG（mutual dual-task generator）（Zhang等，2024）进行比较。其中，RFR、CTSDG、LG-Net、W-Net和MDTG对比方法的数据来自Zhang等人（2024）实验结果，MMT来自本工作复现结果。

从表2（CelebA-HQ）的平均指标来看，本文算法在所有3个指标上均达到了SOTA水平。相较于最新的MDTG模型，本文方法在PSNR、SSIM和LPIPS上均取得了更优的结果，特别是在衡量感知质量的LPIPS指标上，实现了5.88%的相对降低。这表明，本文方法生成的修复结果不仅像素更准确，在人类视觉感知上也更接近真实图像。

对于Cityscapes数据集，同样使用评估指标SSIM、PSNR和LPIPS进行衡量。表3列出了不同方法的修复评估结果。同样，其中RFR、CTSDG、LG-Net、W-Net（Zhang等，2023a）和MDTG（Zhang等，2024）对比方法的数据来自Zhang等人（2024）实验结果，MMT来自本工作复现结果。

从表3（Cityscapes）的结果可以看出，本文算法的优势得到了进一步扩大。与同样利用语义信息的SOTA方法MDTG相比，本文方法在PSNR、SSIM和LPIPS的平均指标上全面领先，平均SSIM和LPIPS

表 2　在 CelebA-HQ 数据集上不同图像修复方法定量对比

Table 2　Quantitative comparison of different image inpainting methods on CelebA-HQ dataset

方法	PSNR/dB ↑				SSIM ↑				LPIPS ↓			
	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均	(0.0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均	(0.0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均
RFR(Li 等,2020)	30.89	25.91	23.77	26.86	0.947	0.870	0.804	0.874	0.031	0.079	0.122	0.073
CTSDG(Guo 等,2021)	31.92	27.49	25.15	28.19	0.956	0.907	0.859	0.907	0.041	0.068	0.096	0.068
MMT(Yu 等,2022)	32.08	27.22	25.12	28.14	0.958	0.904	0.849	0.904	0.018	0.045	0.059	0.041
LG-Net(Quan 等,2022)	32.32	27.00	24.43	27.92	0.962	0.903	0.842	0.902	0.023	0.060	0.097	0.060
W-Net(Zhang 等,2023a)	33.22	27.87	**25.62**	28.85	0.962	0.907	0.857	0.909	0.016	0.040	0.061	0.039
MDTG(Zhang 等,2024)	33.61	27.88	25.46	28.98	0.966	0.913	**0.863**	0.914	0.013	0.034	0.054	0.034
本文	**33.82**	**27.93**	25.51	**29.13**	**0.971**	**0.918**	0.860	**0.916**	**0.012**	**0.032**	**0.053**	**0.032**

注：加粗字体表示各列最优结果。↑表示值越高越好，↓表示值越低越好。

表 3　在 Cityscapes 数据集上不同图像修复方法定量对比

Table 3　Quantitative comparison of different image inpainting methods on Cityscapes dataset

方法	PSNR/dB ↑				SSIM ↑				LPIPS ↓			
	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	平均
RFR(Li 等，2020)	31.11	25.79	23.33	26.74	0.942	0.860	0.780	0.861	0.038	0.090	0.140	0.089
CTSDG(Guo 等,2021)	30.34	26.07	**24.28**	26.90	0.941	0.882	0.822	0.882	0.069	0.104	0.140	0.104
MMT(Yu 等,2022)	31.12	25.89	23.15	26.83	0.947	0.873	0.798	0.873	0.039	0.092	0.159	0.097
LG-Net(Quan 等,2022)	31.38	26.17	23.69	27.08	0.946	0.867	0.791	0.868	0.037	0.087	0.131	0.085
W-Net(Zhang 等，2023a)	31.04	25.49	22.45	26.33	0.949	0.875	0.791	0.872	0.042	0.099	0.158	0.099
MDTG(Zhang 等,2024)	31.68	26.61	23.44	27.24	0.950	0.884	0.817	0.884	0.027	0.066	0.103	0.065
本文	**32.03**	**26.95**	23.31	**27.43**	**0.964**	**0.899**	**0.832**	**0.898**	**0.022**	**0.063**	**0.098**	**0.061**

