

结构约束下的生成对抗深度图修复

卢奇¹, 龚勋^{1,2,3}

(1. 西南交通大学计算机与人工智能学院, 成都 610031; 2. 可持续城市交通智能化教育部工程研究中心, 成都 610031; 3. 四川省制造业产业链协同与信息化支撑技术重点实验室, 成都 610031)

摘要: 不同于纹理图像, 深度图像中的像素点代表采集设备到场景各点的距离, 直接使用通用图像修复方法并不能有效恢复深度图像中缺失区域的场景结构, 本文提出一个两阶段编解码结构的生成对抗网络以解决深度图像修复问题。与常见生成对抗网络(Generative adversarial networks, GAN)模型不同, 本文的生成器网络包括深度生成 G_1 和深度修复 G_2 两个模块。 G_1 模块从RGB图像得到预测深度, 替换待修复深度图像缺失区域, 保证修复区域局部结构一致性。 G_2 模块引入RGB图像边缘结构, 保证全局结构一致性。针对现有图像修复方法没有考虑到修复区域间的一致性问题, 设计结构一致注意力模块(Structure coherent attention, SCA)加入到 G_2 中改善修复效果。本文提出的深度图像修复模型在主流数据集上进行了验证, 利用结构约束并经过两阶段的生成器模型和判别器模型的共同作用, 有效改善了深度图像修复效果。

关键词: 深度图像; 图像修复; 两段式生成对抗网络; 边缘结构信息

中图分类号: TP391 **文献标志码:** A

Generate Adversarial Depth Repair Under Structural Constraints

LU Qi¹, GONG Xun^{1,2,3}

(1. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 610031, China; 2. Engineering Research Center of Sustainable Urban Intelligent Transportation, Chengdu 610031, China; 3. Manufacturing Industry Chains Collaboration and Information Support Technology Key Laboratory of Sichuan Province, Chengdu 610031, China)

Abstract: Unlike RGB images, pixels in depth images represent the distance from the acquisition device to the points of the scene, and the direct use of inpainting methods for the natural image can not effectively restore the scene structure of missing areas in deep images. This paper proposes a two-stage code structure generation counter-network to solve the problem of deep image inpainting. Unlike standard generative adversarial network (GAN) models, the generator network in this paper includes depth build G_1 and depth repair G_2 modules. G_1 obtains the predicted depth from the RGB image, replacing the missing area of the depth image to be repaired, and ensuring the local structure consistency of the repair area. G_2 introduces RGB image edge structure to ensure global structure consistency. The consistency of the missing areas, which is not considered in the existing image inpainting methods, is solved by a structure consistency attention module (SCA) embedded into G_2 . The proposed depth image repairing model is verified on

基金项目: 国家自然科学基金(62376231); 四川省重点研发项目(2023YFG0267, 2023YFS0202); 中央高校基本科研业务费科技创新项目(2682021ZTPY030, 2682022KJ045)。

收稿日期: 2022-03-24; **修订日期:** 2023-03-11

several mainstream data sets, showing that the effect of structural constraints, and the combination of the generator and discriminator is evident.

Key words: depth image; image inpainting; two-stage generative adversarial networks; edge information

引言

目前,深度传感技术已广泛应用于自动驾驶、增强现实以及机器人任务(例如避障、建模)等各个领域。传统的深度传感器利用激光雷达实现,能够获取高精度的目标深度信息。但由于其造价昂贵,因此主要用于关键场景,如自动驾驶汽车。随着深度传感技术的不断发展,近年出现了多种廉价的、消费级的深度相机用于高速深度图像生成,比如微软的 Kinect、谷歌的 Tango 和因特尔的 RealSense^[1]。但这些消费级的深度相机在实际应用时容易受到环境影响,比如当处理半透明、深色、光滑表面物体或者超出量程等都会造成深度图缺失,导致深度图像的进一步使用受到很大限制。

图像修复能够合成图像中缺失区域的内容,可以用于修复损坏、遮挡区域的内容(目前的图像修复算法多集中于彩色图像)。传统图像修复方法利用图像中的低级语义信息,在简单背景图像上具有不错的修复效果。不过在图像缺失区域过大或者针对人脸等某些强语义图像时,传统方法由于缺乏对高级语义的理解,修复区域会出现视觉模糊、缺乏细节等现象。随着神经网络以及生成对抗网络的快速发展,利用对抗网络联合训练的编解码器结构能够提升图像修复的纹理结构一致性。Nazeri 等^[2]在编解码器结构基础上,提出了利用边缘结构信息作为先验信息填充缺失区域。边缘可以看作物体最具辨识度的信息,引入边缘结构信息,有效解决了修复区域所存在的结构扭曲现象。

