



文章编号:1001-9081(2018)12-3557-06

DOI:10.11772/j.issn.1001-9081.2018051097

基于双鉴别网络的生成对抗网络图像修复方法

刘波宁*, 翟东海

(西南交通大学 信息科学与技术学院, 成都 610031)

(*通信作者电子邮箱 boningliu@qq.com)

摘要:针对现有神经网络图像修复方法的修复结果在视觉连通性上存在结构扭曲、训练过程中易陷入过度学习等问题,提出了一种基于双鉴别网络的生成对抗网络(GAN)图像修复方法。该方法的修复模型使用了修复网络、全局鉴别网络和局部鉴别网络。修复网络将待修复图像破缺区域用相似信息填充后作为输入,极大地提高了生成图像的速度与质量;全局鉴别网络综合采用图像全局的边缘结构信息和特征信息以保证修复网络输出的修复图像结果符合视觉连通性;而局部鉴别网络在鉴别输出图像的同时,利用在多个图像中寻找到的辅助特征块来提高鉴别的泛化能力,很好地抑制了修复网络在特征过于集中或单一时容易过度学习的问题。实验结果表明,所提修复方法在人脸类图像上具有较好的修复效果,且在不同种类图像上有非常好的适用性,其峰值信噪比(PSNR)和结构相似性(SSIM)指标比当前基于深度学习且修复效果较好的几种方法更优。

关键词:生成对抗网络;边缘结构;人脸修复;缓冲池;卷积神经网络

中图分类号: TP391.41 **文献标志码:**A

Image completion method of generative adversarial networks based on two discrimination networks

LIU Boning*, ZHAI Donghai

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu Sichuan 610031, China)

Abstract: The existing image completion methods have the problems of structural distortion on visual connectivity and easy to overfitting in the process of training. In order to solve the problems, a new image completion method of Generative Adversarial Network (GAN) based on two discrimination networks was proposed. One completion network, one global discrimination network and one local discrimination network were used in the completion model of the proposed method. The broken area of image to be completed was filled by a similar patch as input in the completion network, which greatly improved the speed and quality of the generation images. The global marginal structure information and feature information were used comprehensively in the global discrimination network to ensure that the completed image of completion network conformed visual connectivity. While discriminating the output image, the assisted feature patches found from multiple images were used to improve the generalization ability of discrimination in the local discrimination network, which solved the issue that the completion network was easily overfitting with too concentrated features or single feature. The experimental results show that, the proposed completion method has good completion effect on face images, and has good applicability in different kinds of images. The Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) of the proposed method are better than those of the state-of-the-art methods based on deep learning.

Key words: Generative Adversarial Network (GAN); marginal structure; face completion; cache pool; Convolution Neural Network (CNN)

0 引言

在传统的图像修复方法中,修复方法一般可以分为两种:第一种是适用于小尺寸的图像修复方法,如快速行进算法^[1]、Criminis 算法^[2-3]等。第二种方法是适用于较大面积的图像填充修复,如 PatchMatch 方法^[4]和统计块概率修复方法^[5]。总体来说,传统图像修复方法基于原图,由于在内容上缺乏高层次的连贯性,使得修复结果不令人满意;而且一旦图像的缺失比例过大,导致重要的参考信息丢失,也同样无法让人取得满意的效果。

利用特征学习的深度学习修复方法^[6-12]恰好填补了传统图像修复方法内容缺乏高层次连贯和难于应对大块缺失区域的缺陷。随着图形处理器(Graphics Processing Unit, GPU)技术的发展,深度学习技术在图像分类^[13]、图像检测识别^[14-15]、图像目标定位^[16-18]、图像生成^[19-20]等方面的应用逐渐广泛。

在深度学习的图像修复中,图像生成是图像修复基础,本文采用的生成对抗网络(Generative Adversarial Network, GAN)^[10]模型中的生成网络也来源于图像生成的研究^[21]。在图像生成中,Vincent 等^[6]引入去噪的自动编码器,重建并

收稿日期:2018-05-28;修回日期:2018-07-12;录用日期:2018-07-17。 基金项目:国家自然科学基金资助项目(61461048, 61661047)。

作者简介:刘波宁(1993—),男,广东梅州人,硕士研究生,主要研究方向:数字图像处理; 翟东海(1974—),男,山西芮城人,副教授,博士,主要研究方向:数字图像处理、海量数据挖掘。



学习干净的输入信号来生成图像。随后的研究中, Dosovitskiy 等^[7]通过解码器网络反向表达深度卷积网络的特征(例如 VGG(Visual Geometry Group)^[13] 网络)来重建对象图像, 这使得生成的图像在真实性上有了重要的突破。随后的研究中, 大多都着力于图像的真实性上, 比如, Kingma 等^[8]提出的变分自动编码器(Variational Auto-Encoder, VAE)通过在隐含层之前加入规则化的编码器, 使得可以通过从隐含层采样或插入隐含层来生成更加真实的图像。然而, 由 VAE 生成的图像因为其基于像素的高斯似然性训练, 输出的部分图像是模糊的。于是, 在进一步的研究中, Larsen 等^[9]通过增加一个对抗训练来改善 VAE, 这就可以产生更真实的图像。

