



基于谱归一化条件生成对抗网络的图像修复算法

雷蕾¹, 郭东恩^{1,2}, 靳峰¹

(1. 南阳理工学院 软件学院, 河南 南阳 473000;

2. 重庆邮电大学 计算机科学与技术学院 计算智能重庆市重点实验室, 重庆 400065)

摘要: 基于生成对抗网络的图像修复算法在修复大尺寸缺失图像时, 存在图像失真较多与判别网络性能不可控等问题, 基于谱归一化条件生成对抗网络, 提出一种新的图像修复算法。引入谱归一化来约束判别网络的判别性能, 间接提高修复网络的修复能力, 并根据控制判别网络性能对谱归一化进行理论分析。通过类别信息约束特征生成, 保证修复图像的内容不变性, 引入扩展卷积算子对待修复图像进行像素级操作, 解决修复图像缺乏局部一致性的问题。在此基础上, 运用 PSNR、SSIM 等图像评价方法及分片 Wasserstein 距离、Inception 分数、流形距离度量、GAN-train 和 GAN-test 等流形结构相似度评价指标对修复图像进行综合评价。实验结果表明, 与 CE、GL 等算法相比, 该算法获得的修复图像在主观感受和客观评价指标上均有明显提高。

关键词: 谱归一化; 条件生成对抗网络; 图像修复; 判别性能; 图像评价

开放科学(资源服务)标志码(OSID):



中文引用格式: 雷蕾, 郭东恩, 靳峰. 基于谱归一化条件生成对抗网络的图像修复算法[J]. 计算机工程, 2021, 47(1): 230-238.

英文引用格式: LEI Lei, GUO Dongen, JIN Feng. Image inpainting algorithm based on conditional generative adversarial network with spectral normalization[J]. Computer Engineering, 2021, 47(1): 230-238.

Image Inpainting Algorithm Based on Conditional Generative Adversarial Network with Spectral Normalization

LEI Lei¹, GUO Dongen^{1,2}, JIN Feng¹

(1. School of Software, Nanyang Institute of Technology, Nanyang, Henan 473000, China;

2. Chongqing Key Laboratory of Computational Intelligence, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

[Abstract] To solve the problem of large image distortion and uncontrollable discriminative network performance in image inpainting based on Generative Adversarial Network (GAN), this paper proposes a new image inpainting algorithm based on conditional generative adversarial network with spectral normalization. Spectral normalization is introduced to constrain the discriminative performance of discriminative network, and thus bring an improvement to the inpainting network performance followed by detailed theoretical analysis of spectral normalization in controlling the discriminative network performance. Category information is used to constrain feature generation to ensure that content of the inpainted image is close to that of the original image. The extended convolution operator is also introduced to perform pixel-level operation on the to-be-inpainted image to address the lack of local consistency in image inpainting. On this basis, PSNR, SSIM and other image evaluation methods, as well as slice Wasserstein distance, Inception score, manifold distance measurement, GAN-train, GAN-test and other manifold structure similarity evaluation indicators are used to comprehensively evaluate the inpainted image. Experimental results show that compared with CE, GL and other algorithms, the proposed algorithm can significantly improve the subjective and objective evaluation indicators of the inpainted images.

[Key words] spectral normalization; conditional Generative Adversarial Network (GAN); image inpainting; discrimination performance; image evaluation

DOI: 10.19678/j.issn.1000-3428.0056287

基金项目: 国家自然科学基金(61671091); 重庆市自然科学基金(cstc2017jcyjBX0037, cstc2017jcyjA0982); 河南省高等学校重点科研项目(19A520030); 重庆市教委一般项目(CYB19173); 重庆邮电大学博士创新人才项目(BYJS201812)。

作者简介: 雷蕾(1986—), 女, 硕士研究生, 主研方向为图像处理、机器学习、大数据分析; 郭东恩, 副教授、博士; 靳峰, 讲师、博士。

收稿日期: 2019-10-14 **修回日期:** 2019-12-31 **E-mail:** 61069597@qq.com

0 概述

图像修复属于图像复原领域,是数字图像处理和计算机视觉中的一个重要研究方向。图像修复主要是对图像上损坏或者缺失的部分根据像素变化的规律进行填充,从而恢复图像原始面貌^[1]。早期的图像修复技术源于文艺复兴时期对艺术作品的修复,随着各类学科的不断发 展,人们对修复图像在视觉感知、纹理结构和语义理解上有了更高的要求。计算机硬件设备以及深度学习等相关技术的不断发展,派生出一系列通过深度学习方法来进行图像修复的算法,这类算法的核心在于训练一个深度神经网络来获取受损图像和原始图像之间的非线性映射。在众多基于深度学习的图像修复算法中,基于生成对抗网络(Generative Adversarial Network, GAN)的修复算法在修复效果上取得了较大进展,修复图像具有良好的主、客观评价,因此该算法受到国内外学者的广泛关注。

本文提出基于谱归一化条件生成对抗网络的图像修复算法,利用条件生成对抗网络修复彩色图像,并引入谱归一化约束判别网络参数梯度的 Lipschitz 上界,可间接提高修复网络的修复能力。此外,从理论上分析谱归一化,将判别网络参数梯度的 Lipschitz 上界约束为 1,通过实验来验证该算法能够获得较好的修复结果和梯度收敛。

1 相关研究

图像修复算法大致可分为基于扩散的方法、基于纹理合成和基于深度学习的方法。基于扩散的方法是一种古老、经典的方法,包括变分和偏微分算法。这类算法能够较好地修复图像损失的细节,适用于小尺寸缺失的图像修复^[2-3],但更多时候仅限于老、旧照片的修复及去除文字水印等。基于纹理合成的方法适用于修复大尺寸缺失的图像^[4-5],这类方法也称为基于补丁的修复方法^[6-7],原理在于通过使用一个较大的图像数据库来进行查询填充,然后从匹配的图像中复制相应的区域并将它们粘贴到对应缺失部分来完成图像修复。此外,研究人员还提出了结合变分偏微分与纹理合成的算法,称为基于图像分解的修复方法^[8-9],通过将已损坏的图像分解为大尺度缺失和小尺度缺失的部分,然后对这两部分缺失区域分别采用基于变分偏微分方程的修复算法和基于纹理合成修复算法进行修复。