深度图像修复在一定程度上类似于为普通(彩色)图像修复设计的传播方法。但是,如果将为彩色图像设计的算法简单地移植到深度图像上,其结果质量会受到严重限制。目前许多的用于填充深度图像中缺失区域的方法,其大致划分为仅使用深度的单通道补全算法和结合 RGB 或其他信息的多通道输入深度修复算法^[3]。使用深度通道的深度修复算法将待修复深度输入到网络获取修复后的深度图像,由于缺乏丰富的语义信息,因此修复效果并不理想。与仅使用深度通道的修复方法不同,多通道修复方法加入了 RGB 图像。模型可以充分利用图像中丰富的语义和几何结构信息。由于 RGB 图像容易获取,而且能提供丰富的场景结构信息,因此当前深度补全多集中在基于 RGB 指导下的多通道深度补全。Zhang 等^[4]介绍了一种纹理辅助修复技术,将目标区域划分为平滑类和边缘类,并将不同的偏微分方程分布到每个类别中。Abarghouei 等^[5]对彩色图像进行语义分割来获取目标边缘,使深度值在对应目标能一致传播。这类工作主要是针对每个未知像素进行生成操作。Shvakumar 等^[6]提出了 DFuseNet 的双分支编码器设计,从彩色图像和稀疏深度图像学习局部到全局的上下文信息,结合两个分支获取的特征信息。此外,也有使用其他信息来增强深度完成结果,如曲面法线^[7]和语义信息^[8]。利用自动编码器和生成对抗网络(Generative adversarial network, GAN)体系结构来修复彩色图像十分普遍,但目前利用 GAN 进行深度修复工作还较少。现有的基于深度学习的深度图像修复方法,通常只学到了邻近像素插值或复制粘贴深度值,从可视化的深度图像可以看出具有明显的模糊边界和结构。

针对现有深度修复方法所存在的上述问题,本文提出一个用于深度图像修复的纹理-深度生成对抗网络模型(Color-depth GAN, C-D GAN),结合色彩信息、深度信息和边缘结构信息,利用两阶段网络来优化修复深度,改善深度图修复效果。由于经过深度生成网络得到的粗糙修复深度图像会缺少部分结构信息,导致修复图像存在结构偏差。本文创新性地引入 RGB 图像的边缘结构图作为约束来弥补丢失的结构信息,保证修复图像全局与局部结构一致性。此外,受注意力机制启发,在深度修复网络中,考虑待修复区域之间的补丁相关性,本文设计和引入结构一致注意力模块以获取更好的深度修复结

果。在多个数据集上的综合实验表明,本文提出的模型在主流评价指标上均具有较高精度。

1 本文模型

1.1 网络结构

本文提出了C-D GAN网络针对两个目标改进:估计精确的深度值并构建准确的图像结构。C-D GAN是一个两阶段编解码的生成对抗网络,网络模型如图1所示。本文的两阶段修复网络模型分别称为深度生成网络 G_1 和深度修复网络 G_2 。

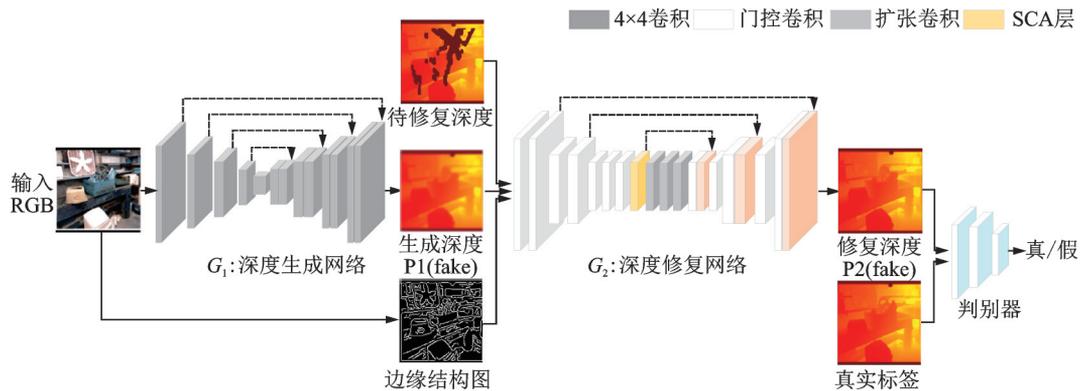


图1 C-D GAN网络模型

Fig.1 C-D GAN network model

深度生成网络实现了单目深度估计任务:色彩图像到深度图像的转换。网络输入RGB图像 x ,输出得到粗糙的预测深度图像 P_1 表示,有

$$P_1 = G_1(x) \quad (1)$$

目前单目深度估计模型主要分为基于回归/分类的监督模型,基于双目训练/视频序列的无监督模型以及基于生成学习的图像风格迁移模型。本文的深度生成网络基于图像风格迁移的方式来生成深度图像。风格迁移可以被认为是一个从内容图像到目标样式图像的分布对齐过程,将一个图像(源域)的样式转移到另一个图像(目标域)上,本质是为了最小化源域和目标域分布之间的距离。本文的深度生成网络利用这一思想,实现色彩图像到深度图像的转换。

