基于生成式对抗神经网络的手写文字图像补全

李农勤[1]， 杨维信[2,3]

(1. 东华理工大学经济与管理学院，江西 南昌 330013；
2. 华南理工大学电子与信息学院，广东 广州 510641；
3. 牛津大学数学研究所，牛津 OX26GG)

摘 要：手写文字图像补全是图像补全问题中一个重要研究分支，其难点在于图片中具有无约束书写风格的文字的结构关系补全。为了模拟实际中复杂和困难的应用情景，在图像补全研究工作的启发下，针对大类别、小样本、多风格、未知语种等复杂情况下进行手写象形文字图像补全。采用全局和局部一致性保持的生成式对抗神经网络(GLC-GAN)。在大类别多风格的手写文字图像补全中，补全图片往往因可能的补全候选很丰富而导致补全区域模糊不清。为此，提出两级补全系统：第一级粗补全模块考虑文字结构的完整性，第二级细补全模块实现文字的清晰化、细致化。通过在大类别手写汉字数据库 CASIA-HWDB1.1 上的实验，验证了该两级系统的有效性，同时分析系统在不同书写风格和不同缺失区域情况下的补全效果。

关 键 词：生成式对抗网络；手写文字；图像补全；结构补全；自监督学习
中图分类号：TP 391 **DOI：**10.11996/JG.j.2095-302X.2019050878
文献标识码：A **文 章 编 号：**2095-302X(2019)05-0878-07

Handwritten Character Completion Based on Generative Adversarial Networks

LI Nong-qin[1], YANG Wei-xin[2,3]

(1. School of Economics and Management, East China University of Technology, Nanchang Jiangxi 330013, China;
2. School of Electronic and Information Engineering, South China University of Technology, Guangzhou Guangdong 510641, China;
3. Mathematical Institute, University of Oxford, Oxford OX26GG)

Abstract: Handwritten character completion is an important research topic in image completion. Its challenge comes from the completion of the structural relationships in handwritten characters with unconstrained handwritten styles. To simulate the complicated and difficult situations in the real-world applications, the paper focuses on handwritten pictographic characters with large category, small sample size, multiple unconstrained handwritten styles, and unknown language (i.e., with no access to the class label of each character). Inspired by the progress in natural image completion, the generative adversarial network with global and local consistency was leveraged to achieve handwritten character completion. Under the circumstances of large category and various writing styles, the completion areas of character completion suffer from low-fidelity because of the large number of potential completion candidates. To solve this problem, a two-stage character completion system was proposed: the first stage is coarse-grained completion module ensuring the completeness of the character; the second stage is fine-grained completion module improving the sharpness and details of characters. Extensive experiments were conducted on CASIA-HWDB1.1 to validate the

收稿日期：2019-04-22；**定稿日期：**2019-06-09
第一作者：李农勤(1960–)，男，江西广昌人，副教授，硕士。主要研究方向为运作管理等。E-mail：nqli@ecit.en
通信作者：杨维信(1990–)，男，广东广州人，研究员，博士后。主要研究方向为机器学习、计算机视觉等。E-mail：wxy1290@163.com

effectiveness of the two-stage system and analyze the completion performance under different writing styles and different conditions of missing area.

Keywords: generative adversarial network; handwritten character; image completion; structure completion; unsupervised learning

