

基于 GAN 网络图像语义指导的取证修复图像模型

王金伟^{1,2,3}, 张子荷^{1,2}, 罗向阳^{3,4}, 马 宾⁵

¹ 南京信息工程大学 数字取证教育部工程研究中心 南京 中国 210044

² 南京信息工程大学计算机学院、软件学院、网络空间安全学院 南京 中国 210044

³ 数学工程与先进计算国家重点实验室 郑州 中国 450001

⁴ 中国人民解放军战略支援部队信息工程大学 郑州 中国 450001

⁵ 齐鲁工业大学网络空间安全学院 济南 中国 250353

摘要 大数据时代的到来使得网络上图像数据信息越来越丰富,篡改图像也越来越多,为了区分篡改图和原图,图像篡改检测方法层出不穷。虽然篡改检测可以定位图像的篡改区域,但是仍无法了解原图的图像以及原图所表达的语义信息。为了解决上述问题,本文提出了一种基于生成对抗网络(Generative adversarial network, GAN)图像语义指导的取证修复图像模型。首次在篡改环境下,针对于 copy-move 和拼接篡改这两种篡改手段进行图像修复。该模型利用 GAN 网络对抗式进行训练,最终得到训练好的生成器进行图像取证修复工作。模型利用图像语义识别模块,提供原图图像的语义特征,对篡改环境下的图像修复进行指导和约束,这样使得修复图像遵循原始图像的语义,且生成器不会过度对图像进行美化,从而达到还原真实图像的目的。该模型在 GAN 网络中,将生成器修复的图像进行对应位置掩模拼接后再放入判别器,使得图像真实像素区域较多,这样可以训练出更加严格的判别器,最终使得生成器生成的图片更加真实。模型在语义识别模块和生成器特征拼接前设置了遗忘门,使得模型根据自己的需要选择适当的语义指导强度,使得修复图像更加真实。我们在三个数据集上的大量实验证明了所提模型在像素级修复图片方面的可行性和有效性。

关键词 GAN; 取证修复; 图像语义识别; 图像篡改; 语义指导

中图法分类号 TP391 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2025.09.07

A Forensic Restoration Image Model Based on GAN Network for Image Caption Guidance

WANG Jinwei^{1,2,3}, ZHANG Zihe^{1,2}, LUO Xiangyang^{3,4}, MA Bin⁵

¹ Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China

² Department of Computer Science, Department of Software, Department of Cyberspace Security, Nanjing University of Information Science and Technology, Nanjing 210044, China

³ State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

⁴ PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China

⁵ School of Cyberspace Security, Qilu University of Technology, Jinan 250353, China

Abstract The arrival of the big data era has made image data information on the network increasingly rich, and there are also more and more tampered images. In order to distinguish between tampered images and original images, image tamper detection methods are emerging one after another. Although tamper detection can locate the tampered area of an image, it still cannot understand the original image and the semantic information expressed by the original image. To address the aforementioned issues, this paper proposes a forensic restoration image model based on Generative Adversarial Network (GAN) image semantic guidance. For the first time in a tampering environment, image restoration was carried out using two tampering methods: copy move and concatenation tampering. This model utilizes GAN network adversarial training to obtain a trained generator for image forensics and restoration work. The model utilizes an image semantic recognition module to provide semantic features of the original image, guiding and constraining image restoration in a tampered environment. This ensures that the repaired image follows the semantics of the original image, and the generator does not ex-

通讯作者: 王金伟, 博士, 教授, Email: wjwei_2004@163.com。

本课题得到国家自然科学基金(No. 62072250, No. 62172435, No. U1804263, No. U20B2065, No. 61872203, No. 71802110, No. 61802212); 中原科技创新领军人才项目(No. 214200510019); 江苏自然科学基金(No. BK20200750); 河南省网络空间态势感知重点实验室开放基金(No. HNTS2022002); 江苏省研究生研究与实践创新项目(No. KYCX200974); 广东省信息安全技术重点实验室开放项目(No. 2020B1212060078); 山东省计算机网络重点实验室开放课题基金(No. SDKLCN-2-22-05); 人文社会科学教育部项目(No. 19YJA630061) 国家重点研发计划项目(No. 2021QY0700)资助。

收稿日期: 2023-09-18; 修改日期: 2024-03-23; 定稿日期: 2025-08-14

cessively beautify the image, thus achieving the goal of restoring the real image. In the GAN network, this model concatenates the images repaired by the generator with corresponding position masks and then puts them into a discriminator, making the image have more real pixel regions. This can train a more rigorous discriminator, ultimately making the images generated by the generator more realistic. The model sets a forget gate before the semantic recognition module and generator feature concatenation, allowing the model to choose appropriate semantic guidance intensity according to its own needs, making the repaired image more realistic. Our extensive experiments on three datasets have demonstrated the feasibility and effectiveness of the proposed model in pixel level image restoration.

Key words GAN; forensic repair; image caption; image tampering; semantic guidance

1 引言

1.1 图像篡改检测

图像作为信息载体在人类长期的发展与交流中扮演着重要的角色。但随着图像处理技术的发展,篡改工具层出不穷,篡改图像在网络上越来越多。这些篡改图片造成的社会舆论影响了人们正常生活,甚至对国家信任体系造成影响。因此,图像篡改检测技术应运而生。篡改检测可以判断图像的真假,对保护图像真实性和有效性做出了巨大贡献。但篡改检测只能定位出篡改区域,无法还原图像了解图像真实的语义。而篡改图像的目的一般都是破坏图像语义信息,使得图片语义发生改变,误导人们的思维。因此,只进行篡改检测,判断图像真假还不能达到最终目的。为了进一步减弱篡改图像对社会国家的影响,还原原始图像语义信息变得非常重要。

近年来,图像篡改检测技术作为研究重点方向,研究学者提出了很多图像篡改检测方法。

2017年, Ambili 等人^[1]提出了一种适用于图像被篡改模糊情况下的拼接区域定位算法。2019年, Wu 等人^[2]提出了一种统一的深度神经体系结构 ManTranet。它们将伪造定位问题巧妙地转化为局部异常检测问题,使用类似人类的思维进行图像篡改定位检测。2021年, Jiang 等人^[3]提出了一种基于篡改边缘两侧和篡改区域内外不一致性引导下的图像篡改检测网络。2021年, He^[4]提出了基于渐进式监督注意力机制嵌套 U-net 的图像复制粘贴伪造检测网络。这些篡改检测方法虽然能识别篡改图像,且可以实现篡改区域定位,但仍不清楚未篡改图像表达的语义信息以及未篡改的真实图像。

1.2 图像修复

为了了解未篡改图像表达的语义信息,需要进行图像取证修复,还原篡改原图。图像修复是在给定相应掩模的情况下填充受损图像中的缺失像素。目前,研究学者提出了很多图像修复方法。

2018年, Liu 等人^[5]发现基于深度学习的图像修复使用标准卷积网络,会造成伪影。所以他们建议使

用部分卷积,并重新规范化,从而仅限于有效像素。他们还使用了带有自动掩模更新机制的部分卷积层。2019年, Nazeri 等人^[6]提出了一个两阶段模型,将修复问题分为结构预测和图像完成两部分,由边缘生成器和图像完成网络组成,两者都遵循对抗模型。2020年, Wan 等人^[7]通过利用真实照片和大量合成图像对,提出了一种新颖的三重域翻译网络来恢复严重退化的旧照片。2021年, Wan 等人^[8]使用变换器进行外观先验重建,使用 CNN 进行纹理补充。前一个变换器恢复了多元相干结构以及一些粗糙纹理,而后一个 CNN 增强了由高分辨率掩模图像引导的粗糙先验的局部纹理细节。2019年, Yu 等人^[9]通过在密集图像块上应用光谱归一化鉴别器,提出了一种基于块的 GAN 损失,称为 SNPatchGAN,可以使得训练更加快速稳定。2018年, Yu 等人^[10]提出了一种新的基于深度生成模型的方法,该方法不仅可以合成新的图像结构,而且可以在网络训练期间明确地利用周围的图像特征作为参考,以做出更好的预测。2019年, Zeng 等人^[11]提出了一种金字塔上下文编码器网络,该网络通过对全分辨率输入的上下文语义进行编码,并将学习到的语义特征解码回图像来恢复图像。2019年, Zheng 等人^[12]提出图像多样化修复,为图像完成生成多个和多种可行的解决方案。2016年, Pathak 等人^[13]提出了一种由基于上下文的像素预测驱动的无监督视觉特征学习算法,实验发现,上下文编码器学习的表示不仅捕获外观,还捕获视觉结构的语义。