基于深度学习的修复方法通过训练深度神经网络来学习受损图像与原始图像之间的非线性映射关系。卷积神经网络(Convolutional Neural Network, CNN)在图像识别和特征学习方面具有较大的优势,和其他的深度神经网络相比,它能更好地学习到图像的 2D 结构和隐含特征,因此有学者利用卷积神经

网络实现图像修复^[10-13]。生成对抗网络^[14]的出现,对图像修复是一个很大的促进。GAN 模型被定义成一个 two-player 的博弈游戏,目的在于生成一个与真实数据接近的分布。训练过程是将随机低维噪声 z 输入到生成器 G 中,然后生成器 G 输出高维样本 $G(z)$,判别器 D 判断给定样本是来自于真实的数据分布还是生成器产生的生成分布, G 和 D 都被参数化为深度神经网络。据此,基于 GAN 的图像修复算法优化过程也可以被参数化为一个最小化最大化问题,即:

$$\min_G \max_D C(D, G) =$$

$$E_{x \sim p_g(x)} [\log_a (1 - D(G(x, m)))] + E_{x \sim p_r(x)} [\log_a D(x)] \quad (1)$$

其中, $p_r(x)$ 和 $p_g(x)$ 分别表示真实图像分布和修复图像分布, $E(\cdot)$ 表示离散型分布期望值, m 表示掩码图像,用于仿真受损图像。

基于 GAN 的图像修复算法^[15]于 2016 年由 PATHAK 等人提出,主要是利用加权重建损失和对抗损失来修复图像。该方法可以生成新的内容和修复缺失较大的区域,克服了基于纹理合成和基于变分偏微分方程的局限性,但是并没有考虑到局部连续性。STAOSHI 等人提出 GL 模型^[16]解决了这个问题,但该模型训练时间长,网络参数多,并且该模型的判别网络包含两个判别器,均没有对其判别性能进行约束。在高维空间中,判别网络的密度比估计在训练期间通常是不准确和不稳定的^[17],因此生成网络通常无法有效学习目标分布的多模结构。此外,当生成分布的支撑集和真实数据分布的支撑集不相交时,存在一个能够完全区分生成分布与真实数据分布的判别网络^[18]。在这种情况下判别网络判别性能“太强”,将会导致生成网络相对于输入的偏导为 0,使得生成网络训练完全停止。因此,需要约束判别网络的性能,使之在短时间内不能被训练得“太好”。考虑到基于生成对抗网络的图像修复算法中判别性能不可控,当缺失部分较多、失真较大时修复结果缺乏局部一致性问题,本文通过引入谱归一化和类别辅助信息到生成对抗网络中,提出基于谱归一化条件生成对抗网络的图像修复算法,称为 SN-CGAN。

本文提出的 SN-CGAN 算法与 CE 算法相比能够获得更好的修复图像,修复图像具有较强的局部一致性,与 GL 模型相比,SN-CGAN 模型训练时间更短,稳定性更强,从根本上避免了梯度消失的问题,与 CA 模型相比,SN-CGAN 模型中判别网络性能可控,能够加速修复网络训练,间接地提高了修复网络的修复性能,进一步提高修复图像质量。

2 本文算法

本文提出的 SN-CGAN 算法模型首先引入类别

辅助信息来约束特征生成,保证修复图像的内容接近原始图像的内容,然后将判别网络中的标准归一化^[18]替换为谱归一化^[19],进一步约束判别网络的判别性能,使之在训练过程中不被训练得太快,从而提高修复网络的修复性能,最后引入扩展卷积算子^[20]来对待修复图像的每个像素进行操作,进一步解决修复图像缺乏局部一致性的问题。如图1所示,修复网络采用常见的编码-解码结构,编码器利用7个卷积层来减少图像分辨率并学习图像特征,并运用4个扩展卷积层(Dilated1~Dilated4)^[20]在不改变卷积核数量情况下通过特征映射传播卷积核,扩展每一层的输入区域。

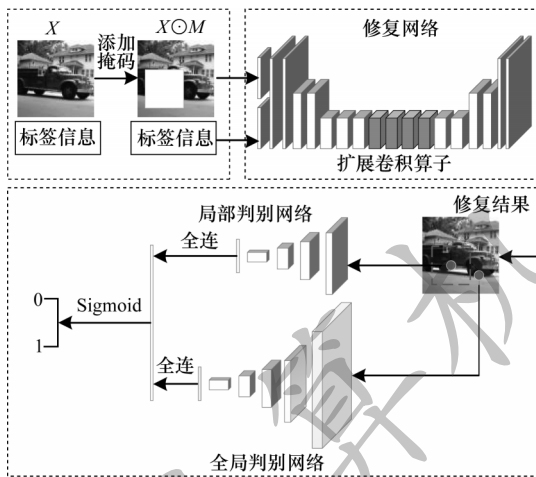


图1 SN-CGAN算法模型示意图

Fig.1 Schematic diagram of SN-CGAN algorithm model

假设2维卷积层处理一个C通道 $h \times w$ 的特征映射得到的结果为C'通道 $h' \times w'$ 的特征映射,那么针对每个像素的扩展卷积算子可以表示为:

$$y_{u,v} = \sigma \left(\sum_{i=-k'_h}^{k'_h} \sum_{j=-k'_w}^{k'_w} W_{k'_h+i, k'_w+j} x_{u+\eta i, v+\eta j} + b \right) \quad (2)$$