注：加粗字体表示各列最优结果。↑表示值越高越好，↓表示值越低越好。

分别取得 1. 58% 的相对提升和 6. 15% 的相对降低。Cityscapes 数据集场景复杂、语义类别多样，极易在初始修复中产生语义错误。本文方法在该数据集上的显著优势，强有力地证明了所提出的半监督语义重校正模块的有效性。该模块能够主动识别并修正初始修复引入的语义偏差，为最终的精修复提供高质量的语义先验，从而在复杂的城市场景中实现了更准确的结构恢复和更真实的纹理生成。

值得注意的是，在部分高缺损率区间，本文方法的 PSNR 指标可能并非最佳，虽然初始修复图像进行了语义校正，但在初始修复中图像细节的丢失并不会被校正。然而，LPIPS 指标始终保持领先，这说明，即使在极端情况下部分底层像素细节的恢复稍有偏差，本文方法所生成的整体语义结构在人类感知层面依然是最为合理的。

此外，本文算法也存在一定局限性。为了进一步探究本文算法的适用边界与鲁棒性，本研究也在语义类别更庞杂、场景更多样化的 ADE20K 数据集上进行了测试。实验发现，模型在该数据集上性能有所下降，并暴露出当前框架的一个核心局限性。该现象的根源在于 ADE20K 数据集中极高的类内方差（intra-class variance）。例如，“建筑”这一语义类别同时包含了规整的城市摩天楼和形态不一的乡村小屋。本文方法的核心机制—跨图像语义一致性在为初始修复图像的“建筑”区域进行语义校正时，会从数据集中检索所有“建筑”样本的特征。当这些样本特征本身存在巨大差异时（如摩天楼 vs. 小屋），模型难以形成一个统一、稳定的语义先验，反而可能引入冲突的纹理信息，导致修复失败。

以图 10 中的典型失败案例为例，在修复建筑立

面受损区域时，本文算法错误地引入了周边纹理特征；而在修复户外场景中的动物区域时，则产生了不符合规律的自然草原纹理。这一发现揭示了本文算法的一个重要适用前提：模型的性能高度依赖于训练数据集中各个语义类别的内部一致性。当一个语义类别包含的场景或形态差异过大时，本文方法中语义校正模块的效果会下降。这一分析不仅解释了模型在特定数据集上的性能表现，也为未来的改进指明了方向，即如何提升模型对高类内方差数据的鲁棒性将是一个关键的优化点。

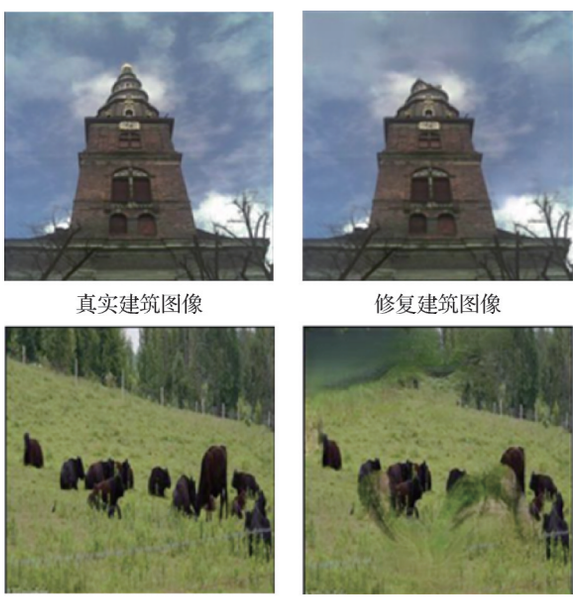

图10　ADE20K数据集的典型失败案例

Fig. 10　Typical failure cases of ADE20K dataset

2.4　基于Cityscapes数据集的消融实验

为验证本文算法中各个核心模块的有效性，本研究在Cityscapes数据集上设计了一系列消融实验。

2.4.1　半监督学习语义校正模块的有效性

初始修复和精修复可以看做粗略修复和精细修复两部分。而半监督学习下语义校正作为中间的桥梁，通过消融半监督学习语义校正，来说明半监督学习语义分割指导的有效性和优势。具体而言，比较初始修复结果和最终修复结果之间差异。