在这些图像生成的研究基础之上, 逐渐提出了一些比较典型的图像修复方法。其中较为出众并广泛使用的一类是基于生成对抗网络^[10]的图像修复方法。比如 Li 等^[11]提出的人脸修复方法在人脸修复上取得了较好的效果, 这种方法在生成网络修复人脸后, 利用了人脸目标检测技术, 检测生成图像的人脸具体器官, 比如鼻子、眼睛等是否符合常理位置, 然后改进网络生成的人脸, 进而起到较好的修复效果。而 Yeh 等^[12]提出的语义图像修复方法, 则利用了潜在的相似图像, 生成网络通过对这些相似图像的特征不断来推断缺失图像的缺失区域, 最后经过迭代优化形成修复图像。

本文所提出的算法是基于 Iizuka 等^[22]提出的全局和局部网络的图像修复方法。后者是一种经典的修复方法, 这种算法利用了生成对抗网络, 在传统的生成对抗网络中, 生成网络负责生成图像, 鉴别网络负责鉴别生成网络生成图像的效果优劣, 而此法则额外增加了一个全局鉴别网络用来负责鉴别破损区域修复后的全局图像是否符合视觉连贯性, 最后综合全局和局部的两种鉴别结果来判断图像的修复效果。整个修复流程为先通过训练生成网络来完成初步的图像修复, 然后不断通过负责鉴别全局图像视觉连贯性的全局鉴别网络和负责鉴别生成网络输出的修复图像真实性的局部鉴别网络与生成网络对抗训练, 在迭代训练中一步一步产生较好修复效果。

但是, 上述算法包括在本文实验中使用的特征卷积算法^[23]都存在一定的问题, 比如: 在 Li 等^[11]提出的人脸修复方法中, 需要定位一些小部件是否符合人的主观, 比如鼻子和眼睛是否真实和清晰, 生成的位置是否正确, 这涉及到一些目标检测识别技术, 而这给算法的复杂度和训练时间开销带来更高的挑战。语义图像修复方法^[12]中利用图像语义的判断技术, 对生成图像的生成效果进行判定, 这同样涉及到了比较复杂的图像语义判别。以上两种方法都对不同的图像数据缺乏适应性。而在 Iizuka 等^[22]提出的图像修复方法中, 虽然模型较为简单, 但这导致在海量的图像训练中, 容易出现特征过度学习的问题。

因此, 为了降低修复模型的复杂性, 增强普适性, 改善特征过度学习的问题, 使得模型图像修复的质量和效果更加出众, 本文提出了一种基于边缘结构信息检测和局部信息增强的生成对抗网络的图像修复方法, 该方法主要有以下工作:

1) 寻找相似块填充破损区域, 通过保留破损区域的图像结构与纹理信息, 来提高修复网络生成图像的质量和网络的训练速度。

2) 在全局鉴别网络中, 限定图像输入区域排除无用特征信息干扰, 利用检测边缘结构信息来约束修复网络生成的图

像, 使其生成具有视觉连通性的修复图像。

3) 局部鉴别网络增加缓冲池技术, 通过在缓冲池中增加更多的特征辅助块, 来避免修复网络掉入过度学习的陷阱。

1 生成对抗网络

生成对抗网络(GAN)^[10]的核心思想是“零和博弈”^[24], GAN 模型如图 1 所示, 目标函数如式(1):

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

GAN 模型主要有两部分组成: 第一部分是生成图像的生成网络 Generator, 通过输入一组无序随机的噪声 z 来映射为一张图像; 第二部分是鉴别图像真伪的鉴别网络 Discriminator, 通过不断学习真实图像 x 特征来加强对某一类图像的鉴别能力。GAN 的目标函数很好地体现了对抗的思想, 即是最小化生成网络生成图像与真实图像的差距, 最大化鉴别网络的鉴别真实图像和生成图像能力, 即最大化生成图像与真实图像的差距。

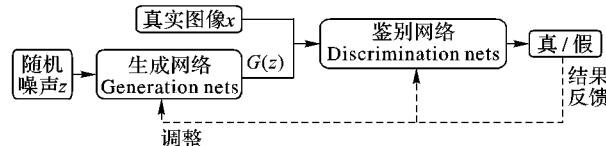


图 1 经典的 GAN 模型

Fig. 1 Classic GAN model

在生成对抗网络训练中, 把随机噪声 z 输入生成网络, 生成网络卷积映射生成一张图像, 即 $G(z)$, 然后把生成图像 $G(z)$ 及其标签“0”和真实图像 x 及其标签“1”一起送入鉴别网络 D 中, 训练鉴别网络 D 的鉴别能力, 使得鉴别网络 D 输出的概率 $D(G(z))$ 趋于 0 而 $D(x)$ 趋于 1。这样, 生成对抗网络就能根据鉴别网络输出的概率数值区分输入图像是来自真实图像还是生成图像。

同时, 这个概率数值 $D(G(z))$ 将反馈给生成网络, 生成网络将根据概率数值结果来调整自己的生成图像, 在下一次训练的时候, 生成更真实的图像欺骗鉴别网络, 使得 $D(G(z))$ 这一概率数值趋于 1。随着训练迭代进行, 生成网络与鉴别网络之间形成一种对抗关系最后达到微妙的平衡, 此时 $D(G(z))$ 趋于 0.5。这时候, 在给定生成网络 G 的情况下, 目标函数式(1)在式(2)得到鉴别网络 D 的最优解:

$$D_G(x) = p_{\text{data}}(x)/(p_{\text{data}}(x) + p_g(x)) \quad (2)$$

在这种情况下, 鉴别网络无法判断生成网络生成的图像是真实图像还是生成图像, 因为生成网络生成了极为逼近真实图像的图像。生成对抗网络通过大量的特征学习, 能够很好地学习到同类图像数据集中的图像特征, 从而生成逼真的图像。

