

文章编号: 2095-2163(2020)04-0009-04

中图分类号: TP391

文献标志码: A

# 基于生成对抗网络的图像修复算法

王可新, 王力

(贵州大学 大数据与信息工程学院, 贵阳 550000)

**摘要:** 图像修复在研究中仍然具有挑战性, 现阶段的图像修复算法往往存在修复边缘模糊, 修复区域与其周边区域存在明显不一致的问题。本文提出一种有效的基于 MSE 损失函数和感知损失函数相结合的图像修复算法。该网络模型由修复网络、全局判别网络和局部判别网络组成, 经过对网络的多次迭代和参数更新后, 最终能够完成人脸修复。经在 celebA 数据集上进行测试, 通过主观和客观的评价标准证明可以获得良好的修复结果。

**关键词:** 人脸修复; 感知损失函数; 局部判别网络

## Image completion algorithm based on Generative Adversarial Nets

WANG Kexin, WANG Li

(Department of Big Data and Information Engineering, Gui Zhou University, Gui Yang 550000, China)

**[Abstract]** Image restoration is still challenging in today's research. At present, image restoration algorithms often have the problem of repairing edge blur, and the repaired area is obviously inconsistent with its surrounding area. This paper proposes an effective MSE loss function and content loss. A combination of functions for image restoration algorithms. The network model consists of a completion network, a global discriminator network, and a local discriminator network. After multiple iterations of the network and parameter updates, the face repair can be completed. The experiment was tested on the celebA dataset, and good completion results were obtained by subjective and objective evaluation criteria.

**[Key words]** the face of a repair; content loss; a local discriminator network

## 0 引言

随着人工智能技术的快速发展, 图像作为携带信息的主要载体之一, 其作用不言而喻。然而现实生活中, 时常会出现图像不完整的现象, 比如年代久远造成文物破损、图像传递过程中隐藏内容等, 都会导致一定程度的图像中所携带的信息丢失, 因此如何对这些受损进行图像修复成为计算机视觉中的一个重要问题。

在计算机的图形学中, 图像的修复有 3 种方法: 一种是传统的基于结构的方法, 这种方法可以合成比较合理的静态纹理特征, 但仅适用于受损区域较小的图像; 第二种是基于纹理合成的图像修复, 此方法可以从图像中其他完整区域选择与缺失区域纹理相同的样本块。但当涉及语义图像损失时, 此方法在缺失区域的细节方面恢复的很差, 因此需要 1 个额外的部件提供合理的想象力, 比如来自机器的幻觉, 即第三种方法, 基于深度学习的图像修复。

本文提出的图像修复算法则是基于深度学习生成对抗网络<sup>[1]</sup>。与以往图像修复算法不同, 该算法包含 2 个判别器网络模型, 即全局判别器和局部判别

器。模型的网络结构都是采用卷积神经网络来实现。

## 1 相关工作

以前已经提出过多种用于图像修复的方法, 传统的基于结构的图像修复方法采用扩散原理, 将可见区域的信息传播到缺失区域。其主要采用高阶偏微分方程模型算法(PDE 算法)。PDE 算法又命名为 BSCB 修补模型, 模型原理是沿着等照度线方向扩散到待修复区域, 但未顾及图像的细节部分, 修复效果较差<sup>[2]</sup>。随后出现了全变分模型(TV 模型), 然而只对填充小的破损区域或者去除杂散的噪声是有用的。

与基于结构的方法相比, 纹理合成的方法可以修复大一些的破损区域<sup>[4-6]</sup>。比如典型的 PatchMatch 方法, 此方法在图像的可用部分搜索类似的纹理结构, 其高质量和高效率迅速成为当时最成功的修复方法<sup>[7]</sup>。但是这种方法假设可以在其他地方找到修复区域的纹理, 这种假设并不是一直会成立, 如果缺失区域很大, 则很难修复。

用来进行图像修复的第一种深度学习方法 Context Encoder, 使用编码器-解码器结构<sup>[8]</sup>。编码

**作者简介:** 王可新(1996-), 女, 硕士研究生, 主要研究方向: 深度学习; 王力(1971-), 男, 教授, 硕士生导师, 主要研究方向: 大数据、深度学习、图像处理。

收稿日期: 2019-10-25

器将缺失区域的图像映射到低维度特征空间,解码器用它来构造输出图像。但是该方法存在修复边缘处不连续的问题,修补痕迹明显,视觉效果较差。上述方法采用监督学习的方式,2014年GAN由Ian Goodfellow首次提出,GAN采用的是一种无监督的学习方式训练。针对Context Encoder的问题,Iizuka等人使用两个判别器使得缺失区域和整张图像在视觉上合理化,本文则是基于Iizuka等人的基础上采用感知损失和MSE损失函数相结合的方法训练生成对抗网络。