定量结果如表4所示，表4中“无”代表仅有初始修复模块，“有”代表完整的本文算法。可以清晰地看到，包含语义校正和精修复的完整模型（表4中“有”）在所有指标上均显著优于仅有初始修复的模型（表4中“无”）。平均而言，PSNR提升了9.8%，SSIM提升了5.15%，而感知指标LPIPS则大幅降低了47.4%。这一结果强有力地证明，初始修复阶段确实会产生大量语义和纹理错误，而本文所提出的语义校正模块能够有效地识别并纠正这些偏差，为后续的精修复提供高质量的语义指导，从而实现修复质量的质变。

综上所述，初始修复与精修复的比较凸显了半监督学习语义校正作为中间桥梁的有效性。借助语义信息的引导，模型不仅可以完成粗略修复，而且能在细节上进行有效优化，明显提升整体修复效果。

表4　在Cityscapes数据集上有无语义重修复结果对比

Table 4　Comparison of semantic correction results on the Cityscapes dataset

缺损率	PSNR/dB ↑		SSIM ↑		LPIPS ↓	
	无	有	无	有	无	有
(0.0, 0.2]	29.96	**32.03**	0.930	**0.964**	0.075	**0.022**
(0.2, 0.4]	23.64	**26.95**	0.859	**0.899**	0.112	**0.063**
(0.4, 0.6]	21.35	**23.31**	0.773	**0.832**	0.161	**0.098**
平均	24.98	**27.43**	0.854	**0.898**	0.116	**0.061**

注：加粗字体表示较优结果。↑表示值越高越好，↓表示值越低越好。

2.4.2　交互式反馈机制的必要性验证

为了验证本文提出的“修复—校正”交互式反馈机制的必要性（即针对修复任务优化语义分割模型），在此将其与一种无反馈的单向引导方法进行对比。具体而言，本研究使用一个在真实标签上预训练好的、固定的语义分割模型直接指导精修复，而不经过本文提出的半监督语义重校正过程。

定量结果如表5所示，表5中“未使用”代表使用固定的预训练分割模型进行引导，“使用”代表完整的本文算法。可以看出，使用本文方法训练的、自适应的分割模型（表5中“使用”）在性能上全面优于使用固定的预训练模型（表5中“未使用”）。这说明，一个通用的、预训练的语义分割模型并不能完美适配充满伪影和错误的初始修复图像。直接使用它进行单向引导，反而可能因为无法识别初始修复的潜在错误而放大语义偏差。相比之下，本文方法的交互式反馈机制能够让分割模型“看到”并“适应”初始修复的结果，从而学习到针对性的校正能力，这对于打破错误累积、提升最终修复质量至关重要。

表 5　Cityscapes 数据集上是否使用校正语义分割模型对比

Table 5　Comparison of whether semantic correction segmentation model is used on the Cityscapes dataset

缺损率	PSNR/dB ↑		SSIM ↑		LPIPS ↓	
	未使用	使用	未使用	使用	未使用	使用
(0.0, 0.2]	30.87	**32.03**	0.944	**0.964**	0.034	**0.022**
(0.2, 0.4]	25.67	**26.95**	0.868	**0.899**	0.091	**0.063**
(0.4, 0.6]	22.69	**23.31**	0.787	**0.832**	0.137	**0.098**
平均	26.41	**27.43**	0.866	**0.898**	0.087	**0.061**

