

一种双判别器 GAN 的古彝文字符修复方法

陈善雄¹ 朱世宇^{1,2} 熊海灵¹ 赵富佳^{1,3} 王定旺^{1,3} 刘云³

摘要 在中国,彝文古籍文献日益流失而且损毁严重,由于通晓古彝文的研究人员缺乏,使得古籍恢复工作进展十分缓慢.人工智能在图像文本领域的应用,为古籍文献的自动修复提供可能.本文设计了一种双判别器生成对抗网络(Generative adversarial networks with dual discriminator, D2GAN),以还原古代彝族字符中的缺失部分. D2GAN 是在深度卷积生成对抗网络的基础上,增加一个古彝文筛选判别器.通过三个阶段的训练来迭代地优化古彝文字符生成网络,以获得古彝文字符的文字生成器.根据筛选判别器的损失结果优化 D2GAN 模型,并使用生成的字符恢复古彝文中丢失的笔画.实验结果表明,在字符残缺低于 1/3 的情况下,本文提出的方法可使文字笔画的修复率达到 77.3%,有效地加快了古彝文字符修复工作的进程.

关键词 彝文,生成式对抗网络,深度学习,梯度下降

引用格式 陈善雄,朱世宇,熊海灵,赵富佳,王定旺,刘云.一种双判别器 GAN 的古彝文字符修复方法.自动化学报,2022,48(3): 853-864

DOI 10.16383/j.aas.c190752

A Method of Inpainting Ancient Yi Characters Based on Dual Discriminator Generative Adversarial Networks

CHEN Shan-Xiong¹ ZHU Shi-Yu^{1,2} XIONG Hai-Ling¹ ZHAO Fu-Jia^{1,3} WANG Ding-Wang^{1,3} LIU Yun³

Abstract Ancient Yi literatures are increasingly lost and damaged seriously. Due to the lack of ancient Yi researchers, the inpainting of ancient books is progressing very slowly. The application of artificial intelligence is successful in the field of image and texts, so it is possible for automatic inpainting of ancient books. In this paper, a generative adversarial networks with dual discriminator (D2GAN) is designed to restore missing part in ancient Yi characters. The D2GAN is based on the deep convolution generating adversarial network, and adds a selection discriminator. The generation networks of ancient Yi character is optimized iteratively through three-stage training, and the character generator of ancient Yi is established. The loss of selection discriminator is used to optimize the model D2GAN iteratively. So, generated characters based D2GAN can restore the missing stroke in the ancient Yi characters. The experimental results show that the method proposed has an inpainting rate of 77.3% for incomplete characters that the incomplete part does not exceed one third. Therefore, our method is effectively for the inpainting of ancient book, it can accelerate the protection progress of ancient Yi literature.

Key words Yi characters, generative adversarial network (GAN), deep learning, gradient descent

Citation Chen Shan-Xiong, Zhu Shi-Yu, Xiong Hai-Ling, Zhao Fu-Jia, Wang Ding-Wang, Liu Yun. A method of inpainting ancient Yi characters based on dual discriminator generative adversarial networks. *Acta Automatica Sinica*, 2022, 48(3): 853-864

收稿日期 2019-10-30 录用日期 2020-04-16

Manuscript received October 30, 2019; accepted April 16, 2020
国家自然科学基金(61603310), 国家社会科学基金(19BYY171), 重庆市自然科学基金(cstc2019jcyj-msxm2550), 模式识别国家重点实验室开放课题(201900010), 中央高校基本科研业务费(XDJK2018B020), 重庆市教育委员会科学技术研究计划青年项目(KJQN201801901, KJQN201801902) 资助

Supported by National Natural Science Foundation of China (61603310), National Social Science Foundation of China (19BYY171), Natural Science Foundation of Chongqing (cstc2019jcyj-msxm2550), Open Projects Program of National Laboratory of Pattern Recognition (201900010), Fundamental Research Funds for the Central Universities of China (XDJK2018B020), and Youth Project of Science and Technology Research Program of Chongqing Education Commission (KJQN201801901, KJQN201801902)