在实现上,一个简单的方案是利用卷积神经网络最小化网络输出的像素值和真实值之间的重建损失(欧氏距离)。然而,由于单目深度估计问题固有的多模态性(多个合理的深度图可以对应于单个RGB图像),使用基于单一重建损失预测深度的模型往往倾向于生成预测结果中所有可能模态的平均值,其结果通常是模糊的。通过采用对抗训练的方式,帮助模型从分布中选择一个单模态,能够产生更真实的结果。参考文献[9],深度生成网络采用类似UNet的编解码结构,实现细节在1.3节介绍。

深度生成网络输出得到预测深度图像 P_1 , P_1 替换待修复深度图像的缺失区域,保证了修复区域局部结构一致性。 P_1 被视为粗糙的生成深度图像,是由于经过生成网络后出现结构信息损失。为了弥补丢失的结构信息,本文利用Canny算法从RGB图像提取边缘结构图像。深度修复网络利用 P_1 ,边缘结构图 m 和待修复深度图 n 作为输入,其输出为精细的生成深度图 P_2 ,有

$$P_2 = G_2(P_1, m, n) \quad (2)$$

1.2 SCA模块

本文在深度修复网络中加入了结构一致注意力(Structure coherent attention, SCA)模块。如图2所

示,先前的修复方法通常只考虑特征图中待修复区域和上下文区域间的联系来重建缺失区域,由于忽略了生成区间间的相关性,会导致修复结果缺乏延展性和连续性。受注意力机制以及 Liu 等^[10]的研究启发,通过考虑缺失区域间的相关性,设计结构一致性注意力模块。如图 3 所示,SCA 注意力机制分为两个阶段:搜索和生成。

搜索:对于 $R(i \in [1, n])$ (其中 n 表示补丁数量)中的每个生成区域,SCA 层会在 \bar{R} 中搜索相似度最高的上下文补丁 \bar{R}_i ,用于初始化 R_i 。然后设置权重参数 D_i^{\max} 和 D_i^{ad} 用于在生成阶段更新 R_i , D_i^{\max} 代表 R_i 和上下文区域相似度最高的补丁 \bar{R}_i 的相似度值, D_i^{ad} 代表相邻两个生成补丁间的相似度值,表达式为

$$D_i^{\max} = \frac{\langle R_i, \bar{R}_i \rangle}{\|R_i\| \|\bar{R}_i\|} \quad (3)$$

$$D_i^{\text{ad}} = \frac{\langle R_i, R_{i-1} \rangle}{\|R_i\| \|R_{i-1}\|} \quad (4)$$

生成:从左上角开始扫描,以从上至下从左至右的方式进行生成,最左上角第一个补丁没有先前的生成补丁, $D_i^{\text{ad}} = 0$, $R_1 = \bar{R}_1$ 。从 R_2 开始,由于具有先前的补丁 $m-1$ 作为参考,根据式(5)来更新 m_2 。整个生成过程会不断迭代,从 R_1 到 R_n 。

$$R_i = \frac{D_i^{\text{ad}}}{D_i^{\text{ad}} + D_i^{\max}} \times R_{i-1} + \frac{D_i^{\max}}{D_i^{\text{ad}} + D_i^{\max}} \times \bar{R}_i \quad (5)$$

1.3 实现细节

深度生成网络结构使用 256 像素 \times 256 像素的输入图片输出具有同样维度的生成深度图片 P_1 。编码阶段包含 8 个卷积层,使用 Leaky Relu 作为激活函数,其内核大小均为 4,通道数分别为 64、128、256、512、512、512、512、512,步长为 2。解码器遵循编码阶段类似的结构,使用反卷积进行上采样,并减少特征映射(128、64、3)。反卷积最后一层使用 tanh 作为激活函数。

深度修复网络结构由编码器和解码器组成。编码阶段使用了的 3×3 普通卷积和 4×4 的扩张卷积。 3×3 卷积核维持空间尺寸但将通道数扩增一倍,能够提高网络获得深度语义信息的能力。 4×4 扩张卷积将空间减少一半但保持通道数不变,能够扩大感受野,防止信息的过度损失。通过实验设置,本文结构一致性注意力层放到修复网络编码器中第 3 层。生成对抗网络中的鉴别器用于判别生成的精细修复深度和原始深图像。参考文献[11]的判别器网络结构,卷积网络作为判别器,C-D GAN 判别器由普通卷积层和最大池化层组成,利用 VGG 网络提取特征图来区分修复完整图像和原始图像。