注：加粗字体表示较优结果。↑表示值越高越好，↓表示值越低越好。

2.4.3　基于消融实验的定性对比

在对 Cityscapes 数据集图像施加随机掩码后，分别去除半监督学习语义校正模块和使用预训练模型进行语义分割（上述两种消融设置），将获得的修复结果与本文算法的修复结果进行了对比，结果如图 11 所示。可以看出，移除半监督语义重校正模块后（“无语义校正”列），修复结果保留了大量初始修复的模糊纹理和错误结构，如第 3 行的墙面和道路边界完全混淆，证明了语义校正对于结构恢复的必要性。使用固定的预训练模型（“预训练模型”列），虽然优于无校正，但在处理初始修复引入的伪影时能力不足，导致了明显的语义混淆和不自然的噪声（如第 3 行道路与墙壁交界处），未能生成清晰的边界。而本文算法凭借其完整的交互式语义校正机制，成功地恢复了清晰的建筑轮廓和合理的道路纹理，生成了在结构、语义和视觉真实感上都最优的结果。

3　结　论

针对现有语义引导修复方法中“引导”单向性、易造成错误累积的问题，本文提出了一种创新的、基于半监督学习的语义重校正图像修复算法。该算法的核心贡献在于：

在图像修复与语义分割模型之间建立了双向的反馈校正通路，使修复模型能够主动引导分割模型的优化，从而打破了传统方法中语义引导的单向局限性。

本文方法创新性地将跨图像语义一致性策略与半监督学习相结合，能够有效利用未标记数据对初始修复产生的语义错误进行识别与纠正，在提升语

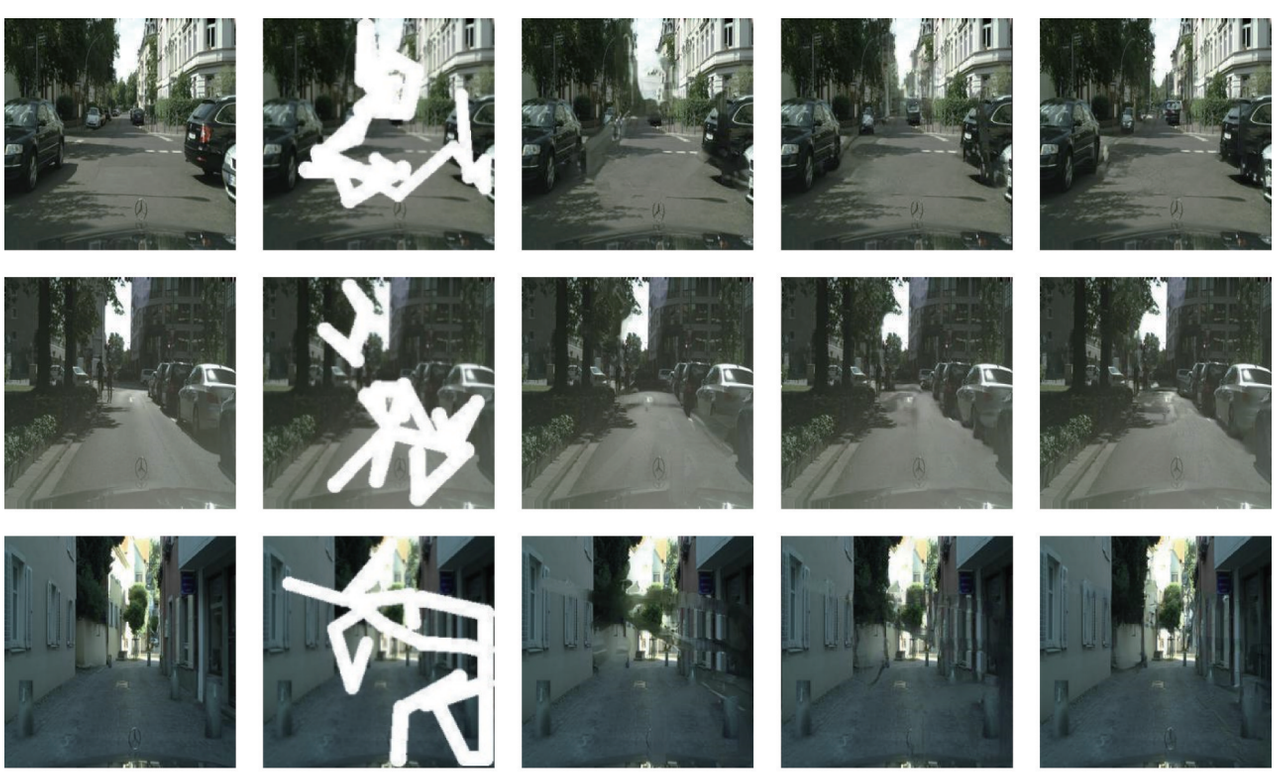

图 11　Cityscapes 数据集上消融实验定性对比

Fig. 11　Ablation experiment qualitative comparison on the Cityscapes dataset

义先验质量的同时，显著降低了对昂贵人工标注的依赖。

在CelebA-HQ和Cityscapes等多个基准数据集上的大量实验表明，本文方法在定量指标和定性效果上均优于近年来的多种先进算法，尤其在处理结构复杂、易产生语义偏差的场景时，优势更为显著。

同时，该研究也认识到本文算法存在一定的局限性。其性能在一定程度上依赖于训练数据集中语义类别的内部一致性，当面临如ADE20K这样具有极高类内方差的数据集时，语义校正效果会下降。此外，多阶段的训练流程也带来了相对较高的计算成本。在模型效率方面，本文提出的“初始修复—语义校正—精修复”三阶段框架是一个非端到端的流程。其中，核心的“语义重校正”模块依赖于一个独立的半监督训练过程（包含伪标签生成与迭代优化），其计算成本主要体现在独立的训练阶段。因此，该框架的整体计算复杂度无法简单地用传统的前向传播FLOPs（floating point operations per second）或参数量进行衡量。这种设计是为了通过分阶段精细处理来换取更高的修复质量，尤其是在纠正复杂语义偏差方面。

未来的研究将围绕两个方向展开：1）探索更鲁棒的特征对齐与校正策略，以提升模型在处理高类内方差数据时的性能；2）研究将校正机制轻量化并整合进端到端训练方案的可能性，以期在保持高性能的同时优化模型的训练与推理效率。总体而言，本工作为解决复杂场景下的语义一致性修复问题提供了一种有效的新范式，具有重要的理论与应用价值。

参考文献（References）

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. 2014. Generative adversarial nets//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2672-2680 [DOI: 10.5555/2969033.2969125]