图像补全(image completion)是对图片中缺失或受污染的部分进行填充，使得填充后的图片尽可能完整和逼真。图像补全是近年来在机器学习领域中非常热门的研究课题，在图片修复、编辑或重构等应用中起到重要作用[1-3]。图像补全需要注重图片中内容的纹理和语义结构关系。目前较多研究针对自然场景图片中的纹理信息的补全，补全后图片比较逼真[1-2,4]。在对图片内容的结构补全上，目前较多的研究工作在人脸图片上展开[2,4]。通过对大量归整后的人脸图片进行训练，人脸上缺失(故意挖空)部分能够被补全，保持了人脸的结构的完整性。然而，目前的大部分方法在面对具有强语义结构关系同时又复杂多变的物体时，补全结果不佳[5]。例如文献[2]最后展示了人体图像补全时出现的错误，其中头部像素难以合理补全。文字补全问题是图像补全的一个分支，与自然场景中纹理信息的补全不同，文字补全重点需要解决的是结构关系信息的补全，可以作为研究结构信息补全的重要研究对象。文字补全除了具有重要科研意义外，还具有很高的社会应用价值。诸如埃及象形文字、苏美尔文、古印度文和中国甲骨文、中国古体字等象形文字，其书写载体一般是石头、石壁、龟甲、竹简等。受到长期自然因素和人为因素的影响，书写载体会有不同程度的侵蚀和破损，载体上的文字也随之出现部分缺失情况，此时，文字图像补全的工作能在一定程度上缓解这种不利影响。进一步地，对于未破译的象形文字，由于文本内容无法获取，人工的文字补全将因为不能运用文本间的语言模型而变得更加困难，此时更能体现无类别标记的文字图片自动补全研究的意义。为了研究图片中的结构关系的自动补全问题，辅助历史考古应用场景中的文字图片分析，本文着重研究手写文字图像补全，简称手写文字补全或文字补全。

现有的针对文字补全的研究[6-7]主要是在 0 到 9 的 10 类数字图片上进行的，主要利用 MNIST 数据集[8]或 SVHN 街景房屋号码数据集[9]，其每类数字均有大量的训练样本，例如 MNIST 中每类数字有 6 000 个手写数字样本；SVHN 数据集共有 73 257 张训练图片，其中每张图片展示了一个或多个印刷体数字。大量的训练数据给深度学习模型的优化带来便利，使得最后数字图像补全效果很好。然而，实际应用中的文字图片情况非常复杂：①常用文字的类别数远远大于 10 类，大类别数使得类与类之间的相似程度也大幅提升；②大类别数的数据集在采集过程中不能保证每类文字都有大量的训练样本；③手写文字与较为规整的印刷体文字相比，拥有各式各样的书写风格，大大增加了补全的复杂程度。

鉴于目前针对上述复杂文字情况下的图像补全研究的缺乏，本文的研究问题限定为大类别、小样本、多风格、未知语种的手写象形文字补全。又鉴于汉字是上古时期各大文字体系中唯一传承至今的文字，本文采用汉字作为研究对象。相比于目前其他语言的文字而言，汉字的类别数特别大，例如 GB18030-2005 收入的汉字个数达到七万多个，其中最常用的汉字也有 3 755 个；而且，汉字是目前世界上使用人数最多的文字，不同人书写风格各异，造成手写汉字数据的多样性，因此也提高了文字补全的复杂性。实际历史文档的应用场景中，往往会出现未知语种的象形文字；为了模拟这种情况，本文研究文字补全只使用每个汉字样本的图片像素信息，并不使用汉字的类别标签信息。最后，大部分历史文档在收集和整理中存在困难，这也往往造成可供研究的文字样本较为缺乏，因此，本文研究采用小样本的文字数据进行实验。

在自然图片的补全中，目前很多研究方法采用了生成式对抗网络[10]。其中，文献[2]中基于全局和局部一致性保持的生成式对抗网络(global and local consistent generative adversarial networks，GLC-GAN)在图片的判别网络中设计了 2 条支路，分别从全局和局部的角度来衡量待填充部分的完整性和逼真程度，促使生成网络得到的填充部分能兼顾全局完整和局部逼真。受到该工作的启发，考虑到在文字补全中全局结构信息的重要性，本文采用

GLC-GAN 模型来实现无约束手写文字的补全。针对大类别无类别标签手写文字补全中遇到的补全区域图片的模糊问题，提出两级补全方案：第一级补全模块结合全局和局部考虑各种可能的补全情况得到模糊的粗补全；第二级补全模块则进一步将图片做细致化、清晰化处理。通过在手写数据集 CASIA-HWDB1.1[11]上进行实验，验证了本文方法的有效性，同时探究和分析了 GLC- GAN 模型在不同的书写风格和不同的缺失区域情况下的填充补全效果。