以上图像修复算法虽然可以对图像进行修复,但其修复图像的目的大多为修复旧照片、进行图像美化或增加图像的可读性。这些修复算法所考虑的场景都是未被篡改的图像,而对于篡改后的图像无法进行修复。这是因为图像篡改,尤其是恶意篡改,一般都会破坏原始图像的语义,且篡改后区域掩模的语义也会发生极大的变化。而普通图像修复不会考虑待修复图本身的真假,而是直接关注缺失掩模轮廓区域,做最大可能的修复。这导致待修复图像会受到篡改掩模轮廓的影响,无法做出正确的修复。还原的修复图仍遵循篡改后图像的语义。这是取证环

境下修复和普通图像修复相比的研究重点。

如图 1 所示, (a)为原始未篡改图像, (b)为篡改后图像, (c)为篡改掩模。如果使用普通图像修复方法, 按照篡改掩模对篡改后图像进行扣除, 再放入修复模型进行修复, 会得到(d)使用普通图像修复方法的修复结果。这是由于普通图像修复算法易受掩模轮廓干扰, 会依据掩模轮廓信息做大致判断修复图像, 并未考虑空缺位置真假, 普通图像修复目的是尽最大可能进行图像修复, 使得图像视觉效果更佳。而取

证修复方法需要排除篡改掩模轮廓的干扰, 因为取证修复的前提是掩模区域本身就是被篡改的。若按照该掩模轮廓进行修复, 得到的仍不是真实原始图像。即普通图像修复默认掩模和原始图像语义相同, 而对于取证修复来说, 掩模和原始图像语义并不同。使用取证修复方法的修复图如图(e)所示。

本文的主要工作是在篡改环境下, 进行图像的语义修复。因此, 需要找到能够引导图像减少关注篡改掩模形状, 增强关注空缺图像语义信息的方法。

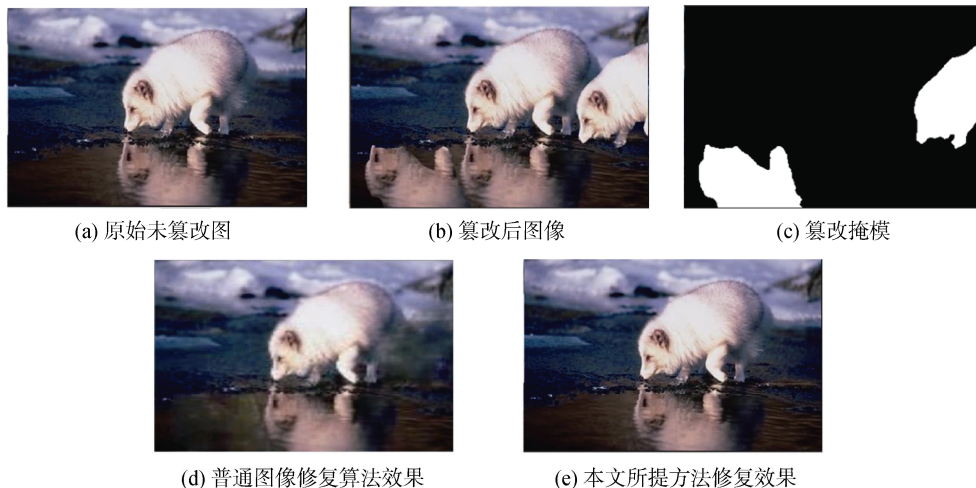


图 1 普通图像修复和取证图像修复的不同效果, 其中(a)为原始未篡改图像, (b)为经过篡改后的图像, (c)为篡改掩模, (d)为使用普通图像修复算法修复后的效果图, (e)为使用本文所提方法, 在篡改环境下的修复效果图

Figure 1 The different effects of ordinary image restoration and forensic image restoration, where (a) is the original unmodified image, (b) is the tampered image, (c) is the tampered mask, (d) is the effect image repaired using the ordinary image restoration algorithm, and (e) is the repair effect image in the tampered environment using the method proposed in this article

1.3 图像文本生成技术

图像文本生成技术的出现给我们带来了启发。图像文本生成技术即将图像概括为文本信息, 将图像中目标信息以及目标间的关系通过一段文字进行描述。而图像篡改就是改变图像描述的文本信息, 因此将图像文本作为指导生成修复图具有研究意义。近年来, 图像文本生成技术的发展使得预测图像语义成为可能, 这一研究为篡改环境下图像修复指明了道路。

图像文本生成技术近年来得到了很大的发展。2015 年 Karpathy 等人^[14]提出了一个生成图像及其区域的自然语言描述的模型。2016 年, Vinyals 等人^[15]提出了一种基于深度递归架构的生成模型, 可用于生成描述图像的自然语句。该模型被训练以最大化给定训练图像的目标描述句子的可能性。2019 年, Biten 等人^[16]提出了一种能够利用与图像相关联的新闻文章文本提供的上下文信息进行新闻文章文本识

别的方法。模型能够在视觉提示的引导下选择性地从文章中提取信息, 并动态地将输出字典扩展到上下文源中出现的词汇表外的命名实体。2015 年, Xu 等人^[17]引入了基于注意力的模型, 还通过可视化展示了模型如何能够在生成输出序列中的相应单词的同时自动学习注视突出对象。

图像文本生成模型可以识别图像的语义内容。例如: 原图表达的文本信息为一个人站在路边打电话。通过拼接篡改增加一个人, 文本信息会从“one people”改为“two people”。这改变了图像表达的语义。所以使用图像文本生成技术作为特征指导是很有必要的。首先使用该先验模型生成图像语义, 再将生成的语义转换成特征图, 使用拼接融合技术, 与修复图像生成器中的特征图融合, 达到图像文本信息指导取证图像修复的目的。这样不仅可以告知生成器原图应该有的语义物体信息, 还可以约束生成器不会过分美化修复的图片。

1.4 GAN 网络

本文采用 GAN 网络架构。近年来, GAN 网络得到广泛的研究。2014 年, Mirza 等人^[18]介绍了生成对抗网络的条件版本, 可以通过向生成器和判别器提供数据标签来构造在图像修复领域也有了一定的应用成果。2016 年, Reed 等人^[19]利用 GAN 网络从文本中自动合成真实的图像, 从而有效地连接文本和图像, 将视觉从字符转换为像素。2022 年, Sun 等人^[20]提出了 SeCGAN, 用于使用语义信息编辑人脸图像, 而不需要指定目标语义掩码。2015 年, Radford 等人^[21]引入了一类称为深度卷积生成对抗网络(DCGAN)的 CNN, 它具有某种结构约束, 并说明了他是一个有力的无监督候选方法。

受到以上研究学者的启发, 本文针对现有图像修复网络模型存在的问题, 提出了一个基于 GAN 网络的图像语义指导的取证修复图像模型。所提模型采用 GAN 网络, 使得生成器和判别器相互对抗训练, 最终达到平衡。模型利用图像语义先验模块, 对图像修复进行指导和约束, 使得修复图像遵循原始图像语义, 达到还原真实图像的目的。模型在将生成器生成的图片放入判别器前, 进行对应掩模的拼接操作, 将拼接后的修复图放入判别器, 一起进行 Loss 计算。这样可以训练得到更加严格的判别器, 使得最终生成的修复图片视觉效果更好。模型在生成器和先验模型进行融合学习前设计了遗忘门, 使得生成器可以根据自己的需要选择合适的文本指导特征应用。该模型的提出, 很好地解决了 copy-move 和拼接篡改修复图像的问题。本文的贡献如下:

(1) 本文首次在篡改环境下对图像语义进行取证修复, 并取得了较好的效果。

(2) 本文运用训练好的图像语义先验模块, 对篡改环境下的图像修复进行文本指导, 同时约束 GAN 网络, 防止图像过分美化。

(3) 本文在语义指导前加设遗忘门, 使得修复模型根据自己的需要选择适当的语义指导, 控制指导力度, 使得修复图像更加真实。

2 相关工作

图像修复就是将图像中损坏的像素部分进行修复填充。传统的图像修复方法有 diffusion-based 和 patch-based 两种。近年来, 深度学习的发展给图像修复提供了新思路。

2016 年, Pathak 等人^[13]提出了一种由基于上下文的像素预测驱动的无监督视觉特征学习算法。上下文编码器需要了解整个图像的内容, 从而对缺失

区域进行合理的假设。实验发现, 上下文编码器学习的表示不仅捕获外观, 还捕获视觉结构的语义。训练好的模型能够生成基于上下文的图像, 提高了修复水平。

2018 年, Liu 等人^[5]发现基于深度学习的图像修复使用标准卷积网络, 会造成伪影。而想要处理这些伪影, 需要经过后处理, 但是成本高且不一定成功。所以他们建议使用部分卷积, 卷积被遮蔽并重新规范化, 从而仅限于有效像素。他们还使用了带有自动掩模更新机制的部分卷积层。

2019 年, Nazeri 等人^[6]提出了一个两阶段模型, 将修复问题分为结构预测和图像完成两部分, 模型首先利用边缘图的形式进行缺失区域图像结构的预测, 然后将其结果指导图像修复的过程。模型整体由边缘生成器和图像完成网络组成, 两者都遵循对抗模型。

2021 年, Wan 等人^[8]将变压器和神经网络两者的优点结合起来, 以实现高保真的自由形式多元图像完成。文章使用变换器进行外观先验重建, 使用 CNN 进行纹理补充。前一个变换器恢复了多元相干结构以及一些粗糙纹理, 而后一个 CNN 增强了由高分辨率掩模图像引导的粗糙先验的局部纹理细节。

3 所提方法

3.1 整体网络框架

本文所提模型利用 GAN 网络进行对抗式训练, 最终得到训练平衡的最优图像修复模型; 利用预先训练好的图像语义先验模型, 实现对生成器的指导和约束。将残缺图文本信息转换为特征图, 和 GAN 网络生成器进行特征融合, 使得生成器避免根据残缺轮廓判别缺失内容, 这是取证修复和普通修复需要注意的不同点。同时防止 GAN 网络过分美化图片, 导致修复结果和原图差距较大; 模型在图像文本特征图和生成器特征图拼接前设计了遗忘门, 这样模型可以根据自己的需要进行文本特征选择, 实现了生成器与先验信息自适应融合; 模型将生成器生成的修复图放入判别器时, 首先进行对应掩模的拼接操作, 使得最终生成的修复图只有掩模对应部分是模型生成的。这样再进入判别器时, 图片会更加真实, 从而训练出更加严格的判别器, 最终提高生成器的生成效果。

本文提出的模型整体架构如图 2 所示。该网络将待修复的空缺图分为两个分支走向。一方面将空缺图放入图像语义先验模型^[17]中, 该模型首先经过 Resnet101 编码器生成 2048×14×14 大小的特征图, 再

将特征图放入解码器中, 得到该空缺图图像文本信息; 另一方面将空缺图放入 GAN 网络中, 进入生成器生成修复图, 根据掩模拼接后, 将修复图和原图放入判别器中, 形成生成器和判别器的对抗式训练。在 GAN 网络生成器中, 我们将图像语义分支生成的图像文本信息转换成 $2048 \times 14 \times 14$ 大小的特征图后, 再经过 padding 操作, 变为 $2048 \times 16 \times 16$ 大小。经过遗忘门模块后, 将得到的 $2048 \times 16 \times 16$ 大小的特征图和生成器中的 $2048 \times 16 \times 16$ 的空缺图特征进行融合拼接, 实现图像语义指导并且约束生成器修复空缺图。

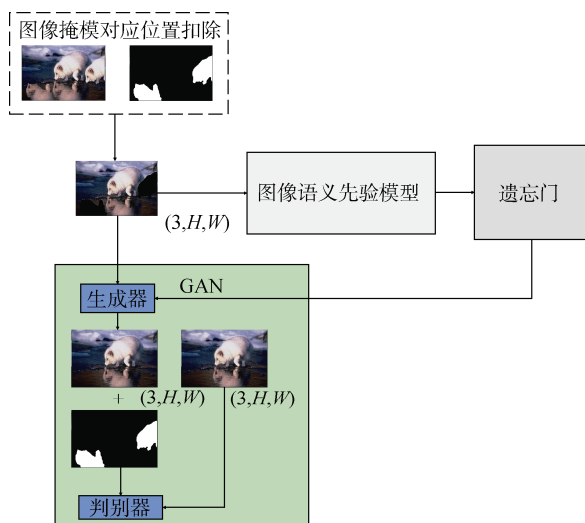


图 2 网络模型整体结构框架

Figure 2 Overall structural framework of network model

3.2 图像语义先验模块

对于普通图像修复来说, 残缺待修复部分若有

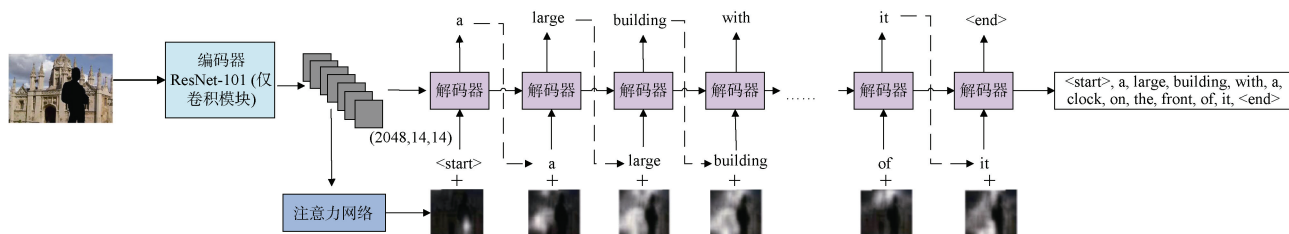


图 3 图像语义先验模型

Figure 3 Image semantic prior model

3.3 GAN 网络

本文使用的 GAN 网络在原基础上进行了改动。原始 GAN 网络一般是经过生成器生成图像后, 直接将生成图和真实图放入判别器进行判断, 得出图像为真或假的结果。本文在 GAN 网络生成图放入判别器前, 增加使用掩模进行对应位置的拼接操作, 将

较清晰的轮廓特征, 会为图像修复提供便利。例如一片草地盛开着很多花, 但其中一朵由于篡改缺失了, 那么模型会根据残缺轮廓对该区域进行修复, 生成一朵花。但是对于 copy-move 或拼接这样的篡改图像来说, 图像篡改检测算法会先判断某一朵花是否为篡改区域, 生成篡改区域对应的掩模。此时会生成花朵形状的残缺掩模区域。但是若修复算法仍然根据普通修复的逻辑把它修复成一朵花, 这样的修复明显是错误的。相反, 应该修复为原图的背景信息。这是普通图像修复不能进行篡改环境下修复的原因。为了解决这个问题, 我们设计了一个残缺图像文本识别模块, 该模块将识别得到的残缺图像文本作为生成器修复的指导信息, 指导并约束其修复过程。图像文本生成模型参考 Xu 等人^[17]的设计。在训练图像语义先验模型时, 数据集一部分为未经过随机掩模扣除的 MSCOCO 图像, 另一部分为 MSCOCO 进行随机不规则掩模扣除后的图像。文本标签均为完整图像的语义标签。这是为了训练图像文本生成模型对残缺图像的原始完整语义信息进行准确表述。