其中, $k'_h = \frac{k_h - 1}{2}$, $k'_w = \frac{k_w - 1}{2}$, $k_h \times k_w$ 为卷积核的大小, W 和 b 表示非线性映射的权重矩阵和偏移量, 由于采用全卷积网络, b 为全0矩阵, x 和 y 分别表示输入层和输出层对应像素特征值, η 为扩展因子, 当 $\eta = 1$ 时, 式(2)为标准的卷积算子, $\sigma(\cdot)$ 表示非线性激活函数。

本文提出的修复模型由修复网络、局部判别网络和全局判别网络三部分组成。修复网络用于修复受损图像缺失的部分以达到修复图像的目的, 局部判别网络用于判别局部修复部分与局部真实部分的真伪, 而全局判别网络则对整个原始图像和修复图像的真伪进行判别, 通过联合训练修复网络以及局部、全局判别网络, 以此来保证修复图像的全局-局部一致性。

2.1 内容损失

内容损失用来提高修复图像的质量, 减少真实图像与修复图像之间的失真。内容损失包括像素方均误差损失(pixel-wise Mean Square Error, pMSE)、视觉感知损失(perception loss)^[21]和全变分损失(Total Variation loss, TV)3个部分。

pMSE损失是图像处理中广泛使用的一种代价损失函数, 常用于快速图像修复、图像超分辨及图像盲复原等, 基于pMSE的损失函数可表示为:

$$L_{\text{pMSE}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H (I_{x,y}^o - G_{\theta_g}(I'_{x,y}))^2 \quad (3)$$

其中, $I_{x,y}^o$ 和 $I'_{x,y}$ 分别表示原始图像 I^o 和待修复图像 I' 中第 (x,y) 像素点的像素值, W 和 H 分别表示 I^o 和 I' 的图像大小, θ_g 表示生成网络的参数。

尽管利用pMSE代价函数通常能够产生较高的PSNR值, 但修复图像往往缺乏高清细节, 容易导致纹理结构过度平滑, 造成较差的视觉感知。因此, 利用视觉感知损失代价^[21]来获取视觉上满意的修复结果, 视觉感知损失定义如下:

$$L_{\text{VGG}_{i,j}} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^o) - \phi_{i,j}(G_{\theta_g}(I')))^2 \quad (4)$$

其中, $\phi_{i,j}$ 表示预训练VGG19网络^[22]中第 i 个最大池化层之前及第 j 个卷积层后(激活之后)获得的特征图, I^o 和 I' 分别表示原始图像和待修复图像, $W_{i,j}$ 和 $H_{i,j}$ 表示VGG网络中各特征映射的维度。

已有文献证明了在图像修复过程中, 全变分损失能够较好地抑制噪声。因此, 本文引入全变分损失来抑制图像的噪声, 全变分损失定义为:

$$L_{\text{TV}} = (I^o - G_{\theta_g}(I'))^2 \quad (5)$$

利用内容损失代价函数能够减少修复图像与原始图像的失真, 尽管引入了视觉感知损失来保证视觉感知效果, 但修复结果的高频细节依然比较缺乏。本文提出条件对抗损失来获取修复图像的高频细节。

2.2 条件对抗损失

文献[23]证明了原始的生成对抗网络^[14]是不稳定的, 容易出现梯度消失的问题, 并提出利用Wasserstein距离度量^[24]替换原始生成对抗网络中的Kullback-Leibler散度。因此, 式(1)也可以表示为:

$$\min_G \max_D C(D, G) = -E_{x \sim p_r(x)} [\log_a D(x)] + E_{x \sim p_g(x)} [\log_a D(G(x, m))] \quad (6)$$

对抗损失驱动修复网络通过“愚弄”判别网络来生成符合原始图像流形结构的特征。为使得修复图像更好地接近原始图像, 引入类别辅助信息来限制图像特征生成^[25]。根据所有训练样本 N 在判别网络上的概率定义条件对抗损失:

$$L_{\text{CGen}} = \frac{1}{N} \sum_{n=1}^N \log_a D_{\theta_d}(G_{\theta_g}(I^n | l)) \quad (7)$$

其中, l 表示类别辅助信息, G 表示修复网络, D 表示

判别网络, 包含局部判别网络 D_1 和全局判别网络 D_2 。因此, 式(7)可以进一步表示为:

$$L_{CGen} = \frac{1}{N} \sum_{n=1}^N \log_a D_{\theta_{d1}}(G_{\theta_g}(I^{I1}|l)) \times \frac{1}{N} \sum_{n=1}^N \log_a D_{\theta_{d2}}(G_{\theta_g}(I^{I2}|l)) \quad (8)$$

其中, θ_{d1} 和 θ_{d2} 分别表示局部判别网络和全局判别网络的参数, I^1 和 I^2 分别表示图像待修复区域部分和整体待修复图像。式(8)中没有包含谱归一化的相关范式, 因为谱归一化式子仅在判别网络中利用谱范数来对权重矩阵进行归一化。

2.3 整体代价函数

通过引入超参数 α 、 β 和 λ 与上述损失函数相结合, 得出本文算法的整体代价函数为:

$$L_{Total} = L_{CGen} + \alpha L_{MSE} + \beta L_{VGG/i,j} + \lambda L_{TV} \quad (9)$$

式(9)中的超参数均用来权衡各个损失之间的比例。不难证明式(9)是一个多参数的非凸函数, 直接对这类函数进行求解是一个NP难问题。因此, 采用随机梯度方法和交替迭代的策略进行参数学习, 令批次大小 $m=64$ 。具体的学习算法步骤如算法1所示, 算法1采用Adam优化器^[26]进行优化。当所有网络交替训练好后保存修复网络, 即能修复受损图像。