1.4 损失函数

本文将损失函数分为两部分:对抗损失和重建损失。受文献[12]的启发,本文未直接使用 L_1 范数作为重构损失函数,而是将重构损失划分为缺失区域重构损失 L_{rec} 和上下文区域重构损失 L_{context} 两部分(其中 \cdot 表示按元素相乘),表达式为

$$L_{\text{in}} = |G_2(G_1(x), m, n) - x| \cdot m \quad (6)$$

$$L_{\text{context}} = |G_2(G_1(x), m, n) - x| \cdot (1 - m) \quad (7)$$

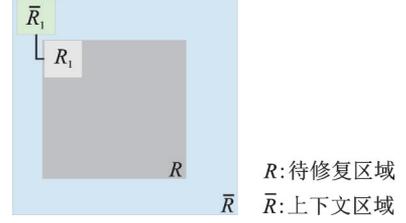


图 2 先前的修复方法

Fig.2 Previous inpainting methods

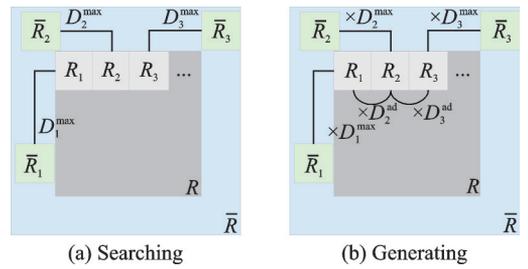


图 3 SCA 模块

Fig.3 SCA module

通过给定所有修复区域内像素的重建损失指定一个较小的恒定权重,避免了使用 L_1 范数重构损失所带来额外的空间开销,整个重构损失定义为

$$L_{\text{rec}} = \alpha_1 L_{\text{in-hole}} + \alpha_2 L_{\text{context}} \quad (8)$$

式中 α_1 和 α_2 分别是对应损失的权重系数($\alpha_1=1, \alpha_2=1.2$,参数设置在实验部分引入)。

本文使用WGAN-GP^[13]作为对抗损失。WGAN-GP是WGAN损失函数的改进版本,通过引入梯度惩罚来处理WGAN引起的梯度爆炸问题。WGAN-GP进一步解决了GAN训练不稳定,梯度消失的现象。WGAN-GP对抗损失函数定义为

$$L_{\text{adv}} = E_{\tilde{x} \in p_g} [D(\tilde{x})] - E_{x \in p_r} [D(x)] + \sigma E_{\hat{x} \in p_{\hat{x}}} \left[\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|_2 - 1 \right]^2 \quad (9)$$

式中: D 为判别器输出; σ 为惩罚占比; x, \tilde{x}, \hat{x} 分别是真实图像、生成图像以及对应插值; $p_g, p_r, p_{\hat{x}}$ 则分别是其对应的分布。

联合重构损失和对抗损失,本文损失函数定义为

$$L_g = \alpha L_{\text{rec}} + \beta L_{\text{adv}} \quad (10)$$

式中: α 为重建损失权重系数; β 为对抗损失权重系数(初始 $\alpha = 0.05, \beta = 0.95$),利用Adam梯度优化算法更新参数。

2 实验与结果分析

2.1 数据集和预处理

为了验证本文方法的有效性,本文在两个不同类型的数据集上进行评估:NYU Depth Dataset V2^[14]和Bosphorus^[15]。NYU Depth Dataset V2数据集是使用微软的Kinect从464个不同室内场景收集RGB和深度图像组成。按照官方数据分割采样约3万张图像。原始尺寸640像素 \times 480像素的图像下采样一半,RGB和Depth图片均上下添加8像素,然后进行中心裁剪。数据集图像被划分为256像素 \times 256像素的RGB图和对应的深度图。Bosphorus数据库专门用于研究三维和二维人脸处理任务。数据库中包含105个人的4666张三维人脸。数据集图像被划分为256像素 \times 256像素的RGB图和对应的深度图。

掩膜 为了公平评估,本文对上述数据集分别在中心掩膜和随机掩膜上进行了实验。对于中心掩膜,实现在代码中将深度图中心像素值全部设置0;不规则的掩膜设置使用了文献[16]的工作,根据孔洞与图像的面积比(如0%~10%,10%~20%)来划分不同类型的掩膜。掩膜图像的大小为256像素 \times 256像素。