图像文本识别模块如图 3 所示。该模块首先将输入图像经过 Resnet101 编码器网络, 生成 $2048 \times 14 \times 14$ 大小的特征图。然后经过 LSTM, 根据注意力网络每一步提供的注意力图和上一轮生成的单词, 进行下一步单词的预测生成。最终连词成句, 生成图像文本信息。当图像逐字生成文本信息, 经过注意力网络时, 可以发现模型的视线在图像上移动, 图像上的亮度代表不同区域的注意力强度。图像越亮的地方注意力越集中。模型中的注意力机制, 使得模型专注于与下一步生成单词最相关联的图像部分。

生成图中对应残缺掩模部分的结果扣出, 并放入残缺图的对应位置。本文所提整体 GAN 网络如图 4 所示。待修复的空缺图首先经过生成器, 生成修复好的完整图像。之后将掩模和生成器生成的图像进行对应位置的扣除, 取待修复图的空缺生成部分, 这一部分才是真正希望生成填充的。而对于图像其余部

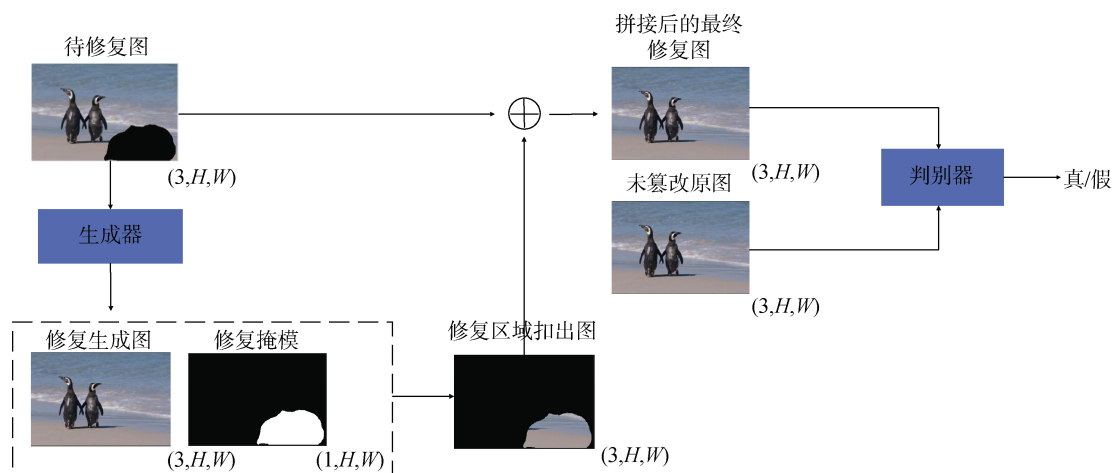


图 4 GAN 网络流程框架
Figure 4 GAN network process framework

分, 并不希望它产生变动。因此, 本文将扣出的修复区域和待修复空缺图进行对应位置拼接, 生成最终修复图。再将其和未篡改图像放入判别器进行判断, 这样放入判别器的修复图视觉效果更佳, 真实区域较多, 和完整图像差异相对全部由生成器生成的修复图较小。对于判别器来说, 相对较难区判别别, 有利于训练出更加严格的判别器。由于判别器和生成器是相互对抗训练的, 判别器更加严格, 那么对于 GAN 的生成器来说, 生成的图像也更加逼真, 且空缺修复边缘过渡填充视觉效果也更好。最终, 生成器生成的图像修复效果更佳, 视觉效果更好。

3.4 遗忘门设置

遗忘门的作用是决定语义先验模型指导 GAN 网络生成器时, 哪些特征应被保留, 哪些应被过滤或遗忘。通过研究发现先验模块在指导空缺较大的待修复图或有较少纹理特征的区域时, 指导效果较明显。但在空缺区域较小或者纹理特征连续性较明显

的图像上, 指导过多反而会干扰 GAN 网络生成器的生成效果。分析认为: 由于纹理特征连续性较明显的图像在提取特征时能够较清晰地了解空缺区域应填入的纹理特征, 而空缺区域较大或纹理特征与空缺区域联系较少的图像提取特征时, 对于空缺区域的特征提取较为困难, 故此时需要语义先验模块的指导特征加入生成器进行图像修复。由于这两种情况在图像修复时都会发生, 所以在语义先验模型连接到 GAN 网络生成器前设置了遗忘门操作。让模型根据当前修复图像的空缺纹理情况自主决定语义先验模块的指导力度大小。对于空缺纹理较清晰的待修复图, 指导力度相对较小; 对于纹理不连续的图, 指导力度相对较大。遗忘门设计如图 5 所示。使用语义先验模型时, 我们希望完整图像的文本信息能够在一定程度上对 GAN 网络生成器起到指导和约束作用, 但又不希望过多地干扰生成器生成。所以本文在先验模块指导信息输入 GAN 网络生成器前设置了遗

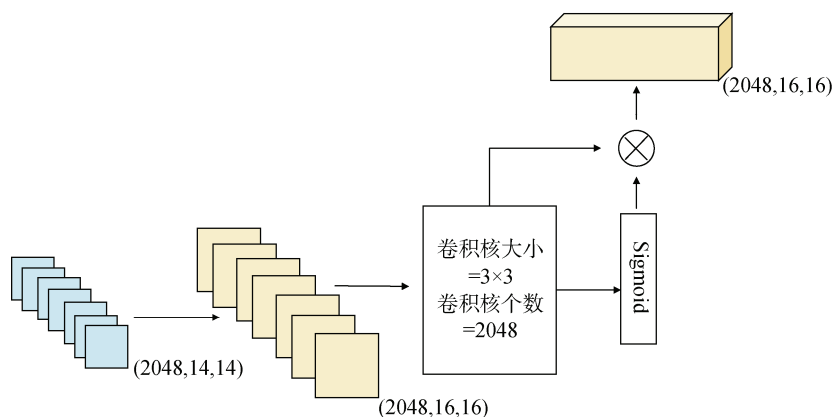


图 5 遗忘门设计示意图
Figure 5 Schematic diagram of forgetting door design

忘门。先验模块生成的文本信息转化成 $2048 \times 14 \times 14$ 大小的特征图, 再经过 padding 变为 $2048 \times 16 \times 16$, 以便和生成器中深层特征图进行通道维度的拼接。padding 后的特征图经过 3×3 大小卷积核融合特征, 再进行 sigmoid 操作, 使得特征图信息转换为 $0 \sim 1$ 。这些数字代表每个位置特征的权重, 权重表示在该位置特征的重要性。权重越靠近 1, 表示该位置特征对后续的作用越大; 越靠近 0, 作用权重越小。将经过 sigmoid 的权重特征图和经过 padding 的特征图进行对应位置相乘操作, 使得特征图中的无用信息权重占比降低, 从而减小不重要图像特征对 GAN 网络生成器的干扰, 使得指导特征信息更明确。

4 实验结果及分析

本节将通过大量实验验证所提模型在篡改环境下图像修复的有效性。在 Casia、RTD、Nist16 三个篡改数据集上对所提模型的修复能力进行评估, 并且和普通图像修复进行了比较。图像语义先验模型的训练数据集使用 MSCOCO 完整图和 Casia 的篡改掩模进行不规则扣除, 得到空缺图和其对应的完整 MSCOCO 标签文本信息。实验同时证明了使用图像语义先验模型对 GAN 网络生成器进行指导的作用、在拼接指导信息前使用遗忘门的作用以及在放入判别器前进行图像拼接操作的有效性。