算法1 SN-CGAN算法

输入 真实图像集 X , 随机掩码图像 M , 标签类别 L

输出 修复网络 G , 局部判别网络 D_1 , 全局判别网络 D_2

步骤1 对真实图像集和标签进行采样, 获得分批图像和标签。

步骤2 对每一批图像中的每一张图像添加随机掩码, 获得仿真的受损图像。

步骤3 计算局部判别网络的梯度:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{k=1}^m [-\log_a D_{\theta_{d1}}(I^{O1}|l)^k + \log_a D_{\theta_{d1}}(G_{\theta_g}(I^{I1}|l)^k)]$$

步骤4 利用梯度上升方法来更新局部判别器。

步骤5 计算全局判别网络的梯度:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{k=1}^m [-\log_a D_{\theta_{d2}}(I^{O2}|l)^k + \log_a D_{\theta_{d2}}(G_{\theta_g}(I^{I2}|l)^k)]$$

步骤6 利用梯度上升方法来更新全局判别器。

步骤7 计算修复网络的梯度:

$$\begin{aligned} \nabla_{\theta_g} \frac{1}{m} \sum_{k=1}^m & \left[\log_a D_{\theta_{d1}}(G_{\theta_g}(I^1|l)^k) + \right. \\ & \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_g}((I^1|l)^k))^2 + \\ & \left. \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}((I^1)^k) - \phi_{i,j}(G_{\theta_g}(I^1|l)^k))^2 + \right. \\ & \left. ((I^1)^k - G_{\theta_g}((I^1|l)^k))^2 \right] \end{aligned}$$

步骤8 利用梯度下降方法来更新修复网络。

重复步骤1~步骤8, 直到梯度收敛或达到预设的训练次数结束训练, 保存模型。

3 判别网络的判别性能分析

本文通过引入谱范数来归一化判别网络中的权重矩阵, 进一步约束其判别性能。假设判别网络为:

$$f(x, \theta) = W^L a_{L-1}(W^{L-1} a_{L-2}(\dots a_1(W^1 x))) \quad (10)$$

其中, $\theta := \{W^1, W^2, \dots, W^{L-1}, W^L\}$ 为判别网络的参数集(以全卷积为基础), 且 $W^l \in \mathbb{R}^{d_l \times d_{l-1}}$, $W^L \in \mathbb{R}^{1 \times d_L}$ 。 $a_l (1 \leq l < L)$ 是判别网络中第 l 层激活函数, x 是输入。判别网络的最终输出表示为 $D(x, \theta) = A(f(x, \theta))$, A 是该网络最后一层的激活函数。

GAN常用的归一化如WGAN-GP中的权重限制, 它通过使用 $L: h_{in} \rightarrow h_{out}$ 来渐进地约束每一层, 以此控制可微函数的Lipschitz上界。

根据定义, Lipschitz范数可以表示为:

$$\|L\|_{Lip} = \sup_h \sigma(\nabla L(h)) \quad (11)$$

其中, $\sigma(A)$ 是矩阵 A 的范数。对于线性层 $L = Wh$, 归一化过程可以表示为:

$$\|L\|_{Lip} = \sup_h \sigma(\nabla L(h)) = \sup_h \sigma(W) \quad (12)$$

根据Lipschitz不等式, 即:

$$\|L_1 \cdot L_2\|_{Lip} \leq \|L_1\|_{Lip} \cdot \|L_2\|_{Lip} \quad (13)$$

可以获得判别网络的上界 $\|f\|_{Lip}$ 为:

$$\begin{aligned} \|f\|_{Lip} & \leq \|a_{L-1}\|_{Lip} \cdot \|(h_L \rightarrow W^L h_L)\|_{Lip} \cdot \\ & \|a_{L-1}\|_{Lip} \cdot \|(h_{L-1} \rightarrow W^{L-1} h_{L-1})\|_{Lip} \cdot \dots \cdot \\ & \|a_1\|_{Lip} \cdot \|(h_1 \rightarrow W^1 h_1)\|_{Lip} = \\ & \prod_{l=1}^L \|a_l\|_{Lip} \cdot \|(h_l \rightarrow W^l h_l)\|_{Lip} \end{aligned} \quad (14)$$

当选择ReLU或者leaky ReLU作为激活函数时, 式(14)可以简化为 $\prod_{l=1}^L \|(h_l \rightarrow W^l h_l)\|_{Lip}$, 因为这两种激活函数拥有 $\|a_l\|_{Lip} = 1$ 的属性。由于谱范数被

定义为: $\sigma_s(A) = \max_{h: \|h\|_2=1} \|Ah\|_2 = \max_{\|h\|_2 \leq 1} \|Ah\|_2$, 并且等于矩阵 A 的最大奇异值, 因此式(12)可以进一步表示为:

$$\|L\|_{Lip} = \sup_h \sigma(\nabla L(h)) = \sup_h \sigma(W) = \sigma_s(W) \quad (15)$$

谱归一化定义为 $W_s(W) := \frac{W}{\sigma(W)}$, 因此, 式(14)

可以被简化为 $\prod_{l=1}^L \sigma(W^l)$ 。谱归一化使判别网络满足

Lipschitz约束 $\sigma(W_s(W)) = \sigma\left(\frac{W}{\sigma(W)}\right) = \frac{1}{K} \sigma(W) = 1$, 其中 K 为矩阵 W 的最大奇异值, 因此, $\|f\|_{Lip} = 1$ 。

当数据分布的重叠空间位于实际样本空间和生成的样本空间之间时, WGAN-GP^[27]中的权重剪切

效果很好,但该约束条件将判别网络的 Lipschitz 上界确定为常数 c ,限制了其应用范围。其他的归一化方法如基于 F 范数或 L 范数的归一化,减少了奇异值(平方)的总和,即对输入扰动不敏感的模型是有效的,但由于训练好的权重矩阵 W 都能对该方向上的梯度进行操作,因此训练好的模型可能会丢失关于输入的重要信息。相反,谱归一化仅关注最大的奇异值,并且每个训练好的权重矩阵 W 在与第一奇异向量正交的方向上并不显著地收缩。因此,谱归一化能够较好地控制判别网络的判别性能。