## 2 网络结构

### 2.1 生成对抗网络

网络主体采用生成对抗网络(GAN),GAN的产生其中思想为“二人零和博弈”<sup>[9]</sup>,与其他普通网络不同,生成对抗网络由两个模型组成,一个是用来生成数据的生成模型G,另一个是用来判别生成样本真假性的判别模型D。如图1所示生成模型G以1个随机噪声作为输入,将其生成1张伪造的样本,判别器模型D以真实样本或者伪造样本作为输入,预测该样本是真实的还是伪造的。生成网络的目的是能够欺骗判别器网络,因此随着逐步的迭代训练,生成网络生成越来越逼真的样本像,甚至可以达到以假乱真的效果,以至于判别器无法判断其真伪性;在这个过程中,判别器也在不断的提高自己辨别真伪的能力,使得生成器能够生成更加真实的样本。

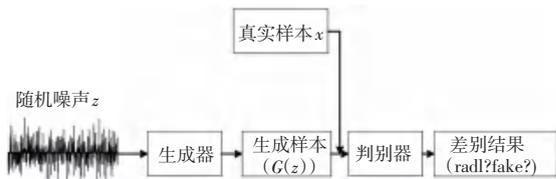


图1 GAN模型

Fig. 1 GANmodel

### 2.2 生成器

生成器采用Iizuka等提出的网络结构模型,网络模型采用典型的编码器-解码器结构。编码器采用带掩码的图像作为输入,通过下采样进行图像的特征提取,解码器通过这些特征恢复为原始的分辨率,生成图像中掩码部分的内容<sup>[10]</sup>。在图像修复任务中,通常想要了解更多的图像中潜在的特征,因此网络模型中间采用了扩张卷积,扩张卷积在基础的卷积核上加入间隔,在不损失特征图尺寸的情况下增大了感受野,从而获取图像中更多的信息。总的来说,生成器包括六次下采样的编码器,接着是三次扩张卷积和将图像上采样到原始尺寸的解码器。生

成器的整体组成如图2所示。应当注意的是,网络结构中并未使用池化层,而是将卷积步幅设置为2,将待修复图像的大小缩小为原来的一半,生成器网络模型仅仅降低了两次分辨率,使用大小为原始大小四分之一的卷积,其目的是为了降低最终图像的纹理模糊程度。此外,随着卷积网络层数的加深,数据的波动浮动较大,会大大的降低网络训练的速度。因此,本文除了网络的最后一层不使用批量归一化处理,其余所有层都使用批量归一化处理<sup>[11]</sup>;同时除最后一层卷积使用Tanh函数外,其余所有层都使用Relu激活函数。

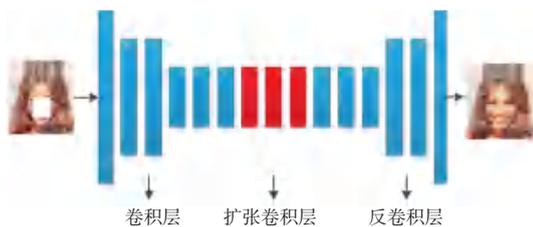


图2 生成器模型

Fig. 2 Generator model

### 2.3 判别器

通过对生成器使用损失函数训练后,生成器对缺失区域进行补充。但是补充的效果并不能保证与真实图像保持一致。为了图像生成的更加真实合理,在本文中引入全局判别器和局部判别器来鉴定图像是真实图像还是生成器修复的图像。判别器的主要作用是通过与生成器不断的对抗训练,最终能够更准确的识别真假图像。同样地,判别器也由一系列卷积层组成,如图3所示。

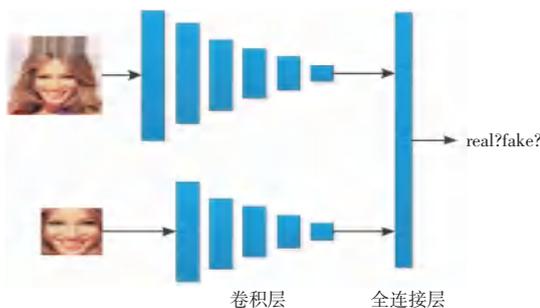


图3 判别器模型

Fig. 3 Discriminator model

全局判别器的输入为完整的整幅图像,网络由5个卷积层和1个全连接层组成。全局判别器从整体上判断生成的图像是否真实合理,即整幅图像的上下文信息是否一致。局部判别器的输入为整幅图像的一半,网络由4个卷积层和1个全连接层组成,局部判别器只检测以补充区域为中心的部分,用来判断局部修复的纹理特征是否合理。局部判别器与