4.1 实验参数设置

为了评估所提模型, 我们在 Casia、RTD、Nist16 三个数据集上进行修复效果评估实验。对于图像语义先验模型, 使用 MSCOCO 与 Casia 不规则掩模生成空缺图, 但仍使用完整图像文本信息。据此训练模型可以根据空缺图进行完整图像文本信息预测。

在训练先验模型时, 本文将 MSCOCO 空缺图按照 0.8: 0.2 的比例划分为训练集和测试集。训练整体模型架构时, 将公开篡改数据集提供的标签掩模和其对应的图像进行对应位置扣除, 模拟篡改图像进行篡改检测定位后, 修复前的扣除操作, 得到待修复空缺图。此空缺图作为模型的输入, 模型训练标签为对应的未经过篡改的原图图像。

(1) Casia 图像篡改检测数据集包含原始图像和不同类型的篡改图像, 如拼接、复制粘贴、涂抹等, 共计约 800 张图像。篡改图像均经过精心制作, 篡改区域经过了仔细的选择, 并进行滤波和模糊处理, 以淡化篡改痕迹。同时, 该数据集还提供了与每个篡改图像对应的真实标注信息, 以便进行模型训练和性能评估。

(2) RTD 图像篡改数据集包含拼接、复制粘贴、

涂抹等篡改操作。数据集提供了原始图像、篡改图像和篡改图像所对应的篡改掩模。该数据集中的篡改图像以 TIF 格式提供。

(3) Nist16 图像篡改检测数据集覆盖了各种类型的篡改, 包括复制、粘贴、图像合成、图像旋转、图像裁剪、缩放、噪声添加等。该数据集中每个图像都有一个相应的掩模标签, 且提供每张图像的原图。这些图像是以成对的方式提供的, 即每个篡改图像都有一个对应的原始图像。原始图像和篡改图像都以 TIFF 格式提供。

(4) MSCOCO 图像描述数据集包含超过 8 万张图像, 每张图像都配有 5 个人工编写的文本描述。这些描述涵盖了图像中物体的种类、位置、数量、场景、动作等方面, 具有多样性和准确性。而且涵盖了多种场景和语言风格, 适用于不同文化和语言背景的研究。因此, 它已成为图像自动描述算法研究的重要数据集之一。该数据集的广泛应用包括图像自动描述、图像检索、视觉问答等方向。

本文训练网络设置的初始学习率为 $1e-4$, 训练一共设置 300 个 epoch, 每 10 个 epoch 学习率下降 50%, batch-size 设置为 64, 网络模型接受的输入图片大小为 $224 \times 224 \times 3$, 本文的所有实验均在单个 NVIDIA GeForce RTX 3090 Ti GPU 上进行。

4.2 损失函数设置

为了衡量图像语义先验模型预测的性能, 指引模型训练沿着正确的方向前进, 本文图像语义模型使用 CrossEntropyLoss 作为损失函数, 该损失函数内部会自动加上 softmax 层。先验模型完成训练后加入整体修复模型的预测训练, 先验模型参数不再变化。

为了衡量整体模型的预测性能, 本文为模型选用 BCELoss 作为损失计算方式。BCELoss 表达式如式(1)所示。

$$L = \text{Loss}(o, t) = -\frac{1}{n} \sum (t[i] * \log(o[i]) + (1 - t[i]) * \log(1 - o[i])) \quad (1)$$

其中, t 表示标签, 即目标值, 它在 $[0, 1]$ 。 o 表示模型的预测值, 在计算损失之前需要经过 sigmoid 函数处理。 $t[i]$ 表示标签掩模第 i 个位置的值。 $o[i]$ 表示模型预测第 i 个位置的值。

4.3 消融实验

为了验证本文所提模型各个模块的有效性, 对不同模块进行消融实验。数据集使用 Casia 数据集的篡改图像以及其对应的掩模进行扣除分割, 得到空缺图、未篡改的原图和掩模。训练集和测试集的分割比例为 0.8: 0.2。实验结果如表 1 所示。

表 1 所提模型各个模块对训练结果的影响

Table 1 The impact of each module of the proposed model on the training results

数据集	Loss 前拼接	Loss 后拼接	先验 模块	遗忘门	PSNR	SSIM
Casia	√				22.534	0.921
		√			20.863	0.901
	√		√		27.759	0.945
	√		√	√	30.268	0.961

经过实验发现, GAN 网络生成器生成的图片进入判别器计算 Loss 前, 进行基于掩模的拼接操作, 可以提高最终图像生成效果。分析认为: 进行拼接操作后的图像, 真正的生成区域仅为掩模对应区域, 图像中生成像素较少, 整体视觉效果更好。将拼接后视觉效果较好的图像输入判别器, 有利于训练生成更加强大的判别器。由于判别器和生成器训练是相互对抗的, 因此, 判别器越强大, 生成器也越强大, 最终训练完成后生成器生成的图片效果会更好, 可以提升整体的训练效果。

本文采用的实验效果评估方式为 PSNR 和 SSIM 两个指标。其中, PSNR 是一种图像质量评价指标, 他基于对应像素点间的误差, 即为基于误差敏感的图像质量评价指标。公式如式(2)所示。其中 MSE 表示当前图像 X 和参考图像 Y 的均方误差, MSE 计算公式如式(3)所示。 H 、 W 分别为图像的高和宽, n 为每像素的比特数, 一般取 8, 像素灰阶数为 256。PSNR 数值越大表明失真越小, 图像质量越好。

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (2)$$

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (3)$$

SSIM 也是一种图像质量评价指标, 该指标从亮度、对比度、结构三个方面来度量图像相似性。公式如式(4)所示, 其中 $l(x, y)$ 为亮度比较, $c(x, y)$ 为对比度比较, $s(x, y)$ 为结构比较, 具体计算公式如式(5)所示。其中 μ_x 、 μ_y 表示图像 x 和 y 的均值, σ_x 、 σ_y 表示图像 x 和 y 的方差, σ_{xy} 表示图像 x 和 y 的协方差。 $C1$ 、 $C2$ 、 $C3$ 为常数, 可避免分母为 0 的情况。SSIM 取值范围为 $[0, 1]$, 值越大, 图像失真越小, 图像质量越好。

$$\text{SSIM} = l(x, y)c(x, y)s(x, y) \quad (4)$$

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x, y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (5)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

根据表格结果可以发现, Loss 前进行拼接操作的 PSNR 相比于 Loss 后进行拼接提升了 1.67%, SSIM 提升了 0.019%。因此, 本文后续的实验都采用 Loss 前拼接操作; 图像语义先验模块的有无对训练整体效果有一定的影响。从实验数据可以看出, 使用先验模块和 Loss 前拼接的 PSNR 为 28.759, 比不使用先验模块提升了 5.225%。分析认为: 对于空缺位置较大或没有什么纹理特征的缺失区域, 图像修复时无法学习到相关空缺位置的大量特征信息, 不利于图像修复。普通的图像修复方法对于缺失区域较小或纹理特征较明显的空缺图像有着比较好的修复效果。但对于空缺区域较大的部分, 语义指导发挥了较大作用; 不仅如此, 我们还发现有时过度的图像语义指导反而会造成图像修复效果下降。观察此类图像发现, 这些待修复图像的缺失区域相对较小, 且有着较规律的分布特征。因此本文在图像语义指导信息进入 GAN 网络前设置遗忘门, 作为先验指导特征的筛选, 让模型自主决定保留多大权重的先验指导。根据实验数据, 添加遗忘门的模型比无遗忘门的模型最终模型修复效果得到了提升, PSNR 提升了 2.509%, SSIM 提升了 0.016%。

综上所述, 本文最终采取带有先验模型和遗忘门的模型, 且在 GAN 网络生成图像进入判别器前进行掩模对应位置的拼接操作。实验结果证明了所提方法在提高图像修复效果上的有效性。