4 实验结果与分析

为验证所提算法的有效性和可行性,利用两个自然图像数据集来进行仿真实验。使用 Python 3.6 基础编译语言和 Chainer 编译框架,模型使用“四核 Intel® Core™ i-76850K CPU @3.6 GHz”处理器和 4 块“NVIDIA Titan V”的硬件环境。

4.1 数据集与参数设定

实验选择 StreetView 和 Places2 两个数据集: 1) StreetView 数据集是一个由 Google 建立的超过 20 个国家 5 000 个街道的景观数据集,该数据集图像像素变化明显,纹理结构清晰,图像分辨率为 128 像素×128 像素; 2) Places2 数据集包含超过 400 个场景种类,总计超过 100 000 张自然图像,该数据集含有 128 像素×128 像素的低分辨(Places2)和 256 像素×256 像素的高分辨图像(Places2-H),这些图像均符合人类视觉认知的准则。

参考 CE 和 GL 算法的训练方式,在本文实验过程中,将数据集以 7:2:1 的比例划分为训练集、测试集和验证集。训练过程随机选取数据集中每个类别的 70% 图像进行训练,训练完成后随机选择 10% 的图像进行交叉验证来确定参数。式(9)中的超参数 α 、 β 和 λ 用来平衡各个损失之间的比例,在实验中并不设置固定的值,而是将这 3 个超参数的取值范围设置为 $[10^{-6}, 10^2]$,在验证集上采用步长为 10 的交叉验证方法来自适应地确定具体的值。得到最优参数后,再利用剩余 20% 测试集中的图像进行测试。与 CE 算法训练固定中心缺失块不同,本文训练任意位置大小为 $[64, 128]$ 的随机块。训练次数为 12 万次,批次大小 $n=64$ 。采用 Adam 优化器进行优化,并设置衰减系数 $\beta_1=0.9$ 和 $\beta_2=0.99$ 。

4.2 对比算法与评价指标

本文算法属于基于生成对抗网络的图像修复算法,根据对修复图像的评估发现,基于深度学习的算法在

效果上优于基于扩散和基于纹理合成的方法,而在基于深度学习的方法中,基于生成对抗网络的修复算法性能较好,可扩展性较强和使用范围较广。因此,最终实验结果也与这类算法进行比较。近年来,研究人员提出了 4 种基于深度神经网络的修复模型:即 CE 模型^[15]、GL 模型^[16]、CA 模型^[17]和 Pconv 模型^[28]。前 3 种模型基于生成对抗网络,适用于修复规则缺失块,Pconv 基于 U-Net 网络结构,训练难度大,对硬件设备要求高,因此本文主要对比前 3 种模型。

评价指标主要包含主观和客观评估,前者用于评价修复图像的自然度,实验过程中邀请 3 位观察者独立地对修复图像的自然度进行判断,给每位观察者出示 100 张图像,观察者判断 100 张图像中原始图像和修复图像的数量。在客观评价方面,首先选择常用的图像评价方法 PSNR 和 SSIM,然后利用切片 Wasserstein 距离^[27](Sliced Wasserstein Distance, SWD)、Inception 分数^[29](Inception Score, IS)、流形距离^[30](Frechet Inception Distance, FID)、GAN-train 和 GAN-test^[31]对修复图像的流形结构进行评价。

4.3 修复结果主观评价

CE 模型^[15]主要修复中心缺失图像,缺失部分的大小是固定的。首先对比本文算法和 CE 算法在中心缺失受损图像的修复结果,如图 2 所示。其中,第 1 行图像选自 StreetView 数据集,第 2 行选自 Places2 数据集,第 3 行选自 Places2-H 数据集。

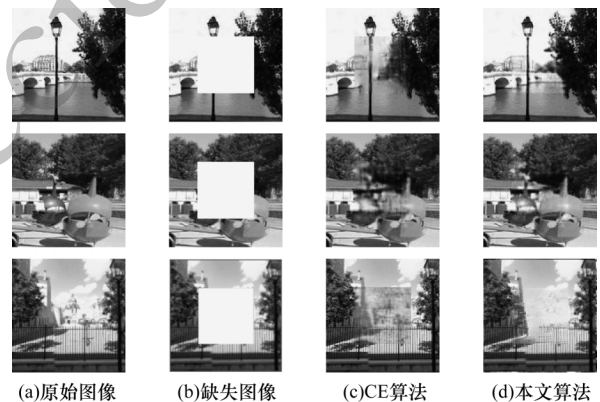


图 2 中心缺失图像的修复结果

Fig.2 Inpainting result of center missing images

由图 2 可以看出,从视觉感受角度出发,本文算法修复中心缺失图像效果明显优于 CE 算法。随着图像大小增加,CE 算法的修复结果有所下降,而所提算法修复中心缺失受损图像获得较好的修复效果。为扩展实验范围,将 GL 算法和 CA 算法在随机缺失图像上进行实验对比,结果如图 3 所示。