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 西南大学计算机与信息科学学院 重庆 400715 2. 重庆工程学院计算机与物联网学院 重庆 400056 3. 贵州工程应用技术学

古彝文作为一种重要的少数民族文字,距今有八千多年历史,可与甲骨文、苏美尔文、埃及文、玛雅文、哈拉般文 5 种文字并列,是世界六大古文字之一,一直沿用至今,并在历史上留下了许多珍贵的典籍.这些用古彝文书写的典籍具有重要的历史意义和社会价值^[1-2],且种类繁多,内容广泛,涉及历史、文学、哲学、宗教、医药、历法等各个领域,具有较高的传承价值.而作为彝文古籍的载体,石刻、崖画、木牍和纸书由于年代久远,往往模糊不清,或者

院彝学研究院 毕节 551700

1. College of Computer and Information Science, Southwest University, Chongqing 400715 2. College of Computer and Internet of Things, Chongqing Institute of Engineering College, Chongqing 400056 3. Institute of Yi Studies, Guizhou University of Engineering Science, Bijie 551700

残缺不全,这给彝文古籍文献的保存和传播带来极大的困难.目前,从各地收藏单位收集到彝文古籍文献来看,纸质文件存在泛黄变脆,甚至出现残边、虫蛀等损毁问题;一些碑刻、木刻的古彝文也由于长期的侵蚀,字迹出现了模糊,腐蚀等情况^[3-5],如图1所示.得益于图像修复技术的发展,我们可以通过技术手段对残缺文字进行修补,还原文字的真实形态.

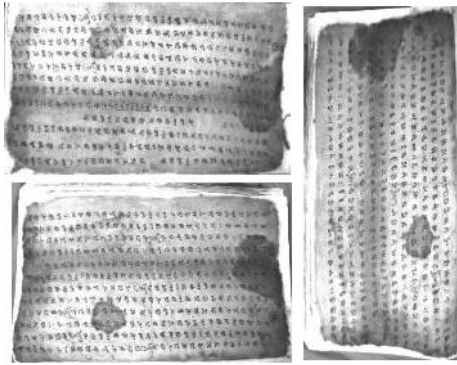


图1 彝文残卷

Fig.1 The incomplete literature of the ancient Yi

传统的文字图像修复,专业研究人员是通过语境信息和感知信息进行,即利用图像周围的像素以及综合标准文字中的各个特征要素来完成字符推演.但古彝文没有标准文字让计算机参照学习,手写古彝文存在不确定性问题.同时计算机也很难具备人的语境信息,对于文字的认知,需要很多的背景知识,这些是人类在一定环境中长期积累形成,很难系统地加以描述和组织.近年来,深度学习在图像语义修复、情感感知、模式识别以及特征分类等领域展现出令人振奋的前景^[6-7],特别在图像生成方面,表现出优越的性能.基于深度学习的图像生成算法相对于传统的基于结构和纹理的生成算法能够捕获更多图像的高级特征,常用于进行纹理合成和图像风格化迁移^[8-10].2014年由Goodfellow等提出的生成式对抗网络(Generative adversarial network, GAN)在图像生成领域取得了开创性进展^[11-12],在图像生成的过程中,生成式对抗网络相对于传统的编码-解码器而言能够更好地拟合数据,且速度较快,生成的样本更加锐利,但该方法也存在不足,如数据训练不稳定、网络自由不可控、训练崩溃等问题.2015年底,Radford等在深度卷积分类网络基础上^[13],提出深度卷积生成式对抗网络(Deep convolutional GAN, DCGAN)^[14].DCGAN融合卷积神经网络(Convolutional neural network, CNN)和GAN,通过设计独特的网络结构,使得训练更加稳定,这是首篇表明向量运算可以作为从生成器中学习的固有属性进行特征表达的论文.DCGAN的

成功使GAN拓展出多种应用^[15-20],如图像合成、风格迁移、超分辨率重构、图像修复以及图像转换等.