2 双鉴别网络修复算法

本文运用生成对抗网络来修复图像, 修复模型分为两部分, 如图 2 所示, 分别是负责修复图像的修复网络和负责鉴别图像修复效果并反馈给修复网络的鉴别网络。与传统的生成网络不同的是, 本文采用的修复网络, 不再是从一组噪声作为输入并通过不断地卷积等手段生成一张图像, 而是直接将一张图像作为输入, 通过对称的卷积层和反卷积层, 将一张破损的图像修复为一张符合视觉的图像。



在经典的生成对抗网络中,用 $D(G(z))$ 来衡量生成网络生成图像的能力。而在本文中,为了使得修复图像符合视觉连贯性,引入衡量更为多样化的图像损失 loss 的概念,取代了在生成对抗网络中, $D(G(z))$ 对生成图像的简单判断。通过最小化图像损失 loss,使得修复网络尽可能生成接近人眼视觉连贯性的图像,取得最好的修复效果。修复图像的生成对抗网络目标函数如式(3) :

$$\min_c \max_d V(D, C) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - Loss)] \quad (3)$$

其中: C 代表 Completion Network,也就是修复网络,负责生成修复的图像; D 代表 Discrimination Network,包含两个鉴别网络,分别是局部鉴别网络和全局鉴别网络,负责鉴别修复图像的修复效果; x 是完好的图像,从完好的训练数据集 P 中抽取出来; z 是等待修复网络 C 修复的破损图像。式(3) 中定义了一个全新的图像损失函数 Loss,来衡量修复网络的修复效果,计算式如下:

$$Loss = \lambda_1 Loss_{\text{global}} + \lambda_2 Loss_{\text{local}} \quad (4)$$

其中: λ_1 和 λ_2 是两个权重系数,用来平衡全局鉴别网络和局部鉴别网络之间的权重占比;而 $Loss_{\text{global}}$ 代表全局鉴别网络的损失, $Loss_{\text{local}}$ 代表局部鉴别网络所输出的损失。详细损失构成在 2.2 节中论述。

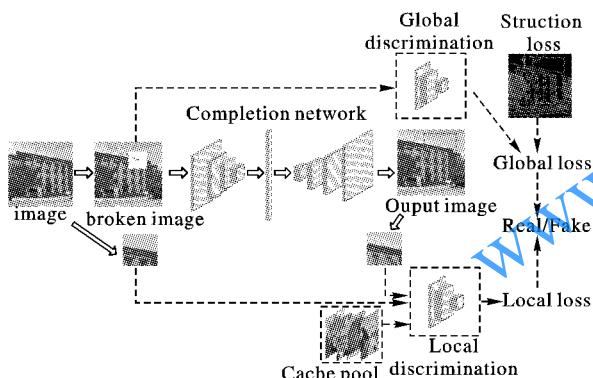


图 2 本文提出的 GAN 模型

Fig. 2 Proposed GAN model

2.1 修复网络相似块搜寻

本文的修复网络采用了破损图像作为输入,但是为了最大化遏止修复网络在生成修复图像时,由于拟合无序信息而产生损失,在待修复的破损区域,还采用了相似填充的方法。

在文献[10]的研究中,直接将一组随机的噪声作为输入,在生成网络的拟合下,这组噪声慢慢映射生成一张逼真的图像。后来在文献[11]、文献[12]、文献[21]的图像生成方法中,都用图像替换了噪声输入,其中,文献[11]方法和文献[12]方法使用的是破损图像,文献[21]方法使用的是人工手绘图像,这些图像的使用提升了生成网络生成图像的效率和质量。

破损图像替换了噪声输入后能提升生成图像的效率和质量的根本原因是噪声是一组无序的随机数,它没有保留任何有助于生成网络生成图像的信息,生成网络为了拟合这组无序信息,建立无序噪声与图像之间的映射,会产生许多不必要的损失。文献[11]方法和文献[12]方法虽然采用了破损图像作为输入,在一定程度上减少了映射的损失,但是并未对最为重要的破损区域加以处理,同样还是对生成网络造成了很大的损失。

本文采用相似填充方法将训练用的破损图像进行预处理,通过把破损区域周围的完好图像信息传播到破损区域内,以此保证待修复图像内没有像噪声一样的无用干扰信息,从而降低修复网络为了拟合输入图像产生的损失。该方法通过比对周围邻域块的相似程度,从破损区域的周围找出一块合适的填充块,填充到破损的区域里面。填充的相似块能够提供给修复网络一些结构和纹理信息,减少特征映射时的损失,从而提高修复网络生成图像的质量和速度。

为了尽量保留与完好块相似的结构与纹理信息,在寻找相似块的时候,本文参考了 PatchMatch 方法^[4] 来寻找一个最好的相似块。同时,考虑到神经网络的耗时性,为了提高相似块的搜索速度,本文简化了 PatchMatch 方法,如图 3 所示,具体操作如下:

1) 简化偏置信息。随机选择破损区域相邻的一个完好块,并在图像内给完好块随机分配一个偏置块,记录这个完好块与破损块之间的位置信息。

2) 传播。用偏置信息和位置信息更新破损块。

3) 随机搜索。以随机偏置块为基准,以 0.5 的收敛速度随机搜索更好的偏置块,更新破损块。

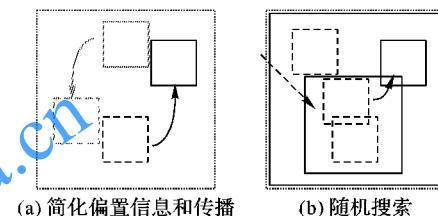


Fig. 3 PatchMatch simplified method

为了在寻找的效率和质量上有更好的表现,本文进一步对随机的迭代作出了一定的限制,设定随机搜索的次数为 2 次。

2.2 双鉴别网络

针对修复区域容易与边缘结构纹理信息不一致的问题,Iizuka 等^[22] 提出的图像修复方法中增加了负责鉴别全局图像区域的全局鉴别网络。在这种方法中,修复的图像既要在局部上产生合适的图像,又需要在全局上拥有较好的视觉连通性。因此,为了维持这种全局的一致性,采用了破损周边权重衡量的方法。但是,这种权重方法在对生成的修复图像约束的同时,过于突出周围某些权重信息,从而对待修复区域的修复产生了负面影响;同时,在采用了全局图像作为输入时,也为全局鉴别网络引入了过多的特征,从而形成了信息干扰。

本文同样也采用了全局鉴别与局部鉴别的双网络结构,即是由局部鉴别网络来负责破损区域的图像生成结果鉴别,由全局鉴别网络来负责整体视觉连通性的修复结果鉴别。但是针对上述两个问题,本文采用了基于结构边缘信息约束的方法来解决权重失衡的问题,同时,限制了全局鉴别网络的图像输入区域,以便排除因为过多特征产生的干扰,最终使得全局鉴别网络能够更加关注于全局图像的视觉连通性;而局部鉴别网络,能更加关注于当前破损区域的生成图像效果。

2.2.1 全局鉴别网络

在本文中,全局鉴别网络的输入区域(图 4 虚线外框)并非是整张图像而是限定在以破损区域(图 4 虚线内框)中心为中心、面积为破损区域两倍的区域,使得这两者的面积符合:



$$2 \times Area_{\text{broken}} = Area_{\text{global}} \quad (5)$$

之所以采用这样的限定,是因为在图像修复的过程中,本文主要关注以下两点:第一,修复网络生成的修复图像要符合视觉连贯性,即修复后的破损区域与周围完好区域的纹理和结构一致;第二,避免引入过多的无关特征信息干扰,这些信息容易导致修复网络学习到无关特征,影响修复效果。采用限定输入图像区域大小的方法,不影响全局鉴别网络对破损图像待修复区域的图像鉴定;同时,因为限定区域以外的图像没有输入全局鉴别网络,也就避免了更多的特征信息干扰。

在限制图像中无关特征信息干扰的同时,对全局鉴别网络增加了一个结构边缘信息约束。这是由于修复网络在无约束时过分追求破损区域与完好区域的平滑连接,生成了过平滑的图像。观察这些过平滑的图像,能观察到一些明显的纹理信息,但是符合视觉连贯性图像注重的结构信息却大部分丢失,如图4(c)所示。所以,本文需要进一步在修复网络中突出结构信息,让结构信息在修复网络生成的图像中得到非常好的再现。

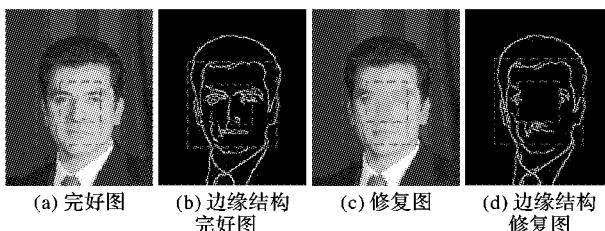


图4 全局鉴别网络不同图像对比

Fig. 4 Different image comparison in global discrimination network

本文将完好图像和修复图像先进行边缘结构检测,然后,通过对比两者之间的结构差异,进而以此来惩罚那些在结构上与完好图像不一致的修复图像,从而提高修复网络对修复图像结构信息的重视程度。

鉴于全局鉴别网络同时考虑了结构和纹理信息,本文对全局损失函数改进如下:

$$Loss_{\text{global}} = loss(D(Area_{\text{global}})) + loss_{\text{structure}} \quad (6)$$

式中 $loss_{\text{structure}}$ 为结构损失惩罚项,用来惩罚修复图像和完好图像像素点间在结构上的差异:

$$loss_{\text{structure}} = \frac{1}{N} \sum_{i=1,2,\dots,N} \| p_i - q_i \| \quad (7)$$

其中: N 是 $Area_{\text{global}}$ 区域(图4虚线外框)内的全部像素点数; p_i 表示完好图像的第 i 个像素点,同时 q_i 表示修复图像中与之对应的像素点。

2.2.2 局部鉴别网络

针对局部鉴别网络在海量图像训练中容易产生特征漂移^[23],即特征过度学习的问题,本文采用缓冲池存储干扰块,避免局部鉴别网络掉入过度学习的陷阱。

在文献[10]的原始局部鉴别网络中,把修复后的图像以及与之对应的完好图像作为局部鉴别网络的输入进行训练,局部鉴别网络根据完好图像的特征与修复后图像的特征的差异构造局部鉴别损失。通过鉴别损失可以很好地衡量图像修复的效果,局部鉴别损失越小说明修复效果越好。