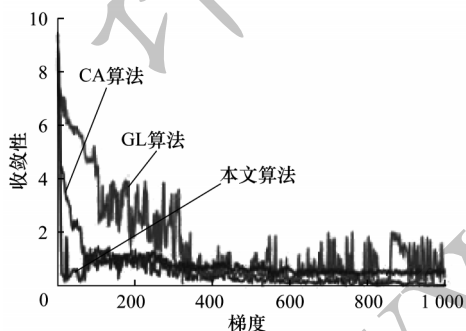


(a)原始图像 (b)缺失图像 (c)GL算法 (d)CA算法 (e)本文算法

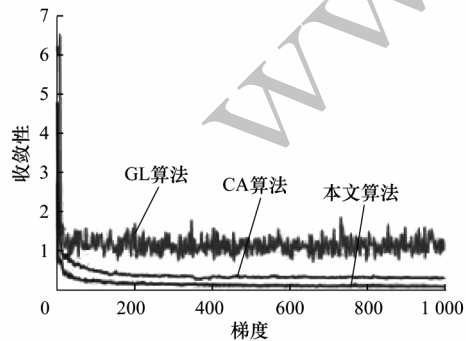
图 3 随机缺失图像的修复结果

Fig.3 Inpainted results of randomly missing images

从图 3 可以看出,本文算法与 CA 算法修复受损图像效果都较好,单从视觉感受角度出发,已无法区分其差别。本文引入谱归一化来约束判别网络性能,为进一步评价本文算法和对比算法的性能,对梯度收敛情况进行分析,结果如图 4 所示。由于 GL 算法是基于 CE 算法的,并且它在训练稳定性和梯度收敛性均优于 CE 算法,因此这部分实验只对比 GL 算法和 CA 算法。



(a)全局判别器梯度收敛



(b)生成器梯度收敛

图 4 梯度收敛性分析

Fig.4 Gradient convergence analysis

从图 4 可以看出,本文算法的梯度收敛性明显好于 GL 算法和 CA 算法。对比图 4(a),本文算法在训练过程中判别网络的梯度行为较好,梯度收敛后无明显抖动。对比图 4(b),本文算法修复网络的梯度行为较好,一直处于下降状态,并且梯度没有消失。

在全局判别网络中,当梯度趋近于某个稳定的值后,本文算法没有出现梯度抖动现象,原因在于引入了谱归一化来约束判别网络参数的 Lipschitz 上界。在局部判别网络中也能得到类似的结果,实验结果证明了谱归一化在训练判别网络过程中起着积极的作用。

4.4 量化分析

图像客观评价指标较多,首先计算大量修复图像的平均 PSNR 和 SSIM 值来对修复算法进行客观分析。为客观地比较增减幅度,本文定义增减量化准则,以 PSNR 值增幅为例可表示为:

$$PSNR(\text{Increment}) = \frac{PSNR_2 - PSNR_1}{PSNR_1} \quad (16)$$

其中, $PSNR_1$ 表示对比算法修复图像的 PSNR 值, $PSNR_2$ 表示所提算法修复图像的 PSNR 值。在 3 个数据集上进行测试的平均 PSNR 值和 SSIM 评价value 如表 1、表 2 所示。

表 1 不同算法在 3 个数据集上的 PSNR 值

Table1 PSNR values of different algorithms on three datasets

数据集	PSNR/dB				平均增量 /%
	CE 算法	GL 算法	CA 算法	本文算法	
StreetView	34.96	39.64	40.44	40.64	8.1
Places2	33.29	36.59	37.02	37.59	6.4
Places2-H	34.00	35.43	37.01	37.07	4.6

表 2 不同算法在 3 个数据集上的 SSIM 值

Table2 SSIM values of different algorithms on three datasets

数据集	SSIM 值				平均增量 /%
	CE 算法	GL 算法	CA 算法	本文算法	
StreetView	0.85	0.89	0.92	0.93	4.7
Places2	0.88	0.91	0.96	0.96	4.5
Places2-H	0.86	0.90	0.94	0.95	5.8

从表 1 可以看出,在所有的数据集中,破损图像经过 4 种算法修复后,本文算法与 CA 算法修复的图像 PSNR 值最高。对比 CE 算法,PSNR 值涨幅为 9.0%~16.2%,对比 GL 算法,PSNR 值涨幅为 6.0%~9.1%。对比 CA 算法,PSNR 涨幅为 0.1%~2.3%。为评估修复图像纹理结构,利用 SSIM 来对修复图像进行评价,结果如表 2 所示。从表 2 可以看出,本文算

方法与CA算法修复的图像SSIM值较高。对比CE算法,SSIM值涨幅为9.40%~10.5%,对比GL算法,SSIM值涨幅为6.04%~9.3%,对比CA算法,SSIM值涨幅为0.0%~1.0%。除了利用PSNR与SSIM评价指标,从大数据量和流形结构评估的角度出发,引入SWD、IS、FID、GAN-train和GAN-test对修复网络的修复性能进行评价。其中,GAN-train和GAN-test是在VGG19分类网络上进行训练获得,结果如表3~表5所示。

表3 不同算法与StreetView数据集的流形结构评价价值

Table 3 Manifold structure evaluation value of different algorithms and StreetView dataset

数据集与算法	SWD	FID	IS	GAN-train/%	GAN-test/%
StreetView	13.1	9.9	9.6	94.1	—
CE[15]算法	19.3	15.4	4.8	88.3	86.5
GL[16]算法	17.6	13.7	6.7	89.7	87.3
CA[17]算法	18.4	14.1	5.9	89.2	86.9
本文算法	16.8	13.2	7.1	90.4	89.4
平均增量/%	9.8	10.2	26.9	1.6	2.9

表4 不同算法与Places2数据集的流形结构评价价值

Table 4 Manifold structure evaluation value of different algorithms and Places2 dataset

数据集与算法	SWD	FID	IS	GAN-train/%	GAN-test/%
Places2	19.5	12.9	8.8	90.3	—
CE[15]算法	26.9	19.8	4.1	86.7	83.2
GL[16]算法	24.6	18.0	5.8	87.9	84.3
CA[17]算法	25.9	19.0	6.1	87.0	83.9
本文算法	23.9	17.1	6.3	88.1	84.6
平均增量/%	7.7	10.5	31.1	0.9	1.0

表5 不同算法与Places2-H数据集的流形结构评价价值

Table 5 Manifold structure evaluation value of different algorithms and Places2-H dataset