我们在深度卷积和生成对抗网络以往的工作之上^[13, 15, 21-24],针对彝文字符的特殊性,提出一个双判别器生成对抗网络模型用于古彝文字符修复.本文设计的双判别器网络是在DCGAN的基础上,增加一个筛选判别器模型,实现从手写彝文古籍文献中获取古彝文字符图像的概率分布,通过已获得的概率分布去预测待修复古彝文字符图像,根据预测图像完成修复任务.其基本流程为:首先使用古彝文字符对DCGAN网络进行训练,使该网络的古彝文生成器模型能够生成古彝文字符图像,然后建立一个筛选判别器模型,对生成出的古彝文字符进行比较,通过对生成的古彝文字符与待修复古彝文字符的差异建立损失函数,并对双判别器模型进行优化,最后能约束生成器模型,使其生成的古彝文字符不再是随机字符,而是与待修复的古彝文字符一致.本文使用手写古彝文字符图像数据集对上述方法上进行测试,对残缺1/3以下的字符能达到77.3%修复率.

1 双判别器生成式对抗网络模型

1.1 模型结构

本文提出了一种基于双判别器生成对抗网络(Generative adversarial networks with dual discriminator, D2GAN)的古彝文字符的修复方法.有别于传统图像修复,在文字修复中,如果已有字库的字符能覆盖待修复文字,则直接采用待修复字符图像和字库比对的方式.然而,古彝文存在字符库不全且手写体形式多样等问题,因而直接进行比对较为困难.针对古彝文缺乏完整字库样本,本文通过GAN生成器来实现一个动态的古彝文字库,并利用双判别器网络来获取目标古彝文字符,进一步用于残字修复.本文关键在于设计出能够生成出古彝文字符的网络,并加入筛选判别器形成双判别器的网络结构.该结构中包含2个判别器,一个用于判断生成器生成的字符是否属于古彝文;另一个用于判断待修复的古彝文字符和生成字符的相似性.D2GAN中筛选判别器的作用为判别生成器生成的古彝文字符是否是待修复的古彝文字符,如果生成字符与待修复字符一致,则两者进行融合,这样就可以实现残缺古彝文字符的修复.D2GAN模型结构如图2所示.

该网络由一个字符判别器、一个筛选判别器和一个字符生成器组成.网络的训练分为三个阶段:

第1阶段对古彝文字符判别器进行训练.将生成器生成的字符输入古彝文字符判别器网络,正向传播后,得到输出,本文期望输出为“假”;将真实

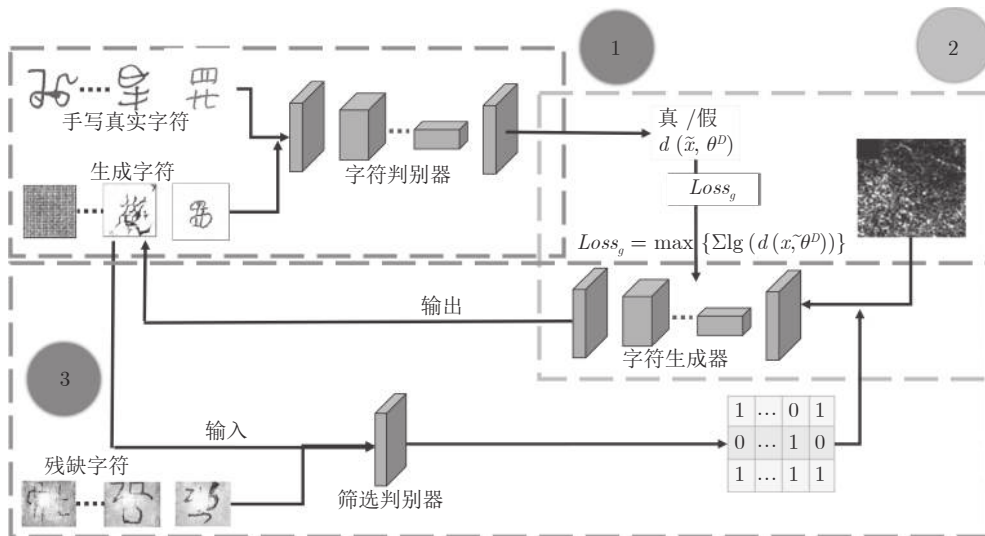


图 2 本文双判别器生成式对抗网络结构

Fig.2 Generative adversarial networks with double discriminator in the paper

的古彝文字符输入字符判别器网络, 正向传播后, 得到输出, 本文期望输出为“真”. 将 2 个输出结合起来建立损失函数, 进行反向传播, 优化古彝文字符判别器网络.

第 2 阶段对古彝文字符生成器进行训练. 将服从均匀分布的 100 维随机向量输入古彝文字符生成器网络, 通过正向传播, 得到输出, 输出数据的形状为 64×64 的矩阵. 将输出结果输入古彝文字符判别器网络, 在古彝文字符判别器网络内进行正向传播后, 得到输出, 本文期望输出为“真”. 通过输出结果建立损失函数, 进行反向传播, 优化古彝文字符生成器网络.

第 3 阶段, 重复第 1 阶段和第 2 阶段, 直到对古彝文字符生成器网络完成优化, 然后停止古彝文字符生成器的优化, 开始进行对筛选判别器进行训练. 将真实的古彝文字符和生成器生成的字符输入到筛选判别器网络, 通过网络的正向传播, 得到输出, 筛选判别器输出是一个 100 维的向量. 将该向量输入到古彝文字符生成器网络, 通过生成器正向传播, 得到输出, 本文期望该输出尽可能地接近需要修复的残缺彝文字符, 通过对二者一致性的比对, 建立损失函数, 通过在筛选判别器网络中反向传播, 对筛选判别器网络进行优化.

经过上述三个步骤, 完成针对古彝文字符图像的双判别器生成式对抗网络结构的训练, 得到一个稳定的模型. 在文献 [25] 中, 也提出了一个双判别器生成对抗网络 (D2GAN), 如图 3 所示. 在两个判别器固定的情况下, 生成器的学习将朝着同时优化 Kullback-Leibler (KL) 散度和反向 KL 散度的方向发展, 从而有助于避免模式崩溃. 该模型与本文差

异在于: 1) 文献 [25] 中, 两个判别器的模型结构, 完全一致, 作者改变的是两个判别器损失值的计算方法. 在本文的双判别器框架中, 两个判别器的模型结构相差甚远, 并且数据输入的方式也并不相同. 2) 双判别器模型的训练方法不一样. 文献 [25] 中的训练方法为生成器与两个判别器同步训练, 通过两个判别器, 来优化生成器, 同时通过生成器优化判别器. 本文的训练方法为生成器和第 1 个判别器同步训练, 相互优化, 在生成器优化结束后, 通过第 2 个判别器对生成器的生成结果进行筛选. 3) 增加的判别器用途不一样. 文献 [25] 中增加的判别器是为

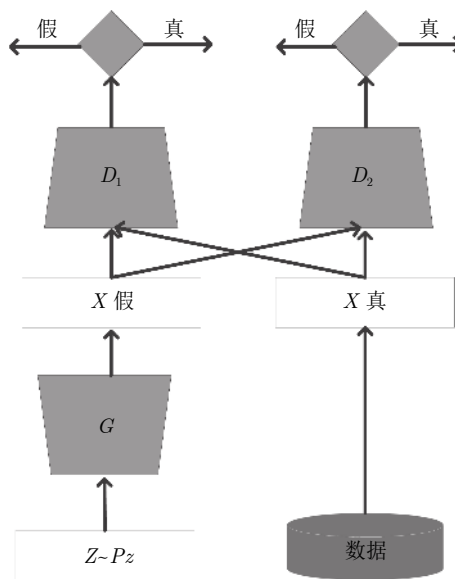


图 3 文献 [25] 中 D2GN 结构

Fig.3 The structure of D2GN in [25]

了能更好地优化生成器的模型参数. 本文增加的判别器, 是为了从生成器生成出的结果中寻找最优解, 并没有对生成器的参数再进行优化.