1 两级生成式对抗网络的文字补全

1.1 基于全局和局部一致性的生成式对抗网络

生成式对抗网络(generative adversarial network, GAN)最初是为了训练基于卷积神经网络的生成网络而设计的一种自监督学习方法[10]。其中自监督是指该网络的训练无需额外数据标签信息作为监督信号，仅通过判断生成产物的真实程度来优化模型。一般来说，GAN 由 G 网络和 D 网络 2 部分组成，即生成网络(generative network)和判别网络(discriminative network)。图像补全中，生成网络 G 通过将给定的图片作为输入，输出一张补全图片。判别网络 D 则负责判别由 G 补全的图片是否“真实”。GAN 的训练过程是 G 和 D 两者的相互迭代博弈和促进的过程，最终理想状态是判别网络 D 难以判定 G 生成图片的真伪，即认为此时的 G 能生成能以假乱真的图片。

对于图像补全任务，不仅需要令填补的部分图片内部更加逼真，而且还需要使填补后的全图也更加连贯和逼真。为此，文献[2]设计了全局和局部一致性保持的 GAN 框架。该框架下，判别网络 D 由 2 个分支组成：全局分支 D_1 和局部分支 D_2。2 个分支网络的最后一层瓶颈特征经过拼接后，再通过一个全连接层，得到一个数值范围在 0 到 1 之间的置信度输出。在 GLC-GAN 的训练中，网络的优化包含了 2 种损失函数：加权均方误差损失函数 L_{MSE}[12] 和 GAN 损失函数[10]。其中，L_{MSE} 的计算公式为

$$L_{MSE}(x,M_c)=\left\|M_c\odot[G(x,M_c)-x]\right\|^2 \qquad (1)$$

其中，x 为完整的无缺失的图片；M_c 为缺失区域位置的掩膜；$G(x,M_c)$为生成网络的输出图片；⊙为对应位置元素相乘。由此可见，L_{MSE} 是衡量真实图片和 G 伪造的图片之间的差异。对于另一部分，GAN 的训练是 min-max 优化过程，目标函数为

$$\min_G \max_D E[\ln(D(x,M_d))+\ln(1-D(G(x,M_c),M_c))] \qquad (2)$$

其中，D 为判别网络；M_d 为随机缺失区域位置的掩膜。结合 2 种损失函数，GLC-GAN 模型最终的目标函数为

$$\min_G \max_D E[L_{MSE}(x,M_c)+\alpha\ln(D(x,M_d))+\alpha\ln(1-D(G(x,M_c),M_c))] \qquad (3)$$

其中，α 为加权的超参数。GLC-GAN 中判别网络 D 同时衡量全局和局部的逼真程度，使得生成网络 G 能够生成局部逼真、全局合理的补全图片。

1.2 两级文字补全系统

文字补全的关键要素是结构合理和图片清晰，尽管通过 GLC-GAN 能够使得结构更加合理，但补全图片的模糊现象仍然显著。对于一般图片的纹理补全任务中，缺失区域四周的邻近纹理信息给补全图片带来很多纹理约束，因此较容易补全得到清晰逼真图片；然而对于注重结构关系的文字而言，图片中缺失区域之外的剩余部分仅能够提供有限的结构约束，换句话说，缺失区域可供选择的补全图片的搜索范围很大。进一步地，对于手写体文字，就算是一般被视为具有相同结构的同类别文字，也会因为多样的书写风格的存在，在缺失区域呈现多样的补全图片候选。

为了使补全区域清晰化，引入基于 GLC-GAN 的两级补全模块：第一级 GAN 注重结构的合理，第二级 GAN 注重图片的清晰。图 1 展示两级手写文字补全网络的总框图。

第一级 GLC-GAN 网络基本上采用 1.1 小节的架构，但对于网络 D 中标记为 1 的样本不再是数据库中真实样本，而是在真实样本基础上，对局部区域进行高斯模糊，得到局部模糊样本。局部模糊区域的抽取规则和缺失区域一致，高斯模糊的窗长参数是在预设定的范围内的随机值(本文在一半的训练迭代中选取[3,15]内的随机奇数值作为高斯模糊的窗长，另一半的迭代中采用原始图片，即不进行高斯模糊操作)。局部模糊样本相对于真实样本而言，在第一级的判别网络 D 中，为生成网络 G 输出图片的清晰程度提供了容忍度，从侧面增大了网络 D 做出判决的难度，使第一级的 D 不再只关注图片的清晰，而更多是去考虑补全后图片在结构上是否合理，进而促进第一级生成网络 G 给出结构合理的补全结果。