全局判别器相同,除网络最后一层外,所有层都使用批量归一化处理。同时,两个判别器除了最后一层使用 sigmoid 激活函数,其余卷积层都是使用 relu 激活函数。最后,将两个判别器最后输出的向量连接到一起,得到 1 个 2048 维的向量,向量最后使用 sigmoid 激活函数,生成 0~1 之间的数值,这个数值的大小决定了输入图像为真实图像的概率。

### 3 损失函数

通过缩小修复图像与真实图像之间的距离来达到修复缺失图像的目的,因此本文采用感知损失<sup>[12-13]</sup>与 MSE 损失相结合的方式训练生成器模型。其次,使用 GAN 损失联合更新判别器和生成器模型,使得生成的图像更加真实。

GAN 损失采用生成对抗网络的对抗训练策略,其本质是求最大最小的优化问题。这体现了生成器不断优化自己,生成更加真实的图像欺骗判别器,判别器也不断更新优化,使得能够正确的区分真伪图像。GAN 损失的表达式(1):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\lg D(x)] + E_{z \sim P_z(z)} [\lg(1 - D(G(z)))] \quad (1)$$

其中,  $P_{data}(x)$  为真实数据  $x$  的分布,  $P_z(z)$  为噪声变量  $z$  的分布,  $G(z)$  为生成器的输出。

生成器的损失函数包含两个部分, MSE 损失函数表示为(2):

$$L_M = \|M \times (G(X, M) - X)\|^2 \quad (2)$$

其中,  $X$  为待修复的图像  $M$  时输入的掩码,  $x$  为像素的乘积。

感知损失函数,使用预训练的 VGG19 的 relu 激活层输出的特征图。在本文实验中,特征图选择 relu1-1, relu2-1, relu3-1 三层的输出。感知损失定义为(3):

$$L_p = E \left[ \sum_j \frac{1}{C_j H_j W_j} \|\phi_j(I_{gt}) - \phi_j(I_{pred})\|_1 \right] \quad (3)$$

$C_j H_j W_j$  表示第  $j$  层特征图的大小,  $\phi_j$  是预训练网络第  $j$  层的特征图。

综合上述 GAN 损失函数和生成器损失函数可得(4):

$$L_{all} = L_M + \alpha \min_G \max_D V(D, G) + \beta L_p \quad (4)$$

其中,  $\alpha, \beta$  为训练时的权重。

## 4 实验

### 4.1 实验平台及内容

本实验基于 windows 操作系统的 tensorflow 框架下运行,python 为编程语言。实验设备 CPU: Inter

i7-8700, GPU: RTX2070S-8G, 内存: 16G。

实验的数据集为 celebA 人脸数据集,原始数据集大小为  $178 \times 218$ ,在模型训练时将图片缩为  $128 \times 128$  的图像,全局判别器的输入为  $128 \times 128$ ,而局部判别器的输入大小为修复区域为中心的  $64 \times 64$ 。实验选取数据集的 100 000 张图像,其中 80 000 张用来模型的训练,20 000 张用来模型的测试。在训练过程中,生成器的网络迭代 20 000 次,判别器网络迭代 10 000 次,生成器和判别器联合迭代 70 000 次。

### 4.2 实验过程

本文的训练分成 3 个阶段。首先,利用感知损失和 MSE 对生成器进行训练,使之生成大致的修复图片;其次,固定生成器,根据上一阶段训练的生成器对判别器进行训练,利用 GAN 损失对判别器迭代更新;最后,对生成器和判别器一起训练,直至训练结束。后续采用泊松图像混合对图像进行处理。

### 4.3 评估标准

实验结果的评估标准由定性分析和定量分析两部分组成,将本文的实验结果与自动编码器、DCGAN 以及 Context Encoder 这 3 种不同的方法进行比较。在定性分析中,3 种对比方法中的视觉图像效果引用文献[15]中的效果图。为了直观的比较,本文把最后的修复效果图裁剪为相同的人脸位置。各个方法的修复图如图 4(从左至右分别为待修复图、自动编码器、DCGAN、Context Encoder、本文、原图)所示。



图 4 对比图

Fig. 4 Comparison picture

由图 4 可知,自动编码器方法的效果最差,这是因为自动编码器只由生成器模型生成,无法将损失降到最小,修复区域出现了结构扭曲现象;对比自动编码器,DCGAN 方法的图像修复效果从内容上保持了一致性,但是图像的局部细节修复的效果较差;Context Encoder 方法在内容和局部细节上都表现较好,但边缘连接区域修复痕迹明显,缺乏视觉连通性。本文方法与 Context Encoder 方法相比,修复效果与原始图像更加相似,边缘修复痕迹也比较淡。