Guo X F, Yang H Y and Huang D. 2021. Image inpainting via conditional texture and structure dual generation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14114-14123 [DOI: 10.1109/ICCV48922.2021.01387]

Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]

Johnson J, Alahi A and Li F F. 2016. Perceptual losses for real-time style transfer and super-resolution//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 694-711 [DOI: 10.1007/978-3-319-46475-6_43]

Li J Y, Wang N, Zhang L F, Du B and Tao D C. 2020. Recurrent feature reasoning for image inpainting//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 7757-7765 [DOI: 10.1109/CVPR42600.2020.00778]

Liao L, Xiao J, Wang Z, Lin C W and Satoh S I. 2020. Guidance and evaluation: semantic-aware image inpainting for mixed scenes//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 683-700 [DOI: 10.1007/978-3-030-58583-9_41]

Liu G L, Reda F A, Shih K J, Wang T C, Tao A and Catanzaro B. 2018. Image inpainting for irregular holes using partial convolutions//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 89-105 [DOI: 10.1007/978-3-030-01252-6_6]

Park T, Liu M Y, Wang T C and Zhu J Y. 2019. Semantic image synthesis with spatially-adaptive normalization//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 2332-2341 [DOI: 10.1109/cvpr.2019.00244]

Pathak D, Krähenbühl P, Donahue J, Darrell T and Efros A A. 2016. Context encoders: feature learning by inpainting//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2536-2544 [DOI: 10.1109/CVPR.2016.278]

Quan W Z, Zhang R S, Zhang Y, Li Z F, Wang J and Yan D M. 2022. Image inpainting with local and global refinement. IEEE Transactions on Image Processing, 31: 2405-2420 [DOI: 10.1109/TIP.2022.3152624]