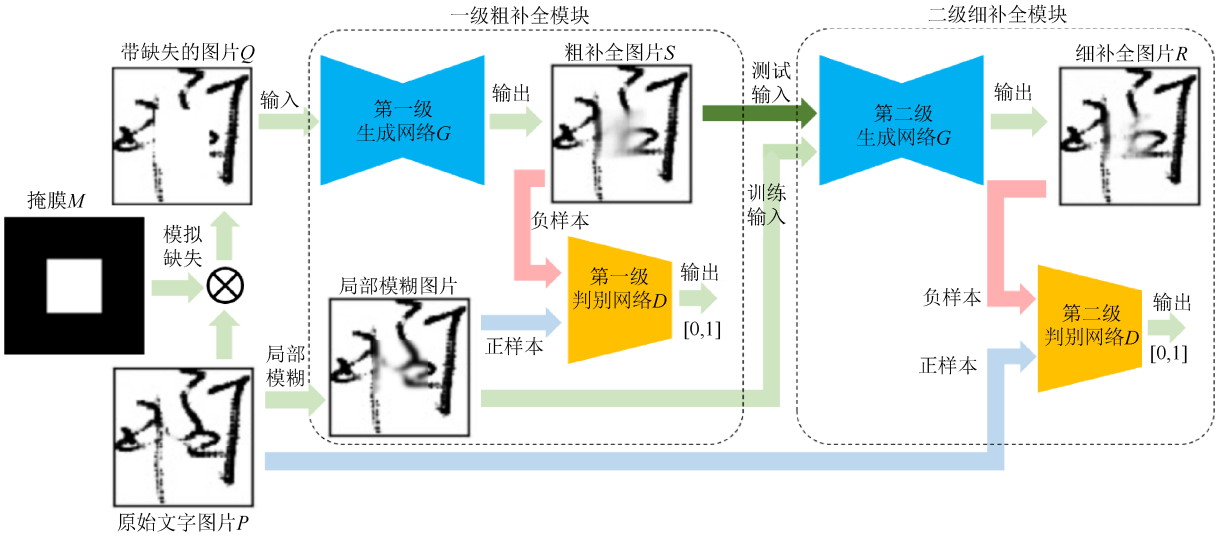


图 1　两级手写文字补全系统的总框图

第二级 GLC-GAN 网络是清晰化补全模块。在训练中，G 的输入是随机生成的局部模糊样本(高斯模糊窗长取[11,51]范围内随机奇数)，监督信号则是数据库中的真实样本。在实际应用中，G 的输入是第一级生成网络输出的较模糊的粗补全图片，输出则作为最终的补全结果。值得一提的是，训练阶段采用随机生成的局部模糊样本是因为该模块只需专注于在不改变图片文字内容的前提下，使图片清晰化，而不需要考虑如何改变文字结构。

图 1 中，给定一个完整书写的 $N\times N$ 大小的原始文字图片 P，随机生成一张 $N\times N$ 大小的二值掩膜 M，掩膜 M 中数值为 1 的正方形区域代表被污染后文字的缺失区域，区域的位置和大小是随机的。此时，带缺失区域的文字图片 Q 便可通过文字图片 P 和掩膜 M 计算得到

$$Q = P\times(1-M) \tag{4}$$

带缺失区域的文字图片 Q 经过第一级 GAN 的生成网络 G 后得到粗补全图片 S。图片 S 经过第二级 GAN 的生成网络 G 后得到细补全图片 R，即最终补全结果。本文所有 GAN 都采用上一小节介绍的 GLC-GAN，即判决网络 D 包括全局分支和局部分支：前者的输入是粗补全图片 S (对于第二级 GAN 则是图片 R)，后者的输入则是由预定义二值掩膜 M 从对应的全局图中抽取出的局部区域图。本文中的 GLC-GAN 模型的相关超参数与文献[2]中的大体一致，不同之处在于：①由于文字图片比自然场景图片未见太多纹理细节，因此图片被统一归一化成 128×128 的大小；②在训练阶段，随机缺失区域的边长大小是 32 到 64 之间的随机值，即缺失区域最大占原图的四分之一，最小占原图的十六分之一。