2.2 参数设置

训练模型采用Adam算法^[17]进行优化,初始学习率为0.0002,超参数梯度均值系数为0.5,衰减乘数因子为0.1,衰减变化阈值为0.01,衰减周期为5。

2.3 主观视觉评价

本文分别从NYU场景数据集和Bosphorus人脸数据集选取500张图像,在中心掩膜和不规则掩膜上进行了测试评估。由于目前并没有有效的将GAN应用于深度图像修复领域的文献,本文使用近两年在RGB图像领域取得显著效果且开放代码的Li^[18]、Lahiri^[16]、Zamir^[19]作为基准,对本文所提出的C-D GAN进行实验对比。图4展示了使用中心掩膜时,不同方法下的定性比较结果;图5展示了使用随机掩膜时,不同方法下的定性比较结果。图4和图5中的第1行和第2行是Bosphorus图像的测试结果,其他行是NYU图像的测试结果。如图4和图5所示,通过可视化深度图修复结果进行主观视觉评价。Bosphorus人脸使用原始深度图修复结果用于评估,其他几种文献都出现了结构扭曲现象,与未修复区域出现了结构断层。NYU数据集修复结果进行热力图变换,感知深度图中的细节变换。可以发现,转

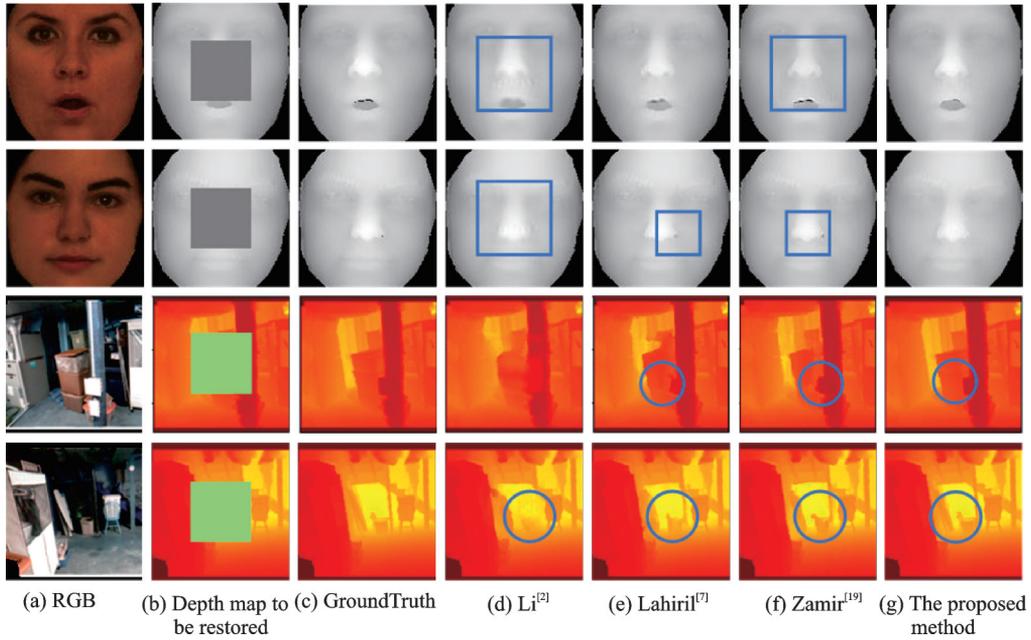


图 4 中心掩膜的定性比较
Fig.4 Qualitative comparison of center masks

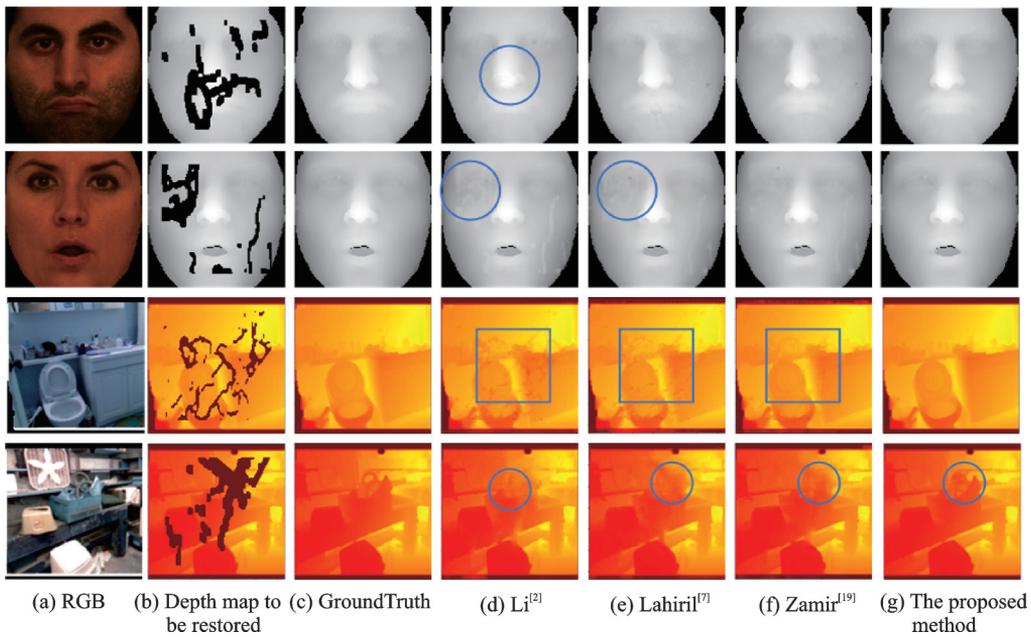


图 5 随机掩膜的定性比较
Fig.5 Qualitative comparison of random masks

换为热力图后,其他文献方法^[16,18,19]均出现纹理模糊和结构不一致问题。

2.4 客观视觉评价

除了主观的视觉效果评估,还需利用度量指标客观展示本文方法的有效性。本文采用的两个评价

指标分别为峰值信噪比 PSNR 和结构相似性指数 SSIM: PSNR 能够测量像素值的差异; SSIM 则能用于估计两幅图像之间的整体相似性,这两个指标越大代表修复效果越好。

表 1 和表 2 分别列出了中心掩模和不规则掩模下的指标度量结果。 \uparrow 代表越大越好(下同)。在使用中心掩时,本文算法均优于其他几种对比算法。PSNR 指数比 Zamir^[19] 模型提高了 2.11%,比 Lahiri^[16] 模型提高了 2.61%,比 Li^[18] 模型提高了 5.53%。

在使用随机掩膜时,本文算法也均优于其他几种对比算法。当掩膜占比为 0%~15%,本文算法 SSIM 指数达到 0.978,PSNR 指数达到 32.54;当掩膜占比为 15%~30% 时,本文算法 SSIM 指数达到 0.914,PSNR 指数达到 29.12;当掩膜占比为 30%~50%,本文算法 SSIM 指数达到 0.813,PSNR 指数达到 26.34。综上,无论在不规则掩膜或中心掩膜的实验结果中,本文的方法均优于其他方法。