Song Y H, Yang C, Shen Y J, Wang P, Huang Q and Kuo C C. J. 2018. SPG-Net: segmentation prediction and guidance network for image inpainting//Proceedings of 2018 British Machine Vision Conference. Newcastle, UK: BMVA Press: #97

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Wu L S, Fang L Y, He X X, He M, Ma J Y and Zhong Z. 2023. Querying labeled for unlabeled: cross-image semantic consistency guided semi-supervised semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(7): 8827-8844 [DOI: 10.1109/TPAMI.2022.3233584]

Xiang H Y, Zou Q, Nawaz M A, Huang X F, Zhang F and Yu H K. 2023. Deep learning for image inpainting: a survey. Pattern Recognition, 134: #109046 [DOI: 10.1016/j.patcog.2022.109046]

Xiong W, Yu J H, Lin Z, Yang J M, Lu X, Barnes C, et al. 2019. Foreground-aware image inpainting//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5833-5841 [DOI: 10.1109/CVPR.2019.00599]

Yang H J, Li L Q and Wang D. 2022. Deep learning image inpainting combining semantic segmentation reconstruction and edge reconstruction. Journal of Image and Graphics, 27(12): 3553-3565 (杨红菊, 李丽琴, 王鼎. 2022. 联合语义分割与边缘重建的深度学习图像修复. 中国图象图形学报, 27(12): 3553-3565) [DOI: 10.11834/jig.210702]

Ye X Y, Zeng M S, Sun W J, Wang L Y and Zhao Z J. 2023. Image inpainting based on multi-scale stable-field GAN. Scientia Sinica Informationis, 53(4): 682-698 (叶学义, 曾懋胜, 孙伟杰, 王凌宇, 赵知劲. 2023. 多尺度稳定场GAN的图像修复模型. 中国科学: 信息科学), 2023, 53(4): 682-698 [DOI: 10.1360/SSI-2022-0065]

Yu J H, Lin Z, Yang J M, Shen X H, Lu X and Huang T. 2019. Free-form image inpainting with gated convolution//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 4470-4479 [DOI: 10.1109/ICCV.2019.00457]

Yu T, Guo Z Y, Jin X, Wu S L, Chen Z B, Li W P, et al. 2020. Region normalization for image inpainting//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 12733-12740 [DOI: 10.1609/aaai.v34i07.6967]

Yu Y S, Du D W, Zhang L B and Luo T J. 2022. Unbiased multi-modality guidance for image inpainting//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 668-684 [DOI: 10.1007/978-3-031-19787-1_38]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/cvpr.2018.00068]

Zhang R S, Quan W Z, Zhang Y, Wang J and Yan D M. 2023a. W-Net: structure and texture interaction for image inpainting. IEEE Transactions on Multimedia, 25: 7299-7310 [DOI: 10.1109/TMM.2022.3219728]

Zhang W D, Wang Y B, Ni B B and Yang X K. 2023b. Fully context-aware image inpainting with a learned semantic pyramid. Pattern Recognition, 143: #109741 [DOI: 10.1016/j.patcog.2023.109741]

Zhang Y L, Liu Y M, Hu R T, Wu Q and Zhang J. 2024. Mutual dual-task generator with adaptive attention fusion for image inpainting. IEEE Transactions on Multimedia, 26: 1539-1550 [DOI: 10.1109/TMM.2023.3282892]

作者简介

叶学义，男，副教授，硕士生导师，主要研究方向为模式识别和信息安全。E-mail: xueyiye@hdu.edu.cn

睢明聪，男，硕士研究生，主要研究方向为数字图像修复。E-mail: smc13777849137@163.com

谭瑞洁，女，硕士研究生，主要研究方向为图像处理。E-mail: 18711706783@163.com

蒋德琦，男，硕士研究生，主要研究方向为目标检测。E-mail: jiangdeqi2002@163.com

陈华华，男，副教授，硕士生导师，主要研究方向为图像处理。E-mail: iseealv@hdu.edu.cn