定量分析使用峰值信噪比 (PSNR) 和结构相似性 (SSIM) 两个指标来衡量图像的修复效果。PSNR

是常用的判断图像质量的评价指标,虽然有时会出现数值结果与人的主观感受不一致的情况,但仍然可以作为一个有效的评判标准。其主要比较图像对应像素之间的差别,PSNR 数值越大,表明失真越小。另一个主要的评价标准 SSIM,从亮度、对比度和结构 3 个方面来衡量,用均值作为亮度的估计,标准差作为对比度的估计,协方差作为结构相似程度的度量。得到的数值大小体现了两幅图像的相似程度,数值越大,代表两幅图像差距越小,修复的效果越好。计算的结果如表 1、表 2 所示。

表 1 不同图像补全方法的 PSNR 值

Tab. 1 The PSNR of different image completion methods dB

| 方法    | 图 1   | 图 2   |
|-------|-------|-------|
| 自动编码器 | 20.68 | 22.40 |
| DCGAN | 23.16 | 23.85 |
| CE    | 22.53 | 21.96 |
| 本文方法  | 25.95 | 25.78 |

表 2 不同图像补全方法的 SSIM 值

Tab. 2 The SSIM of different image completion methods

| 方法    | 图 1   | 图 2   |
|-------|-------|-------|
| 自动编码器 | 0.738 | 0.825 |
| DCGAN | 0.849 | 0.839 |
| CE    | 0.832 | 0.815 |
| 本文方法  | 0.903 | 0.872 |

## 5 结束语

基于生成对抗网络模型对图像进行修复,生成器采用编码器-解码器结构,判别器全局判别器和局部判别器,用来保证整体修复的一致性和局部信息细节的合理性。训练过程使用基于 VGG 预训练特征图的感知损失与 MSE 相结合训练生成模型,修复效果与自动编码器、DCGAN 以及 Context Encoder 方法进行比较。根据定量分析和定性分析可知,本文方法实现了较好的修复效果。然而模型的修复效果非常依赖训练集,对于除人脸修复的其他图像修复,则需要重新训练数据集,对于建筑物及风景的修复则需要进一步的研究与训练。

## 参考文献

[1] GOODFELLOW I J, POUGET - ABADIE, MIRZA M, et al.

Generative adversarial networks[C]//Proceedings of Advances in Neural Information Processing Systems, Montreal, Dec 8 - 13, 2014. Denver: NIPS Press, 2014; 2672-2680.

[2] COLOMABALLESTER, MARCELO BERTALMIÓL, VICENT CASELLES, et al. Filling-in by joint interpolation of vector fields and gray levels [J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200-1211.

[3] SHEN J, CHAN T F. Mathematical models for local nontexture inpaintings[J]. SIAM Journal on Applied Mathematics, 2002, 62(3): 1019-1043.

[4] BERTALMIÓ M, VESE L, SAPIRO G, et al. Simultaneous structure and texture image inpainting [J]. IEEE transactions on image processing, 2003, 12(8): 882-889.

[5] EFROS A A, LEUNG T K. Texture synthesis by non-parametric sampling [C]//Proceedings of the seventh IEEE international conference on computer vision. IEEE, 1999, 2: 1033-1038.

[6] GROSSAUER H. A combined PDE and texture synthesis approach to inpainting [C]//European conference on computer vision. Springer, Berlin, Heidelberg, 2004; 214-224.

[7] BARNES C, GOLDMAN D B, SHECHTMAN E, et al. The patchmatch randomized matching algorithm for image manipulation [J]. Communications of the ACM, 2011, 54(11): 103-110.

[8] 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. 2016; 2536-2544.

[9] 王坤峰, 苟超, 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.

[10] 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.

[11] LONG J, SELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015; 3431-3440.

[12] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [J]. arXiv preprint arXiv:1802.05957, 2018.

[13] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 2414-2423.

[14] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C]//European conference on computer vision. Springer, Cham, 2016; 694-711.

(上接第 8 页)

[6] SILVER D, HUBERT T, SCHRITTWIESER J, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play [J]. Science, 2018, 362(6419): 1140-1144.

[7] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. nature, 1986, 323(6088): 533-536.

[8] ROSIN C D. Multi-armed bandits with episode context [J]. Annals of Mathematics and Artificial Intelligence, 2011, 61(3): 203-230.

[9] MNH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari

with deep reinforcement learning [J]. arXiv preprint arXiv:1312.5602, 2013.

[10] CHASLOT G M J B, WINANDS M H M, VAN DEN HERIK H J. Parallel monte-carlo tree search [C]//International Conference on Computers and Games. Springer, Berlin, Heidelberg, 2008: 60-71.

[11] KOCSIS L, SZEPESVÁRI C. Bandit based monte-carlo planning [C]//European conference on machine learning. Springer, Berlin, Heidelberg, 2006; 282-293.

[12] COULOM R. Efficient selectivity and backup operators in Monte-Carlo tree search [C]//International conference on computers and games. Springer, Berlin, Heidelberg, 2006; 72-83.