但是,在对局部鉴别网络进行大规模的训练时,特征的映射会产生“漂移”,即特征过度学习。局部鉴别网络容易产生特征过度学习的根本原因是训练时输入的图像块所在区域过小,并且来自同一个图像类别,这导致了图像块的特征过

于集中和单一。特征过度学习造成的后果是修复网络为了满足局部鉴别网络的鉴别结果,生成的图像过分地注重某个特定的特征而忽略了全局的连通性,或者是过于注重多个特征从而造成某些特征重复出现的图像。在文献[22]中,研究者通过保留上次迭代训练的一半图像来抑制这种影响。

为了克服特征漂移问题,在本文的算法中,引入了一种既包含主要特征也包含附加特征的辅助特征块。辅助特征块既要包含原本完好图像块的主要特征以保证局部鉴别网络学习的主要特征不会出现严重偏差;同时,为了避免局部鉴别网络过拟合,辅助特征块又要包含丰富的附加特征,这些附加特征提高了局部鉴别网络的泛化能力,避免局部鉴别网络特征过度学习,进而提升修复网络生成图像的真实性。同时,为了提高算法的运行效率,本文采用缓冲池技术来存储和处理辅助特征块。

辅助特征块的搜寻算法如图5所示,主要分为以下步骤:

1) 在同一个训练批次下,提取当前训练图像中修复后的破损区域(*Inpainted image*),并随机选取该批次中的 n 张完好图像。

2) 在 n 张完好图像中按序抽取一张图像,用滑动窗口在该图像中搜索上述修复区域的最佳相似块,并把相似块放入缓冲池。

3) 重复步骤2),直至找到所有相似块。

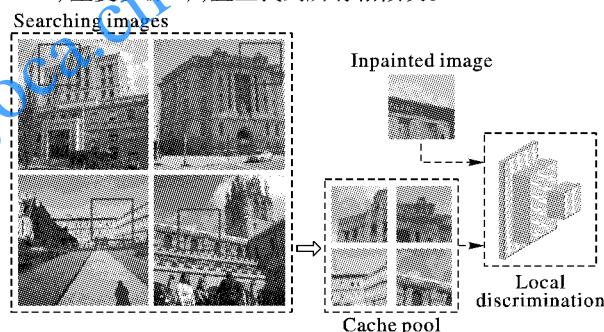


图5 缓冲池存储辅助特征块

Fig. 5 Storage of assisted feature patches by cache pool

在上述算法中寻找到的相似块就是辅助特征块,为了提高搜索效率,同时考虑到破损区域的大小,本文设置滑动窗口步长为划定的破损区域边长的一半。

在训练局部鉴别网络时,把缓冲池中辅助特征块和修复网络的修复块一起作为局部鉴别网络的训练输入。由于这些辅助特征块是在相同图像类别中的不同图像寻找的,比如人脸图像这一类别中的不同人脸实例上的特征,因此,可以保证局部鉴别网络学习到主要特征的同时还学习到很多附加特征。

在文献[22]中,局部鉴别网络的模式是输入一个修复图像,输出一个局部鉴别损失。而在引入了辅助特征块后,局部鉴别网络采用缓冲池存储的多个图像块和一个修复图像块作为输入,由此也产生了多个局部鉴别损失。为了综合反映多个局部鉴别损失的情况,本文重新定义了局部鉴别损失 $Loss_{\text{local}}$,其中,把修复网络生成的修复块输入局部鉴别网络得到第一个局部损失称为 l_{origin} ,同时,把缓冲池内的特征辅助块输入局部鉴别网络计算出相应的 n 个图像损失从而得到它们的累加和。给上述两项赋予不同的权值,组合成一个新的局部鉴别损失,如式(8):



$$\text{Loss}_{\text{local}} = \alpha l_{\text{origin}} + \beta \sum_{i=1}^n l_i \quad (8)$$

其中: l_i 表示第 i 个像素点的损失; α 与 β 为权重衡量系数, 以突显在训练的过程中对修复图像块的重视, 本文把 α 设置为 0.9, β 设置为 0.1, 并限定辅助特征块数为 4 个。

3 算法流程

使用 Mini-batch 的随机梯度下降 (Stochastic Gradient Descent, SGD) 训练生成对抗网络, 共进行 n 轮迭代。根据实验经验, 一般经过 30 轮的迭代时, 实验能取得较好的收敛结果。本文算法流程如下:

- 1) For 迭代次数的第 $1, 2, \dots, n$ 次 do:
- 2) For $k = 1, 2, \dots, K_d$ 步 do:
- 3) 从完好的训练集 $P_{\text{data}}(x)$ 中抽取一个 mini-batch 共 m 个图像 $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, 对抽取的每个图像随机切分固定大小的破损区域, 完成最邻相似填充, 作为输入数据集 $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$
- 4) 数据集输入修复网络生成相应的修复图像 $\{G(z^{(1)}), G(z^{(2)}), \dots, G(z^{(m)})\}$
- 5) 使用 SGD 算法更新局部鉴别网络和全局鉴别网络的参数 Φ_1 和 Φ_2
- 6) $\nabla_{\Phi_1, \Phi_2} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - \text{Loss})]$
- 7) End
- 8) For $k = 1, 2, \dots, K_g$ 步 do:
- 9) 把 m 个图像 $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ 输入修复网络生成相应的修复图像 $\{G(z^{(1)}), G(z^{(2)}), \dots, G(z^{(m)})\}$
- 10) 用上述步骤中训练好的两个鉴别网络 (参数 Φ_1 和 Φ_2) 来鉴别修复网络的修复图像 $\{G(z^{(1)}), G(z^{(2)}), \dots, G(z^{(m)})\}$
- 11) 使用 SGD 算法更新局部鉴别网络和全局鉴别网络的参数 Φ_1 和 Φ_2
- 12) 根据鉴别网络的结果 $\{D(G(z^{(1)})), D(G(z^{(2)})), \dots, D(G(z^{(m)}))\}$, 使用 SGD 算法更新修复网络的参数 θ
- 13) $\nabla_{\theta} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$
- 14) End
- 15) End