为了展示本文双判别器结构的细节, 下面首先对古彝文字符判别网络、筛选判别网络和古彝文字符生成网络进行详细说明.

1.2 古彝文字符生成网络

古彝文字符生成器主要用于获取古彝文字符的概率分布, 在极大似然概率生成网络上产生出彝文, 从而形成动态彝文字库. 古彝文字符生成器 (Generator, 下文简称网络 G) 的优化来自于彝文字符判别器 (Discriminator, 下文简称网络 $D1$), 两个网络之间相互博弈, 从而不断优化, 网络结构示意图如图 4 所示.

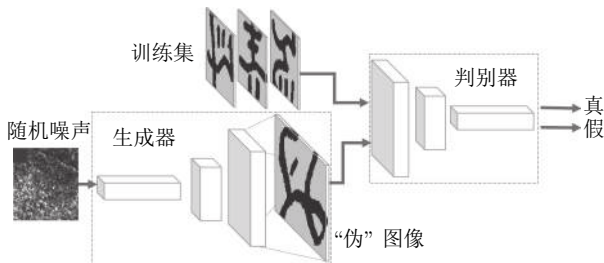


图 4 生成式对抗网络的结构图

Fig. 4 The structure of generative adversarial networks

彝文字符生成器用随机向量 z 作为网络 G 的输入, 将输出送入网络 $D1$, 通过网络 $D1$ 的正向传播获得生成效果, 根据生成效果优化网络 G , 其数学表达式为

$$G: \tilde{x} = g(z, \theta^G) \quad (1)$$

因为网络 $D1$ 可以区分彝文字符的“真”和“伪”, 所以网络 G 为让网络 $D1$ 将其判断为“真”, 会不断优化自身, 使生成的“伪”古彝文字符尽可能与“真实”古彝文字符一致. 通过网络 G 不断优化, 从而学习古彝文字符的本质特性, 刻画出古彝文字符的分布概率, 最终使得网络 G 生成出的数据与古彝文字符高度相似, 形成动态彝文字库. 其中 G 表示彝文字符生成器; θ^G 为待优化参数; $g(\cdot)$ 为需进一步优化的非线性映射函数; z 为 $g(\cdot)$ 的输入, 即 $-1 \sim 1$ 之间的双精度随机数, 是 100 维的向量; x 为真实的古彝文字符数据, \tilde{x} 为生成器输出的结果, 为 64×64 像素的图像, 且取值范围在 $-1 \sim 1$ 之间. 对网络 G 而言, 在彝文字符判别器固定时, 期望生成数据的分布特性尽最大可能与真实古彝文字符一致, 即判别器将生成图像都识别为真. 如式 (2) 所示, 因为期望判别器识别为真, 因此 $D(\tilde{x})$ 的输出越接近于 1

越好, 即 $\lg(D(\tilde{x}))$ 越大越好.

$$loss_g = \max \left\{ \sum \lg(d(\tilde{x}, \theta^D)) \right\} \quad (2)$$

1.3 古彝文字符判别网络

彝文字符判别器的作用是帮助彝文字符生成器优化, 通过不断调整自身的鉴别能力, 从而使生成器的能力也不断提高. 彝文字符判别器分别用彝文数据集中的数据和网络 G 的输出作为网络 $D1$ 的输入, 将判断结果作为网络 $D1$ 的输出. 其数学表达式为

$$D: y = d(x, \theta^D) \quad (3)$$

$$D: y = d(\tilde{x}, \theta^D) \quad (4)$$

其中, D 表示彝文字符判别器; θ^D 为待优化的参数; y 为 $d(\cdot)$ 的输出结果, 即将输入数据判断为真的概率, 且 $y \in [0, 1]$. 判别模型设计为将自然数据判断为真的概率, 以及将生成数据判断为假的概率要高. 通过网络 $D1$ 的正向传播获得判别结果, 根据结果对网络 $D1$ 进行优化, 其损失函数为

$$loss_d = \min \left\{ - \left[\sum \lg(d(x, \theta^D)) + \sum \lg(1 - d(\tilde{x}, \theta^D)) \right] \right\} \quad (5)$$