2　实验结果与分析

2.1　实验数据库

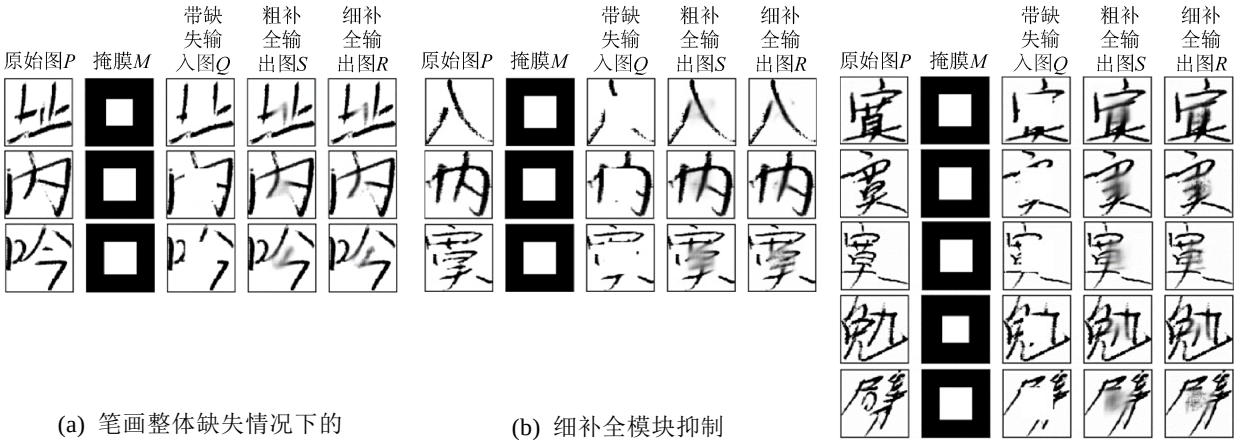
CASIA-HMDB1.1[11]是经典的脱机手写汉字数据库。脱机表示手写汉字数据是以图片形式呈现。该数据库包含了 3 755 类的常用汉字，每类汉字大约由 300 位书写者提供样本。本文实验模拟特殊且困难情况下的文字补全，主要针对大类别、小样本、多风格、未知语种的象形。为了体现大类别，实验考虑了数据库中所有类别(即 3 755 类)的汉字。为了模拟小样本、多风格应用场景，在模拟训练阶段只使用前 20 位书写者的样本，其余书写者的手写样本都用于效果的检验。为了突出未知语种的情况，所有样本的类别标签信息均未用于本文的实验中，本文模型都是通过自监督学习得到。

2.2　不同书写风格的文字补全效果

由于不同个性风格的存在，相同类别(即相同结构)的文字在书写中也会呈现多样化，此时文字补全的难度也有差异。如图 2 所示 9 组补全结果，每组结果来自同类别且由不同书写者提供的汉字。本文掩膜 M 中的缺失区域固定在中心位置，且区域大小为 40×40 或 50×50，粗补全输出图 S 是在只采用一级 GLC-GAN 下得到最终补全效果，即在该 GLC-GAN 的训练中只利用原始真实样本作为判决网络的正样本。