4.3.1 图像语义先验模块有效性

为了进一步说明图像语义先验模块指导的有效性, 我们使用 Casia 数据集进行实验。训练过程中实验结果如图 6 所示。测试结果如图 7 所示。

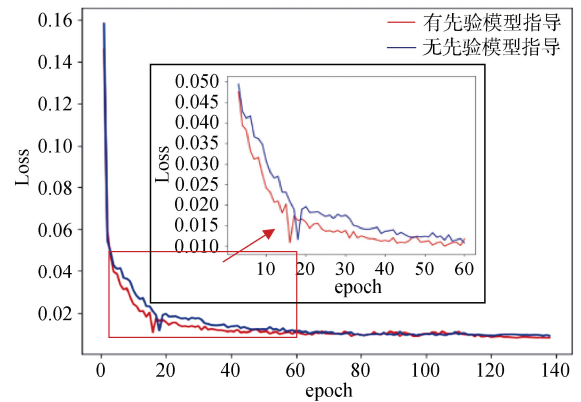


图 6 训练过程中有无先验模型指导模块对训练集 Loss 的影响折线图

Figure 6 Line chart showing the impact of prior model guidance modules on the training set Loss during the training process

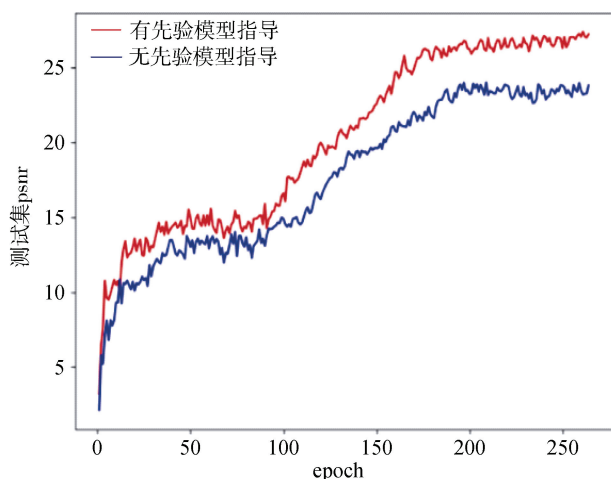


图 7 训练过程中有无先验模型指导模块对测试集 PSNR 的影响折线图

Figure 7 Line chart showing the impact of prior model guidance modules on the PSNR of the test set during the training process

根据实验结果可以看出, 先验模型的存在加快了模型训练的收敛速度, 尤其在训练前期表现明显。在训练过程中测试集上的表现表明先验模型可以提高图像修复能力, 测试集 PSNR 随着训练轮数的增

加而提高。存在先验模型指导的模型测试 PSNR 相比于无先验模型指导有一定程度的提高, 这说明了先验模型指导的有效性。

实验发现对于空缺位置较大的待修复图片, 先验模型的指导作用更加明显。分析认为: 由于大的空缺位置可能缺失的是整个物体信息, 那么仅靠 GAN 网络, 无法判断出缺失区域原本的信息。此时就需要先验语义模型进行先验特征指导, 告知 GAN 网络生成器此处的空缺特征信息, 从而达到最终恢复图像原本语义信息的目的。且先验模型指导信息可以对 GAN 网络生成器进行一定的约束作用, 不会让 GAN 网络为了达到更加逼真的效果背离原图的语义信息, 无限制地修复发挥。语义先验模型从约束和指导两方面作用于 GAN 网络生成器, 使得最终生成修复图像效果得到提升。

4.3.2 先验模型训练集选择消融

本文在训练先验语义指导模块时, 首先尝试使用 MSCOCO 公开图像描述数据集。将训练好的模型进行待修复空缺图像文本预测, 但实际预测效果并不理想。实验结果如图 8 所示。其中第一列为待预测文本信息的空缺图像; 第二列为使用 MSCOCO 完



图 8 图像文本预测效果图, 其中第一列为要预测文本信息的空缺图像, 第二列为使用 MSCOCO 完整图像训练出模型的检测效果, 第三列为使用 MSCOCO 进行矩形扣除后训练模型的检测效果, 第四列为使用 Casia 掩膜进行扣除后训练模型的检测效果

Figure 8 Image text prediction effect diagram, where the first column is the vacant image to predict text information, the second column is the detection effect of the model trained using MSCOCO complete images, the third column is the detection effect of the model trained using MSCOCO after rectangular deduction, and the fourth column is the detection effect of the model trained using Casia mask after deduction

整图像训练出模型的检测效果;第三列为使用 MSCOCO 进行随机矩形扣除后训练模型的检测效果;第四列为使用 Casia 掩模进行随机扣除后训练模型的检测效果。分析结果注意力图发现,由于 MSCOCO 数据集都是完整的图像,模型具备完整图像的文本识别能力。在识别完整图像文本时,注意力图的高亮注意点符合预测下一单词的需求。但本文的目标是预测不完整图像的文本信息,输入图像存在较大黑色空白,这会对模型预测效果产生影响。观察第二列注意力图可以发现,模型会过于关注黑色空缺区域,造成文本预测不准确。为了解决这个问题,本文在 MSCOCO 完整数据图的基础上进行不规则像素扣除,使得模型训练时能够自主根据标签文本信息决定对空缺区域的关注度。我们首先尝试在 MSCOCO 完整图像的基础上进行随机位置矩形像素扣除,仍使用完整图像的标签文本信息。这是因为本文训练先验模型的目的是能够在空缺图像中识别出原本完整图像的文本信息,并将其作为修复图像的依据。根据第三列实验结果发现,该方法对于空缺位置的注意力减弱有一定效果,但仍会在一定程度上由于关注空缺区域造成预测文本不准确。因此,最终本文使用 Casia 数据集中的随机不规则掩模进行 MSCOCO 随机位置的不规则扣除,仍使用原本完整图像的图像文本标签。目的是训练模型能够在不规则扣除的空缺图像中完整表达出图像原本的文本信息,为图像修复提供正确的指导。根据第四列注意力图可以发现,最终训练出的模型不会过于关注

不规则空缺区域,并且能够准确预测出图像原本表达的信息。

4.3.3 遗忘门消融

本文设计了遗忘门模块,从而让模型自主决定指导作用的权重大小。实验发现先验模块指导作用在空缺较大的图或者有较少纹理特征的区域指导效果较为明显。但在空缺区域较小或纹理特征连续性较明显的图像上,指导过多反而会干扰 GAN 网络生成器的生成效果。实验结果如图 9 所示,其中第一列为空缺图,第二列为无遗忘门时修复效果图,第三列为存在遗忘门时修复效果图。第一行为纹理信息较为明显的空缺图修复示例。模型可以根据现有的图像在一定程度上推测出空缺位置的图像纹理信息。此时,图像语义指导的作用并无很明显,反而过度的指导会造成图像修复效果退化。所以此时需要遗忘门对指导的作用强度进行模型自主减弱,达到合适的指导强度,用于修复图像;第二行为空缺区域较大且无明显纹理特征的空缺图像。此时图像语义先验指导的作用效果较明显。由于空缺区域无明显连续纹理修复特征,因此,需要借助先验文本信息指导图像修复,遗忘门此时仍起到控制指导强度的作用。综上,遗忘门对于小面积缺失纹理信息较清晰的图像修复工作有着较为明显的效果,可以防止指导过度造成的修复缺陷。因此,本文在先验特征输入 GAN 网络前添加遗忘门,用于让网络模型自主决定指导作用强度的大小。实验结果表明遗忘门的设置对网络图像修复效果的有效性。



图 9 第一列为空缺图,第二列为无遗忘门时模型的修复效果图,第三列为存在遗忘门时模型的修复效果图
Figure 9 The first column is the vacancy map, the second column is the repair effect map of the model without forgetting gates, and the third column is the repair effect map of the model with forgetting gates present