4 实验结果与分析

4.1 参数设置及数据集

为了验证本文修复方法的效果, 本文对图像修复进行了多轮的实例验证。实验的平台是 Windows 10, Python3.6 和 Tensorflow 结合的编程环境, Intel 4.20 Ghz CPU 时钟频率, 内存 16.0 GB。

本文进行了大量的实验来证明本文的模型在各类图像中的学习能力, 其中重点训练了建筑和脸部图像上合成缺失内容的能力。生成对抗网络训练的参数 (如学习率) 按照文献 [11] 中的建议设置, *learning rate* 为 0.001, *momentum* 为 0.5。为了平衡不同损失的影响, 实验中 λ_1 为 0.4, λ_2 为 0.6, α 设置为 0.9, β 为 0.1。

本文使用了两个数据集来验证本文的方法, 分别是牛津建筑物数据集 (The Oxford Buildings Dataset) 和 CelebA 人脸数据集来训练和评估本文的模型, 将掩模大小设置为中间覆盖, 覆盖范围设置为全图的 30% 进行训练, 以确保至少有一个必要的图像成分丢失。

4.2 实验对比

在建筑类图像修复的效果对比上, 本文采用了近年来基于深度学习并且修复效果比较好的几种算法, 分别是: 1) GlobalLocal 方法^[22]; 2) Context Encoder 方法^[23]; 3) Semantic Inpainting 方法^[12]。同时, 为了直观地展现本文修复方法的有效性和优越性, 本文采用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 指标来表示修复算法修复的图像和原图像的差距, PSNR 值越大, 代表修复的效果越好。为了进一步分析本文方法在图像修复中, 特别在结构上的优异性, 还增加了结构相似性 (Structural SIMilarity index, SSIM) 来评价修复结果, SSIM 值越接近 1, 结构越为相似。

不同方法在牛津建筑物类图像中的修改效果如图 6 所示。由图 6 可以看出, 在图像 1 中, 对比门柱和窗户的修复效果, 可以清晰看出本文的修复方法修复效果最好; 在文献 [22] 方法的修复效果图 6(c) 中, 明显可以看出建筑的修复中颜色出现了偏差, 建筑的结构修复也出现了偏离; 在图 6(d) 的修复中, 特征明显学习过度, 在修复的时候, 出现了特征重复, 图像较为模糊, 修复效果较差; 在图 6(e) 的修复中, 修复的效果较好, 结构比较清晰, 纹理也较好。图 6 中的图像 2 和图像 4 都是属于窗户特征明显的图像: 在图 6(f) 的修复中, 本文方法避免了过度学习导致的图像平滑, 在结构和纹理上修复效果都较为出众; 在图 6(c) 的修复中, 因为采用的是周边权重的恒量, 可以看到明显引入了周边的窗户对修复区域的影响, 表现为窗框重叠、扭曲等; 在图 6(d) 的修复中, 在面对这类全局特征过于明显的图像时, 出现了修复区域过平滑的缺陷, 图像较为模糊; 在图 6(e) 的修复中, 修复效果稍有缺陷, 表现为窗户的修复过于刻板、边缘的修复稍微扭曲的情况。图 6 中图像 3 是典型的曲线修复图像: 在本文方法的修复效果图 6(f) 中, 建筑的二层边缘维持了一个较顺滑的形状, 纹理上也比较符合视觉一致; 而在图 6(c) ~ (e) 的修复中, 则明显缺乏视觉连通性, 修复效果较差。不同修复方法的峰值信噪比 (PSNR) 和相应的 SSIM 值如表 1 所示。

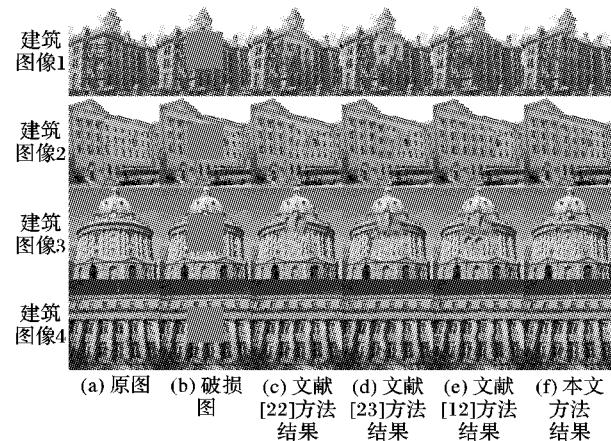


图 6 不同方法在牛津建筑物类图像中的修复效果

Fig. 6 Completion effects of different methods in building images of Oxford

因为 Context Encoder 方法^[23] 在人脸修复上效果欠佳, 本文用 Generative Face Completion 方法^[11] 来代替 Context Encoder 方法^[23]。Generative Face Completion 方法^[11] 采用迭代下降生成损失的方法来逼近最佳效果, 以此修复图像。不同方法人脸类图像的修复效果如图 7 所示。为了验证各算法在人脸大缺失区域修复上的效果, 本文采用了 40% 破损面积