图 2 不同书写风格的汉字样本的两级补全效果

由图 2 每组结果可见，大部分情况下，虽然带缺失区域的图 *Q* 会在一定程度上影响人们对文字的辨别，但经过输入生成网络 *G* 后得到补全图 *S* 都能较好地恢复文字的缺失信息。虽然图 *S* 相比于原图 *P* 常呈现细节差异，但大部分情况下结构信息的补全较为合理。图 3(a)展示了在某些笔画整体缺失的情况下，仍能够补全缺失笔画。再者，通过对比图 *S* 和图 *R*，可见本文提出的两级补全方法能明显抑制补全区域的模糊情况，改善补全图片的清晰度。图 3(b)放大显示 3 个模糊像素被抑制的例子。手写文字补全问题非常具有挑战，图 3(c)展示了一些补全效果不佳的例子，由其中前 2 个的例子可见，“寞”字在缺失中间的“艹”和“曰”后，可供候选的补全文字较多，如“宾”或“实”字，此时补全结果可能和原始文字具有不同的类别。第 3，4 个例子展示了补全区域出现冗余笔画的情况。最后一个例子则展示了在粗补全模块结果非常模糊的情况下，本文的细补全模块无法使图片清晰化。这些问题值得进一步研究和解决。



(a) 笔画整体缺失情况下的文字补全效果

(b) 细补全模块抑制模糊像素例子

(c) 补全效果不佳的例子

图 3 补全效果展示

2.3　不同缺失区域的位置和大小对文字补全的影响

除了文字的多样书写风格外，缺失区域的位置和大小对文字补全的影响也很大。图 4 中，第 1 行展示不同缺失区域位置和大小的掩膜，接下来 4 行则展示了 4 个汉字样本在不同的掩膜作用下的缺失区域的补全效果。缺失区域从小到大依次设定为 30，40 和 50。如图所示，当缺失区域为 30×30 时，生成网络 *G* 能够较好地补全文字，但随着缺失区域扩大至 50×50，补全变得困难，例如“藹”字的“艹”字头缺失笔画段或是“啊”字在缺失“口”部件时均难以填补。其中的原因可能是，“藹”去掉“艹”字头后是“谒”字，“啊”字去掉“口”字旁后是“阿”字，即这类汉字缺失某些区域后并不是无意义的笔画集合，而是变成了另一个类别的汉字，此时带缺失区域的汉字在模型看来仍然是逼真且合理的汉字，因此模型难以对其进行补全。与之相反，如图 4 中的“爱”字，由于在数据集的 3 755 类中具有较少的相似类别，因此就算缺失区域达到 50×50 大小，补全效果仍然较好。另外，当缺失区域是在文字中风格特殊的部件或笔段的位置时，虽然补全后汉字的类别能够大致保持不变，但特殊风格被恢复的难度较大，例如图中“按”字中“扌”的连笔风格在缺失补全后丢失。