4.4 模型性能评估

4.4.1 先验模型性能评估

为了指导 GAN 网络生成器,本文先验模型为提

前训练好的图像文本预测模型。由于我们需要将篡改区域扣除掉的空缺图进行修复,修复成原图图像信息,所以需要先验图像语义模型能够在空缺图像

中预测出原图表达的图像文本信息。因此采用 MSCOCO 数据集中完整图像的图像文本标签以及 MSCOCO 数据集使用 Casia 图像篡改数据集掩模进行随机扣除后的不规则空缺图来作为先验模型的训练数据进行训练。最终训练得到的模型预测图像文本能力如图 10 所示。从结果可以看出, 模型对空缺图的预测效果较好, 能够在空缺图中预测出原图表达的文本信息, 且不会被空缺的掩模轮廓影响, 预测出的文本信息能够指导 GAN 网络的图像修复过程。



图 10 先验模型图像文本预测效果展示图

Figure 10 Display diagram of prior model image text prediction effect

4.4.2 整体模型图像修复效果性能评估

为了验证本文所提模型对图像的修复效果, 我们在三个图像篡改数据集上进行了实验。首先本文根据数据集的篡改图像以及其对应掩模进行对应位置扣除, 模拟图像篡改检测模型判断该图像为假图像并生成篡改掩模的操作。之后为了还原原始图像, 根据生成掩模进行对应位置的像素扣除, 生成空缺图像, 放入本文设计的图像修复模型中进行修复。实验结果如表 2 所示。分析结果发现在 Casia 数据集上 SSIM 可以达到 0.961, PSNR 可以达到 30.268, 肯定了所提模型的图像修复能力。而在 Nist16 和 RTD 数据集上效果并没有在 Casia 上好, 分析认为: Nist16 数据集和 RTD 数据集 PSNR 和 SSIM 较低的原因是: Nist16 数据集和 RTD 数据集的图像分辨率很高, 尤其是 Nist16 数据集图像分辨率高达 3000~4000。RTD 图像分辨率达到 1000~2000。本文的修复算法在生成高精度修复图像方面还需要加强。

表 2 图像修复模型在 Casia、Nist16、RTD 三个数据集上的 SSIM 和 PSNR 分数

Table 2 SSIM and PSNR scores of image restoration models on three datasets: Casia, Nist16, and RTD

数据集	SSIM	PSNR
Casia	0.961	30.268
Nist16	0.938	26.246
RTD	0.948	27.869

4.5 对比实验

4.5.1 修复效果可视化对比

我们将所提模型和现有的普通图像修复模型^[13]进行了比较, 结果如图 11 所示。第一列为普通图像修复算法生成的修复图像, 第二列为本文所提模型生成的修复图像。可以看出本文所提模型在图像修复的视觉效果上有很程度的提升。本文还在此基础上将两模型修复出的图像放入图像文本识别模型中, 各自生成其文本信息, 进一步验证所提模型修复的有效性。这是因为: 图像篡改大部分目的是改变图像所表达的语义信息, 从而使人造成误解或影响社会国家安全。例如将“男孩拿着水枪”改为“男孩拿着手枪”。图像篡改在视觉上是将男孩手中的水枪换成手枪, 但在图像描述方向上就是将“水枪”两字换成“手枪”。所以修复后图像的文本识别效果也



图 11 图像修复效果对比图, 其中第一列为普通图像修复模型修复效果, 第二列为本文所提模型修复效果

Figure 11 Comparison of image restoration effects, where the first column shows the restoration effect of the ordinary image restoration model, and the second column shows the restoration effect of the model proposed in this article

是很重要的考核图像是否还原到位的标准。因此将两模型还原修复好的图像使用图像文本识别模型^[17]进行检测, 实验结果如图 12 所示。从结果可以看出, 本文所提模型修复后的图像能够在文本识别模型中较为准确地识别出图像原本表达的文本信息。而普通图像修复算法修复后的图像仍会生成带有篡改目的的文本信息, 这进一步说明本文所提模型在图像修复方向上的有效性。

4.5.2 修复结果数据对比

本文将所提模型和当前主流的图像修复模型进行比较: Pathak^[13]、Liu^[5]、Nazeri^[6], 在 Casia 数据集上计算 PSNR 和 SSIM 的值, 以进行模型性能评估。实验结果如表 3 所示。

从表 3 可以看出, 本文所提出的模型在修复效果方面相比较现有的修复模型均有较大幅度的提

高。在 Casia 数据集上 PSNR 提高了 6.284%, SSIM 提高了 0.086%。这是由于现有的图像修复方法大都没有考虑篡改环境, 且待修复的区域范围相对较小, 无法针对大范围图像缺失或篡改图像进行修复。

4.5.3 时间复杂度对比

为了验证本文所提模型在时间上的优势, 我们进行了时间复杂度对比实验, 将本文所提模型和当前主流图像修复模型进行对比。实验选用 MSCOCO 数据集中的 20000 张照片, batch-size 大小设置为 64, 训练 100 个 epoch, 计算训练所需的整体时间, 实验均在单个 NVIDIA GeForce RTX 3090 Ti GPU 上进行。实验结果如表 4 所示。

从表 4 可以看出, 本文所提算法训练 100 轮所需时间约为 12.25 小时, 相比其他模型, 所需训练时间较少, 时间复杂度较低, 训练效率较高。

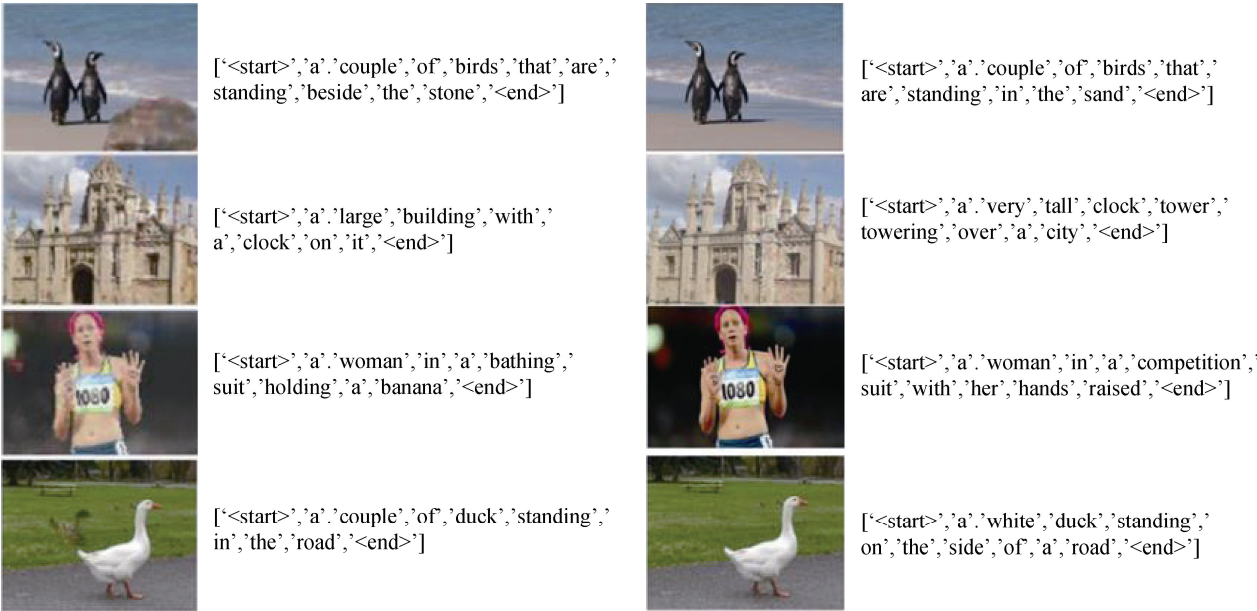


图 12 修复图像的文本识别效果展示图, 第一列为普通图像修复模型文本识别效果, 第二列为本文所提模型文本识别效果

Figure 12 The text recognition effect of the restored image is shown in the first column. The first column shows the text recognition effect of the ordinary image restoration model, and the second column shows the text recognition effect of the model proposed in this article