2.5 实验结果分析

2.5.1 α_1, α_2 权重参数分析

如式(8)所述,本文将重构损失划分为缺失区域重构损失和上下文区域重构损失两部分。为了验证权重 α_1, α_2 对修复性能的影响,本文比较了在中心掩膜下不同参数条件的修复结果。通过对 α_2 设置一个相对 α_1 较大的权重,代表上下文区域重构损失对图像修复过程有更大的贡献。参考文献[12]中的设置, α_1 取值固定为 1, α_2 取值设置[1, 1.5]之间变化。如图 6 所示,在不同权重参数下对修复效果造成了不同的影响(并未选择训练 loss 进行比较,而是利用图像修复指标 PSNR 比较来更具说服力)。最后权重参数 α_1 为 1, 权重参数 α_2 更新为 1.2。

2.5.2 边缘结构信息影响

生成器网络引入 RGB 的边缘结构图提升精细修复网络的修复效果。为验证 RGB 边缘结构图作用,设计 3 组对比试验。其中待修复深度图像的边缘结构信息参考文献[20]设置,对修复深度图使用 Canny 算法提取边缘信息。如图 7 所示,加入边缘结构信息后,修复后的深度图有更好的结构一致性。而相比与使用待修复深度图像的边缘结构信息,利用 RGB 边缘结构信息由于保留较多的场景结构,因此修复后的图像有更细致的结构完整性。表 3 展示了不同结构信息的客观指标结果,使用 RGB 的边缘结构信息展示出更好的评价度量。

表 1 中心掩膜在 NYU 数据集上的对照结果

Table 1 Comparison results of center masks on the NYU dataset

方法	SSIM \uparrow	PSNR/dB \uparrow
Li ^[18]	0.83	23.92
Lahiri ^[16]	0.86	26.84
Zamir ^[19]	0.88	27.34
本文方法	0.92	29.45

表 2 随机掩膜在 NYU 数据集上的对照结果

Table 2 Comparison results of random masks on the NYU dataset

方法	掩膜占比/%	SSIM \uparrow	PSNR/dB \uparrow
Li ^[18]	0~15	0.913	28.17
	15~30	0.866	25.43
	30~50	0.734	21.12
Lahiri ^[16]	0~15	0.923	28.45
	15~30	0.878	26.32
	30~50	0.742	22.43
Zamir ^[19]	0~15	0.944	30.15
	15~30	0.872	27.67
	30~50	0.786	24.12
本文方法	0~15	0.978	32.54
	15~30	0.914	29.12
	30~50	0.813	26.34