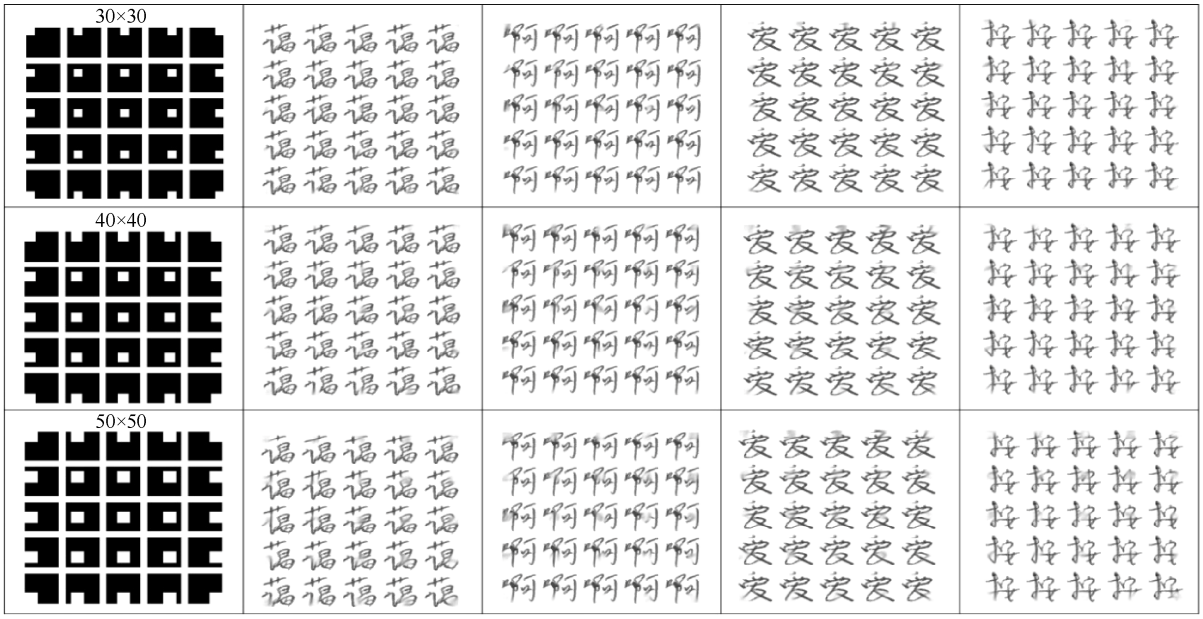


图 4　不同缺失区域位置和大小下的文字补全效果

3　结 束 语

本文针对大类别、小样本、多风格、未知语种的手写象形文字，采用全局和局部一致性保持的生成式对抗网络实现了带缺失区域的文字图像补全。针对结构图片的补全中遇到的模糊问题，本文提出两级补全模块，第一级模块偏重文字的结构补全，第二级模块专注文字补全的清晰化。通过大量的实验，验证了本文解决方案的有效性；同时，对不同大小和位置的缺失区域的实验分析可知，书写风格趋于大众化的、相似字较少的汉字在缺失补全后的效果更佳。

之后的研究工作可从以下 3 个方向展开：①在类别标签辅助下的文字补全研究。文字类别标签的辅助下，可以考虑使用类似 InfoGAN[13]加入分类器，通过分类器的类别监督信息，使生成网络趋向于输出更容易识别的补全文字图片，甚至将补全后文字的易识别性(即识别输出的置信度大小)作为衡量结构信息补全效果的评价指标。②在语料辅助下的文字补全研究。通过对某语种的大量语料进行统计分析，可以得到该语种的语义模型，一元语义模型就可以在单字补全中引入文字在日常使用的频繁程度；多元语义模型能够在篇幅级别的文字补全中考虑上下文信息，从而减少补全时模棱两可的情况，改善补全模糊问题。③书写风格保持的手写文字补全。一般的生成网络会尽可能给出大众化的补全结果，但实际应用中有时会期望补全后文字能够保持风格不变；解决方案可以考虑在判别网络中加

入书写风格一致性的监督信号。

参 考 文 献

[1] YANG C, LU X, LIN Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis [C]// 2016 The IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 3.

[2] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36(4): 1-4.

[3] 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.

[4] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention [EB/OL]. (2018-03-21). [2019-09-17]. https://arxiv.org/abs/1801.07892.

[5] HUY V V, DUONG N Q K, PEREZ P. Structural inpainting [EB/OL]. (2018-03-27). [2019-09-17]. https://arxiv.org/abs/1803.10348.

[6] PATRICIA V, JOAN S, COLOMA B. Semantic image inpainting through improved wasserstein generative adversarial networks [EB/OL]. (2018-12-03). [2019-09-17]. https://arxiv.org/abs/1812.01071.

[7] KALTWANG S, SAMANGOOEI S, REDFORD J, et al. Imagining the unseen: Learning a distribution over incomplete images with dense latent trees [EB/OL]. (2018-08-14). [2019-09-17]. https://arxiv.org/abs/1808.04745.

[8] LECUN Y. The MNIST database of handwritten digits [EB/OL]. (1998-11-01). [2019-09-17]. http://yann.lecun. com/exdb/mnist.

[9] NETZER Y, WANG T, COATES A, et al. Reading digits in natural images with unsupervised feature learning [C]//In Neural Information Processing Systems Workshop on Deep Learning and Unsupervised Feature Learning. New York: Curran Associates, 2011: 5.

[10] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: Curran Associates, 2014: 2672-2680.

[11] LIU C L, YIN F, WANG D H, et al. CASIA online and offline Chinese handwriting databases [C]//2011 IEEE International Conference on Document Analysis and Recognition. New York: IEEE Press, 2011: 37-41.

[12] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016: 2536-2544.

[13] CHEN X, DUAN Y, HOUTHOOFT R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets [C]//In Advances in Neural Information Processing Systems. New York: Curran Associates, 2016, 2172-2180.