进行训练。

由图 7 可以看出,本文方法很好地突显了人脸的结构信息。在图 7 修复的男性人脸图像 1 中:图 7(d)~(e)的修复方法在破损和完好区域的边缘连接出现不同程度的结构扭曲,而图 7(f)的本文方法修复中,特别是在男性右耳部分,则保持了出众的结构。在图 7 修复的女性人脸图像 2 和图像 3 中:图 7(e)的修复方法因为没有采用特征辅助学习,女性右眼都明显出现了特征失真、人脸细节模糊的状况,而图 7(f)的本文方法修复中,在两张女性的眼睛细节和脸部皮肤上都保持了较为完好的纹理信息。用 PSNR 和 SSIM 衡量不同方法的修复效果,结果如表 2 所示。

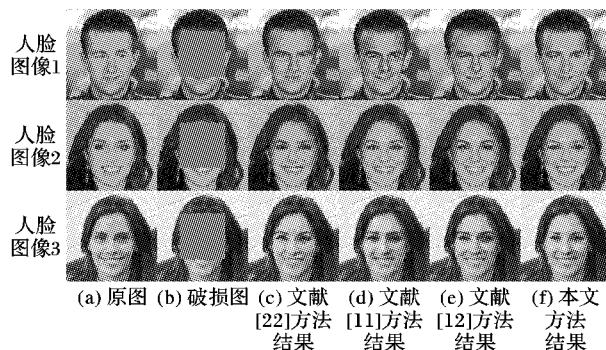


图 7 不同方法在人脸类图像中的修复效果

Fig. 7 Completion effect of different methods in face images

表 1 不同方法建筑修复图的 PSNR 和 SSIM

Tab. 1 PSNR and SSIM of building completed images obtained by different methods

图像	PSNR/dB				SSIM			
	文献[22]方法	文献[23]方法	文献[12]方法	本文方法	文献[22]方法	文献[23]方法	文献[12]方法	本文方法
建筑图像 1	22.4	19.3	21.68	22.56	0.8632	0.7944	0.8465	0.8711
建筑图像 2	23.3	20.9	22.1	24.71	0.8421	0.8054	0.8264	0.8542
建筑图像 3	17.9	17.1	17.5	21.54	0.7964	0.7757	0.7819	0.8358
建筑图像 4	20.5	18.8	18.9	21.53	0.8433	0.8241	0.8219	0.8598

表 2 不同方法人脸修复图的 PSNR 和 SSIM

Tab. 2 PSNR and SSIM of face completed images obtained by different methods

图像	PSNR/dB				SSIM			
	文献[22]方法	文献[23]方法	文献[12]方法	本文方法	文献[22]方法	文献[23]方法	文献[12]方法	本文方法
人脸图像 1	29.7	28.3	27.8	31.59	0.8585	0.7231	0.8617	0.8918
人脸图像 2	26.5	26.9	26.1	27.28	0.8894	0.8835	0.8889	0.8939
人脸图像 3	27.6	28.8	29.3	30.75	0.9061	0.8998	0.9017	0.9128

5 结语

本文提出了一个基于深度学习网络模型来进行图像修复的方法,该模型基于生成对抗网络,采用了双鉴别网络的模型。在网络的修复模型中,本文对破损图像的输入进行完好的相似信息填充处理,减少拟合损失;而在鉴别模型中,本文对全局鉴别网络采用了边缘结构信息的检测突出结构信息,对局部鉴别网络使用了辅助特征块控制过度学习的问题。实验结果表明,本文方法在不同种类数据集图像的修复中,具有良好的适应性,对破损图像也有较好的修复效果。

参考文献 (References)

- [1] TELEA A. An image inpainting technique based on the fast marching method [J]. Journal of Graphics Tools, 2004, 9 (1): 23~34.
- [2] TANG F, YING Y T, WANG J, et al. A novel texture synthesis-based algorithm for object removal in photographs [C]// Proceedings of the 2004 Annual Asian Computing Science Conference, LNCS 3321. Berlin: Springer, 2004: 248~258.
- [3] WEXLERY, SHECHTMAN E, IRANI M. Space-time video completion [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2004: I-120~I-127.
- [4] BARNES C, GOLDMAN D B, SHECHTMAN E, et al. The Patch-Match randomized matching algorithm for image manipulation [J]. Communications of the ACM, 2011, 54 (11): 103~110.
- [5] HE K M, SUN J. Computing nearest-neighbor fields via propagation-assisted KD-trees [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 111~118.
- [6] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders [C]// Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008: 1096~1103.
- [7] DOSOVITSKIY A, BROX T. Inverting visual representations with convolutional networks [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4829~4837.
- [8] KINGMA D P, WELLING M. Auto-encoding variational Bayes [EB/OL]. [2018-03-26]. <https://arxiv.org/pdf/1312.6114.pdf>.
- [9] LARSEN A B L, SØNDERBY S K, LAROCHELLE H, et al. Autoencoding beyond pixels using a learned similarity metric [C]// ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, NY: JMLR, 2016: 1558~1566.
- [10] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2672~2680.
- [11] LI Y J, LIU S F, YANG J M, et al. Generative face completion [EB/OL]. [2018-03-26]. <http://cn.arxiv.org/pdf/1704.05838v1>.
- [12] YEH R A, CHEN C, LIM T Y, et al. Semantic image inpainting with deep generative models [EB/OL]. [2018-03-26]. <http://cn.arxiv.org/pdf/1607.07539v3>.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [EB/OL]. [2018-03-26]. <http://cn.arxiv.org/pdf/1512.03385v1>.