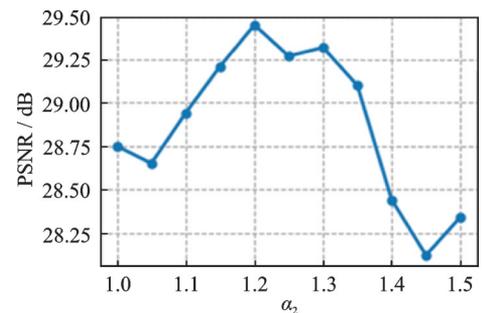


图 6 不同权重系数对修复结果的影响

Fig.6 Effect of weighting factors on restoration results

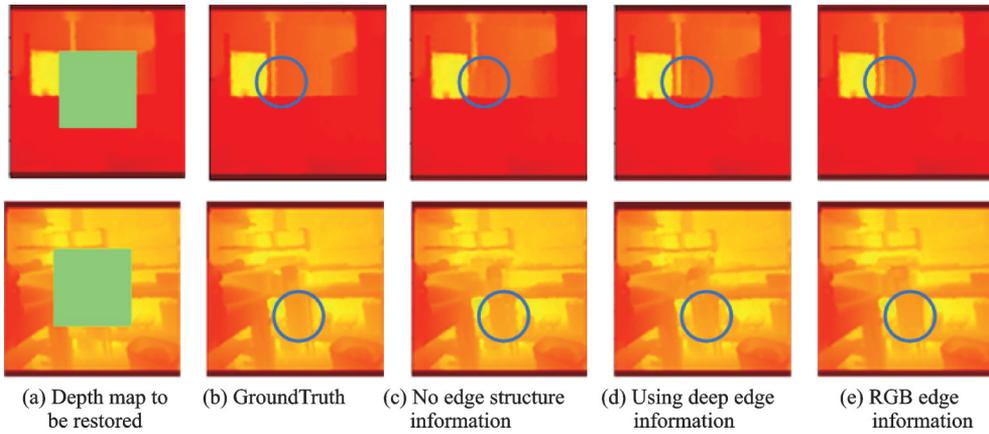


图 7 不同边缘结构信息的影响

Fig.7 Effect of different edge structure information

表 3 边缘结构信息在 NYU 数据集上的对比结果

Table 3 Control results of edge structure information on NYU dataset

方法	SSIM ↑	PSNR/dB ↑
不加入边缘结构图	0.82	21.34
待修复深度的边缘结构图	0.89	28.24
RGB 的边缘结构图	0.92	29.45

2.5.3 SCA 模块影响

在生成器的修复网络中,本文使用结构一致性注意力层来提高网络的修复效果。为了验证该模块有效性,通过将注意力层替换为普通卷积层和上下文注意力层^[21]进行实验对比。如图 8 所示,使用普通卷积层时,缺失区域部分存在严重的纹理模糊和结构扭曲现象。与普通卷积层相比,上下文注意力层提高了性能,但修复结果仍然存在结构不一致情况,与上下文区域边缘不匹配。结构一致性注意力层引入得到了更精细的修复效果。表 4 展示了引入结构一致性注意力层的量化指标评价结果。