表 3 在 Casia 数据集上 PSNR 和 SSIM 数值对比

Table 3 Comparison of PSNR and SSIM values on the Casia dataset

	PSNR	SSIM
Pathak ^[13]	20.657	0.762
Liu ^[5]	22.684	0.836
Nazeri ^[6]	23.984	0.875
Proposed	30.268	0.961

表 4 训练过程时间复杂度对比表

Table 4 Comparison table of training process time complexity

	训练所需时间(h)
Pathak ^[13]	12.95
Liu ^[5]	13.24
Nazeri ^[6]	12.74
Proposed	12.25

为了说明测试过程中的时间复杂度, 我们计算测试过程中修复 100 张照片所需时间, 之后取其平均值, 将结果进行比较, 实验结果如表 5 所示。

表 5 测试过程时间复杂度对比表
Table 5 Comparison table of test process time complexity

测试单张照片所需时间(s)	
Pathak ^[13]	0.035
Liu ^[5]	0.046
Nazeri ^[6]	0.031
Proposed	0.025

根据表 5 可以发现, 本文所提方法计算一张图片所需的时间约为 0.025s, 相比其他模型速度较快。根据训练过程和测试过程的时间复杂度实验, 可以证明所提模型时间复杂度较低, 训练模型所需时间较短, 效率较高。

4.6 实验结果可视化

图 13 展示了本文所提模型的图像修复效果, 从图中可以看出所提模型可以完成图像修复工作, 可以较好地还原图像空缺位置像素, 且符合原图表达的文本信息。本文所提模型是一个端到端的模型, 无须任何预处理和后处理, 直接将空缺待修复图放入

模型即可。实验证明了所提模型在图像修复方面有着不错的修复性能。

5 结论

本文提出了一个基于 GAN 网络图像语义指导的取证图像修复模型。为了生成篡改后空缺图的修复图像, 本文利用 GAN 网络, 使用判别器和生成器进行对抗训练, 最终训练平衡得到具备图像修复能力的生成器; 为了更好地指导 GAN 网络修复图像, 本文使用图像语义先验模型, 对空缺篡改图像进行文本信息识别。将识别的特征信息作为 GAN 网络生成器的输入, 对 GAN 网络生成器的修复图像进行指导和约束, 使得修复图更符合原图表达的语义信息; 为了训练更加严格的判别器, 本文在 GAN 网络生成器生成图像放入判别器前, 对图像根据修复掩模区域进行拼接操作, 使得修复图像更加逼真, 真实像素区域更多, 使得判别器判别能力更强, 从而最终得到更理想的生成器; 为了让模型根据实际情况自主决定先验指导模块的指导强度大小, 本文在文本特征信息进入生成器前设置遗忘门, 使得模型更加灵活自由地利用指导信息。本文在不同数据集上的大量实验证明了所提方法的可行性和有效性。



图 13 本文所提模型的修复效果展示图。第一列为篡改图像, 第二列为根据篡改掩模扣除后的空缺篡改图, 第三列为本文所提算法的修复效果图, 第四列为普通图像修复算法的修复效果图, 第五列为原始未篡改图像

Figure 13 The restoration effect display of the model proposed in this article. The first column is the tampered image, the second column is the vacant tampered image after deducting the tampered mask, the third column is the repair effect image of the algorithm proposed in this article, the fourth column is the repair effect image of the ordinary image repair algorithm, and the fifth column is the original unmodified image

虽然本文在图像取证修复领域取得了较好的结果,但仍需在未来工作中针对高精度高分辨率图像的修复工作进行研究补充,增强其完备性。

参考文献

- [1] Ambili B, George N. A Robust Technique for Splicing Detection in Tampered Blurred Images[C]. *2017 International Conference on Trends in Electronics and Informatics*, 2017: 897-901.
- [2] Wu Y, AbdAlmageed W, Natarajan P. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries with Anomalous Features[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9535-9544.
- [3] Jiang X Y, Liu C X. Edge and Region Inconsistency-Guided Image Splicing Tamper Detection Network[J]. *Journal of Image and Graphics*, 2021, 26(10): 2411-2420.
(蒋小玉, 刘春晓. 边缘与区域不一致性引导下的图像拼接篡改检测网络[J]. *中国图象图形学报*, 2021, 26(10): 2411-2420.)
- [4] He J H. *Research and implementation of image copy-move forgery detection*[D]. Dalian: Dalian University of Technology, 2021.
(何家豪. 图像复制移动篡改检测研究及实现[D]. 大连: 大连理工大学, 2021.)
- [5] Liu G L, Reda F A, Shih K J, et al. Image Inpainting for Irregular Holes Using Partial Convolutions[C]. *Computer Vision – ECCV 2018*, 2018: 89-105.
- [6] Nazeri K, Ng E, Joseph T, et al. EdgeConnect: Structure Guided Image Inpainting Using Edge Prediction[C]. *2019 IEEE/CVF International Conference on Computer Vision Workshop*, 2019: 3265-3274.
- [7] Wan Z Y, Zhang B, Chen D D, et al. Bringing Old Photos back to Life[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 2744-2754.
- [8] Wan Z Y, Zhang J B, Chen D D, et al. High-Fidelity Pluralistic Image Completion with Transformers[C]. *2021 IEEE/CVF International Conference on Computer Vision*, 2021: 4672-4681.
- [9] Yu J H, Lin Z, Yang J M, et al. Free-Form Image Inpainting with Gated Convolution[C]. *2019 IEEE/CVF International Conference on Computer Vision*, 2019: 4470-4479.
- [10] Yu J H, Lin Z, Yang J M, et al. Generative Image Inpainting with Contextual Attention[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 5505-5514.
- [11] Zeng Y H, Fu J L, Chao H Y, et al. Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 1486-1494.
- [12] Zheng C X, Cham T J, Cai J F. Pluralistic Image Completion[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 1438-1447.
- [13] Pathak D, Krähenbühl P, Donahue J, et al. Context Encoders: Feature Learning by Inpainting[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2536-2544.
- [14] Karpathy A, Li F F. Deep Visual-Semantic Alignments for Generating Image Descriptions[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 664-676.
- [15] Vinyals O, Toshev A, Bengio S, et al. Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 652-663.
- [16] Biten A F, Gomez L, Rusinol M, et al. Good News, Everyone! Context Driven Entity-Aware Captioning for News Images[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 12458-12467.
- [17] Xu K, Ba J L, Kiros R, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[C]. *The 32nd International Conference on International Conference on Machine Learning - Volume 37*, 2015: 2048-2057.
- [18] Mirza M, Osindero S. Conditional generative adversarial nets[EB/OL]. 2014: ArXiv preprint ArXiv:1411.1784.
- [19] Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis[C]. *International conference on machine learning*. PMLR, 2016: 1060-1069.
- [20] Sun J, Bhattarai B, Chen Z, et al. SeCGAN: Parallel conditional generative adversarial networks for face editing via semantic consistency[EB/OL]. 2021: ArXiv preprint ArXiv:2111.09298.
- [21] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. 2015: ArXiv preprint ArXiv:1511.06434.



王金伟 于 2007 年在南京理工大学获得博士学位。现任南京信息工程大学教授。研究领域为多媒体版权保护、多媒体取证、多媒体加密和数据认证。CCF 专业会员。Email: wjwei_2004@163.com



张子荷 于 2021 年在南京信息工程大学信息安全专业获得学士学位。现在南京信息工程大学电子信息专业攻读硕士学位。研究领域为信息安全、多媒体取证。Email: 1812627146@qq.com



罗向阳 于 2010 年在解放军信息工程大学获得博士学位。现任中国人民解放军战略支援部队信息工程大学教授。研究领域为图像隐写、隐写分析技术。CCF 专业会员。Email: luoxy_ieu@sina.com



马宾 于 2008 年在山东大学获得博士学位。现任齐鲁工业大学教授。研究领域为可逆信息隐藏、多媒体取证、隐写与隐写分析。CCF 专业会员。Email: sddxmb@126.com