数据集与算法	SWD	FID	IS	GAN-train/%	GAN-test/%
Places2-H	18.3	10.0	7.7	91.2	—
CE[15]算法	25.3	17.6	4.9	87.7	83.9
GL[16]算法	24.3	15.7	6.2	89.0	85.6
CA[17]算法	24.8	14.6	5.9	88.6	84.3
本文算法	23.8	14.6	6.9	89.9	86.0
平均增量/%	4.2	6.1	26.1	1.7	6.1

SWD和FID等评价指标主要用来评价数据集之间的流形结构相似性,SWD、FID评价价值越小,表示数据集之间的流形结构越相似,而IS、GAN-train和GAN-test评价价值越大,表示数据集之间的流形结构越相似。从表3~表5可以看出,对比3种算法,本文算法获得的流形结构相似度评价最高,说明该算法较好地保持了修复图像数据集与原始图像数据集的

流形结构相似性。

综合上述客观评价指标,可以得出本文算法在修复效果、梯度收敛、判别网络性能控制以及保持原始数据集与修复数据集之间的流形结构方面,均取得较好的结果。

为综合评价修复图像的质量,邀请观察者对修复图像的自然度进行主观评价,结果如图5所示。

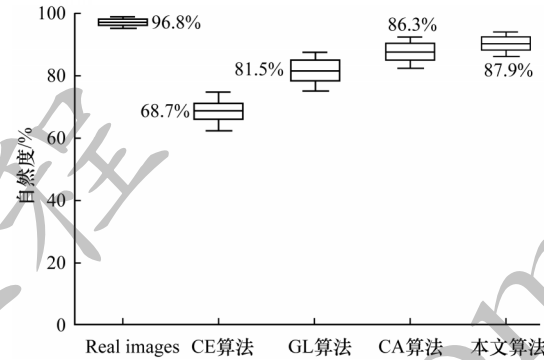


图5 修复图像自然度评估结果

Fig.5 Naturalness evaluation result of inpainted image

从图5可以看出,本文算法修复受损图像能获得较高的图像自然度,相比CE算法、GL算法和CA算法分别提高了19.2%、6.4%和1.6%。

4.5 算法分析

与CE、GL和CA算法在主、客观评价方面的综合对比结果显示,本文算法在流形结构保持、判别网络性能控制方面具有优越性。此外,对本文算法的变体模型进行研究,明确该算法中损失组成部分的作用。变体算法包括:未使用pMSE损失的变体算法SN-CGAN-P,未使用TV损失的变体算法SN-CGAN-T,未使用视觉感知损失的变体算法SN-CGAN-V,未使用标签信息的变体算法SN-CGAN-L及未使用谱归一化的变体算法SN-CGAN-S。在StreetView数据集上流形结构相似度、PSNR和SSIM值的评价结果如表6、表7所示。

表6 变体算法流形结构保持评价结果

Table 6 Manifold structure preservation evaluation results of variant algorithm

变体算法	SWD	FID	IS	GAN-train/%	GAN-test/%
SN-CGAN-P算法	18.1	15.9	5.6	86.8	86.0
SN-CGAN-T算法	17.2	15.3	5.8	89.7	87.9
SN-CGAN-V算法	18.6	16.8	5.0	87.2	86.6
SN-CGAN-L算法	18.0	16.3	5.4	86.8	85.6
SN-CGAN-S算法	19.7	17.6	8.3	84.3	80.7
SN-CGAN算法	16.8	13.2	7.1	90.4	89.4
平均增量/%	9.0	24.1	21.6	4.0	4.8

表7 变体算法的PSNR和SSIM评价结果

Table 7 PSNR and SSIM evaluation results of variant algorithms

变体算法	PSNR/dB	SSIM 值
SN-CGAN-P	34.88	0.91
SN-CGAN-T	36.73	0.90
SN-CGAN-V	38.87	0.91
SN-CGAN-L	39.88	0.93
SN-CGAN-S	33.10	0.89
SN-CGAN	40.64	0.93
平均增量/%	11.20	2.40

视觉感知损失用来获取良好的视觉感知、提高主观评价和解决过度平滑问题。因此,表6中该项损失对流形结构保持有积极的作用。 p MSE损失用来减少修复图像与原始图像之间的失真,TV损失能抑制图像修复过程中产生的噪声。从表7可以看出,未引入 p MSE损失和TV损失时,PSNR和SSIM值均有所下降。

条件生成对抗网络^[25]是通过引入数据标签来提高合成图像的质量,而本文则引入标签来约束图像特征生成,从而保证修复内容与原始图像之间的内容保持。此外,引入谱归一化约束判别网络的判别性能,进一步提高了修复网络的修复能力。根据对判别网络中参数的Lipschitz上界理论分析,并结合图5和表6、表7可以看出,谱归一化对控制判别网络的判别性能、间接提高修复网络的修复能力以及保持修复图像数据集的流形结构有重要作用。

5 结束语

基于变分偏微分和纹理合成等传统图像修复算法无法产生大批量修复图像。为此,本文提出一种基于谱归一化条件生成对抗网络的图像修复算法,基于生成对抗网络修复算法解决判别网络性能不可控以及缺失部分较多、失真较大时修复结果缺乏局部一致性问题。实验结果表明,该算法在主观感受和客观评价上优于SN-CGAN-P、SN-CGAN-T等对比算法,较好地实现了破损图像与原始图像之间的非线性映射。下一步将该学习框架引入到盲复原领域中,学习模糊图像与真实图像之间的映射关系,以实现运动模糊或散焦模糊图像的复原。