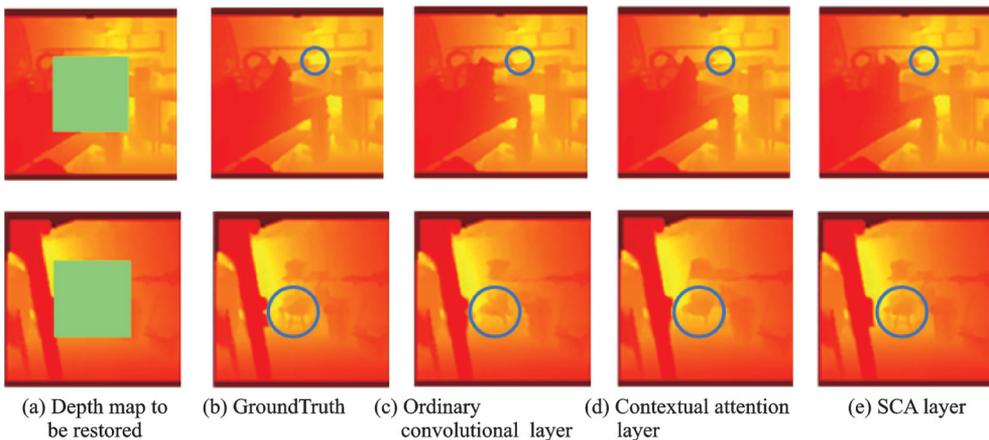


图 8 SCA 模块影响

Fig. 8 Impact of SCA module

表4 SCA模块在NYU数据集上的对比结果

Table 4 Comparison results of SCA module on NYU dataset

方法	SSIM ↑	PSNR/dB ↑
普通卷积层	0.83	25.64
上下文注意力层	0.88	27.83
SCA层	0.92	29.45

3 结束语

本文提出一种基于生成对抗网络的深度图像修复模型C-D GAN。生成器模型是一个两阶段的编解码结构,由深度生成网络和深度修复网络构成。生成网络学习彩色图像到深度图像的映射,输入彩色图像得到粗糙的预测深度图像。修复网络进一步优化缺失深度图像的结构信息,利用边缘结构信息作为约束,与待修复的深度图像堆叠输入到修复网络,实现了较高精度的修复目标。在NYU场景深度数据集和Bosphorus人脸深度数据集上,本文与另外几种基于GAN的图像修复方法进行多次实验对比。无论是主观视觉上的评估还是客观指标度量,本文的深度图像修复模型算法都具有更好的修复效果。虽然本文提出的深度图像修复算法效果上不错,但由于模型参数多,未来工作将利用更好的卷积网络结构,在减少模型参数的同时提取更好的深度图像特征。

参考文献:

- [1] ZHANG Y, FUNKHOUSER T. Deep depth completion of a single RGB-D image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 175-185.
- [2] NAZERI K, NG E, JOSEPH T, et al. Edgeconnect: Structure guided image inpainting using edge prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. [S.l.]: IEEE, 2019: 3265-3274.
- [3] LU K, BARNES N, ANWAR S, et al. From depth what can you see? Depth completion via auxiliary image reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 11306-11315.
- [4] ZHANG Y, KHAMIS S, RHEMANN C, et al. Activestereonet: End-to-end self-supervised learning for active stereo systems [C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 802-819.
- [5] ABARGHOUEI A A, BRECKON T P. Depthcomp: Real-time depth image completion based on prior semantic scene segmentation[C]//Proceedings of British Machine Vision Conference. [S.l.]: BMVA Press, 2017: 58.1-58.13.
- [6] SHIVAKUMAR S S, NGUYEN T, MILLER I D, et al. Dfusetnet: Deep fusion of RGB and sparse depth information for image guided dense depth completion[C]//Proceedings of 2019 IEEE Intelligent Transportation Systems Conference. [S.l.]: IEEE, 2019: 13-20.
- [7] QIU J, CUI Z, ZHANG Y, et al. Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 3313-3322.
- [8] JARITZ M, DE CHARETTE R, WIRBEL E, et al. Sparse and dense data with CNNs: Depth completion and semantic segmentation[C]// Proceedings of 2018 International Conference on 3D Vision. [S.l.]: IEEE, 2018: 52-60.
- [9] 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.
- [10] LIU H, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2019: 4170-4179.
- [11] MA F, CAVALHEIRO G V, KARAMAN S. Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera[C]//Proceedings of 2019 International Conference on Robotics and Automation. [S.l.]: IEEE, 2019: 3288-

3295.

- [12] YI Z, TANG Q, AZIZI S, et al. Contextual residual aggregation for ultra high-resolution image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 7508-7517.
- [13] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs [C]//Proceedings of Advances in Neural Information Processing Systems. Long Beach, USA:[s.n.],2017: 5767-5777.
- [14] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from rgb-d images[C]//Proceedings of the European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 746-760.
- [15] DRAN S A, ALYUZ N, DIBEKLIO H G, et al. Bosphorus database for 3D face analysis[C]//Proceedings of the First European Workshop on Biometrics and Identity Management. Berlin, Heidelberg: Springer, 2008: 47-56.
- [16] LAHIRI A, JAIN A K, AGRAWAL S, et al. Prior guided GAN based semantic inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 13696-13705.
- [17] KINGMA D P, BA J. ADAM: A method for stochastic optimization[EB/OL].(2014-12-22) [2022-01-15]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [18] 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.
- [19] ZAMIR S W, ARORA A, KHAN S, et al. Multi-stage progressive image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 14821-14831.
- [20] 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.
- [21] YAN Z, LI X, LI M, et al. Shift-net: Image inpainting via deep feature rearrangement[C]//Proceedings of the European Conference on Computer Cision. Cham: Springer, 2018: 3-19.

作者简介:



卢奇(1997-),男,硕士研究生,研究方向:计算机视觉、图像修复和深度学习,E-mail: 3293476568@qq.com。



龚勋(1980-),通信作者,男,博士,教授,研究方向:人工智能、视频大数据、多模态人脸识别,E-mail: xgong@swjtu.edu.cn。

(编辑:刘彦东)