(下转第 3595 页)



- corner motion statistics [C]// Proceedings of the 2009 IEEE International Workshop on Performance Evaluation of Trackingand Surveillance. Piscataway, NJ: IEEE, 2009: 31 – 38.
- [8] 曹志通, 李晓华, 周激流. 改进的基于角点检测的视频人数统计方法 [J]. 计算机应用, 2017, 37(S1): 141 – 143, 164. (CAO Z T, LI X H, ZHOU J L. Improved method for people counting in video sequences based on corner detection [J]. Journal of Computer Applications, 2017, 37(S1): 141 – 143, 164.)
- [9] LEMPITSKY V S, ZISSERMAN A. Learning to count objects in images [C]// Proceedings of the 2010 International Conference on Neural Information Processing Systems. North Miami Beach, FL: Curran Associates Inc. 2010: 1324 – 1332.
- [10] ARTETA C, LEMPITSKY V, NOBLE J A, et al. Interactive object counting [C]// Proceedings of the 2014 European Conference on Computer Vision. Berlin: Springer, 2014: 504 – 518.
- [11] ZHANG Y Y, ZHOU D S, CHEN S Q, et al. Single-image crowd counting via multi-column convolutional neural network [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 589 – 597.
- [12] SAM D B, SURYA S, BABU R V. Switching convolutional neural network for crowd counting [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4031 – 4039.
- [13] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60 (2): 91 – 110.
- [14] CHEN K, GONG S G, XIANG T, et al. Cumulative attribute space for age and crowd density estimation [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 2467 – 2474.
- [15] LOY C C, GONG S G, XIANG T. From semi-supervised to transfer counting of crowds [C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2013: 2256 – 2263.
- [16] LOY C C, CHEN K, GONG S G, et al. Crowd counting and profiling: methodology and evaluation [M]// ALI S, NISHINO K, MANOCHA D, et al. Modeling, Simulation and Visual Analysis of Crowds. Berlin: Springer, 2013: 347 – 382.
- [17] CHEN K, LOY C C, GONG S G, et al. Feature mining for localised crowd counting [C]// Proceedings of the 2012 British Machine Vision Conference. Durham, UK: BMVA, 2012: 1 – 11.
- [18] University of Reading. PETS 2009 benchmark data [EB/OL]. [2018-03-26]. <http://www.cvg.reading.ac.uk/PETS2009/data.html>.

This work is partially supported by the National Natural Science Foundation of China (11176018), the Chengdu Industrial Cluster Collaborative Innovation Program (2016-XT00-00015-GX), the Dongguan Social Science and Technology Development Program (2017507102428).

易国宪, born in 1994, M. S. candidate. His research interests include digital image processing, intelligent monitoring.

熊淑华, born in 1969, Ph. D., associate professor. Her research interests include multimedia communication.

何小海, born in 1964, Ph. D., professor. His research interests include digital image processing.

吴小红, born in 1970, Ph. D., associate professor. Her research interests include image processing, pattern recognition.

郑鑫波, born in 1984, Ph. D., senior engineer. His research interests include digital image processing.

(上接第 3562 页)

- [14] UIJLINGS J R R, van de SANDE K E A, GEVERS T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154 – 171.
- [15] ZITNICK C L, DOLLAR P. Edge boxes: locating object proposals from edges [C]// Proceedings of the 2014 European Conference on Computer Vision. Berlin: Springer, 2014: 391 – 405.
- [16] OUYANG W L, LUO P, ZENG X Y, et al. Deep ID-Net: multi-stage and deformable deep convolutional neural networks for object detection [EB/OL]. [2018- 03- 26]. <https://arxiv.org/pdf/1409.3505.pdf>.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2018- 03- 26]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904 – 1916.
- [19] LE Q V. Building high-level features using large scale unsupervised learning [C]// Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2013: 8595 – 8598.
- [20] KINGMA D P, REZENDE D J, MOHAMED S, et al. Semi-supervised learning with deep generative models [EB/OL]. [2018-03-26]. <https://arxiv.org/pdf/1406.5298.pdf>.
- [21] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from simulated and unsupervised images through adversarial training [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2242 – 2251.
- [22] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36 (4) : Article No. 107.
- [23] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2536 – 2544.
- [24] 王坤峰, 荀超, 段艳杰, 等. 生成式对抗网络的研究进展与展望 [J]. 自动化学报, 2017, 43(3): 321 – 332. (WANG K F, GOU C, DUAN Y J, et al. Generative adversarial networks: the state of the art and beyond [J]. Acta Automatica Sinica, 2017, 43(3): 321 – 332.)

This work is partially supported by the National Natural Science Foundation of China (61461048, 61661047).

刘宝宁, born in 1993, M. S. candidate. His research interests include digital image processing.

翟东海, born in 1974, Ph. D., associate professor. His research interests include digital image processing, mass data mining.