参考文献

- [1] ZHANG Yujin. Image engineering in China; 2018 [J]. Journal of Image and Graphics, 2018, 24(5): 665-676. (in Chinese)
章毓晋. 中国图像工程; 2018 [J]. 中国图象图形学报, 2018, 24(5): 665-676.
- [2] SHEN T F, CHAN J H. Mathematical models for local non-texture inpaintings [J]. SIAM Journal on Applied Mathematics, 2002, 62(3): 1019-1043.
- [3] AFONSO M, BIOUCASDIAS J, FIGUEIREDO M. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems [J]. IEEE Transactions on Image Processing, 2011, 20(3): 681-695.
- [4] ZHAO Zhengkang, LIU Ningzhong, LI Wei. Image inpainting algorithm based on synthetic sample block by gray entropy [J]. Computer Engineering, 2016, 42(11): 267-271. (in Chinese)
赵政康, 刘宁钟, 李伟. 基于灰度熵合成样本块的图像修复算法 [J]. 计算机工程, 2016, 42(11): 267-271.
- [5] BATOOLE N, CHELLAPPA R. Detection and inpainting of facial wrinkles using texture orientation fields and markov random field modeling [J]. IEEE Transactions on Image Processing, 2014, 34(9): 3773-3788.
- [6] GUO Qian, GAO Shanshan, ZHANG Xiaofeng, et al. Patch-based image inpainting via two-stage low rank approximation [J]. IEEE Transactions on Vision Computer Graphics, 2018, 24(6): 2023-2036.
- [7] JIANG Fan, LIU Hui, WANG Bin, et al. Image recognition based on CNN-GRNN model [J]. Computer Engineering, 2017, 43(4): 257-262. (in Chinese)
江帆, 刘辉, 王彬, 等. 基于CNN-GRNN模型的图像识别 [J]. 计算机工程, 2017, 43(4): 257-262.
- [8] SOHEI D, ELI S, CONNELLY B. Image melding: combining inconsistent images using patch-based synthesis [J]. ACM Transactions on Graphics, 2012, 31(4): 82-91.
- [9] ZHU Xuan, ZHANG Xufeng, LI Qiuju, et al. Image inpainting method based on sparse decomposition [J]. Computer Science, 2016, 43(1): 294-297. (in Chinese)
祝轩, 张旭峰, 李秋菊, 等. 基于稀疏分解的图像修复方法 [J]. 计算机工程, 2016, 43(1): 294-297.
- [10] QIANG Zhenping, HE Libo, CHEN Xu, et al. Survey on deep learning image inpainting methods [J]. Journal of Image and Graphics, 2019, 24(3): 447-463. (in Chinese)
强振平, 何丽波, 陈旭, 等. 深度学习图像修复方法综述 [J]. 中国图象图形学报, 2019, 24(3): 447-463.
- [11] SHI En, LI Qian, GU Daquan, et al. Convolutional neural network model based on local feature [J]. Computer Engineering, 2018, 44(2): 282-286. (in Chinese)
施恩, 李骞, 顾大权, 等. 基于局部特征的卷积神经网络模型 [J]. 计算机工程, 2018, 44(2): 282-286.
- [12] XIE Junyuan, XU Linli, CHEN Enhong. Image denoising and inpainting with deep neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2012: 341-349.
- [13] YANG Chao, LU Xin, LIN Zhe, et al. High-resolution image inpainting using multi-scale neural patch synthesis [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 4076-4084.
- [14] GOODFELLOW I, ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2014: 2672-2680.
- [15] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2536-2544.

- [16] SATOSHI I, EDGAR S, HIROSHI I. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-14.
- [17] YU J, LIM Z, YANG J, et al. Generative image inpainting with contextual attention[C]//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C. , USA: IEEE Press, 2018: 5505-5514.
- [18] YOSHIDA Y, MIVATO T. Spectral norm regularization for improving the generalizability of deep learning [C]//*Proceedings of the 31th International Conference on Neural Information Processing Systems*. Vancouver, Canada: MIT Press, 2018: 1539-1542.
- [19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//*Proceedings of the 28th International Conference on Machine Learning*. Lille, France: ACM Press, 2015: 1-11.
- [20] REN J, XU L, YAN Q, et al. Shepard convolutional neural networks [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*. Cambridge, USA: MIT Press, 2015: 901-909.
- [21] LEDIG C, WANG Z H, SHI W Z, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C. , USA: IEEE Press, 2017: 105-114.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//*Proceedings of the 31th International Conference on Learning Representations*. San Diego, USA: ACM Press, 2015: 1-14.
- [23] ARIJOVASKY M, BOTTOU L. Towards principled methods for training generative adversarial networks[C]//*Proceedings of the 6th International Conference on Learning Representation*. Vancouver, Canada: [s. n.], 2018: 257-273.
- [24] ARIJOVASKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[C]//*Proceedings of the 30th International Conference on Machine Learning*. Sydney, Australia: [s. n.], 2017: 214-223.
- [25] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C]//*Proceeding of the 34th International Conference on Machine Learning*. Sydney, Australia: [s. n.], 2017: 237-251.
- [26] KINGMA D, BA J. Adam: a method for stochastic optimization [C]//*Proceedings of the 19th International Conference on Computational Statistics*. Paris, France: [s. n.], 2015: 1-15.
- [27] GULRAJANI I, AHMED F, ARIJOVASKY M, et al. Improved training of Wasserstein GANs[C]//*Proceeding of the 31th International Conference on Neural Information Processing Systems*. Long Beach, USA: MIT Press, 2017: 294-303.
- [28] LIU G, RWDA F A, SHIH K, et al. Image inpainting for irregular holes using partial convolutions[C]//*Proceeding of IEEE European Conference on Computer Vision*. [S. l.], IEEE Press, 2018: 89-105.
- [29] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a Nash equilibrium[C]//*Proceeding of the 30th International Conference on Neural Information Processing Systems*. Long Beach, USA: MIT Press, 2017: 2672-2680.
- [30] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs[EB/OL]. [2019-09-10]. <https://arxiv.org/abs/1606.03498>.
- [31] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[C]//*Proceedings of the 5th International Conference on Learning Representations*. Toulon, France: [s. n.], 2017: 1-26.
- [32] SHMELKOV K, SCHMID C, ALAHARI K. How good is my GAN?[C]//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C. , USA: IEEE Press, 2018: 3654-3668.