式 (5) 中对 $\lg(d(x, \theta^D))$ 的损失值取负的物理解释为将 x 判断为真的不确定性越小越好, 其最佳状态为 $d(x) = 1$; 而对 $\lg(1 - d(\tilde{x}, \theta^D))$ 的损失值取负的物理解释为将 \tilde{x} 判断为伪的不确定性越小越好, 即将 \tilde{x} 判断为伪的概率越大越好; 将所有判定的不确定性进行求和, 便得到熵. 根据熵对模型的参数 θ^D 进行优化. 在对判别模型 D 的参数进行更新时, 对于来自真实分布 P_{data} 的样本 x 而言, 因为期望都能够识别为真, 因此 $D(x)$ 的输出越接近于 1 越好, 即 $\lg(D(x))$ 越大越好; 对于通过噪声 z 生成的数据 $G(z)$ 而言, 因为期望都能够识别为假, 因此 $D(G(z))$ 越接近于 0 越好, 即 $\lg(1 - D(G(z)))$ 越大越好.

通过设计好的损失函数 (式 (2) 和式 (5)), 使用梯度下降法对 (θ^G, θ^D) 参数交替优化, 使网络 G 和网络 $D1$ 接近纳什均衡.

1.4 筛选判别器

完成古彝文动态字库的设计后, 针对古彝文字符缺损的修复问题可以描述为图像的比对问题. 通过对古彝文字符生成器的训练, 可以获取 100 维列向量的随机数和古彝文字符的映射关系, 但是生成器输出的古彝文字符是随机的, 要完成特定的彝文字符修复, 还需要建立起 100 维列向量与特定字符之间的映射. 假设待修复的图像为 A , 设置一个初始随机值 z 作为生成器 $G(z)$ 的输入, 正向传播后可

得到一幅图像 B' . 此时的图像 B' 与图像 A 可能毫无关联或相关性不大, 无法用图像 B' 完成图像 A 的修复. 筛选判别器的作用在于找到一个 \mathcal{H} , 使得生成器 $G(z)$ 的输出图像 B' 与图像 A 无限接近. 即使生成器自动生成的古彝文字符与需要修复的特定古彝文字符尽量接近.

本文通过两幅图像的欧氏距离 $d(A, B)$, 即

$$d(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (6)$$

建立起两幅图像间概率分布的差异, 根据差异设计出损失函数对筛选判别器进行优化. 但仅以此作为损失值对 z 进行优化还不够, 还需限定图像 B' 必须是一个古彝文字符, 本文将彝文字符判别器对图像 A 的判定结果, 也作为优化目标之一. 结合两个损失值就可以得到优化 z 的函数, 即

$$loss_z = \min \left\{ \sum \lg(1 - D(z)) + \sum \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \right\} \quad (7)$$

式 (6) 中, a 代表图像 A 中的像素值, b 代表图像 B' 中的像素值. 筛选判别器期望比较出更符合古彝文字符规律的字符图像, 因此 $D(z)$ 的输出越接近 1 越好, 即 $\lg(1 - D(z))$ 越小越好. 生成数据的分布特性尽最大可能与真实古彝文字符数据一致, 即判别器将生成图像都识别为真. 最小化 Lasso, 如式 (5) 所示. 因为期望判别器识别为真, 因此 $D(z)$ 的输出越接近 1 越好, 即 $\lg(1 - D(z))$ 越小越好. $loss_z$ 为凸函数, 通过梯度下降法^[26-27] 对 z 进行优化, 使得 z 无限接近于期望的 \mathcal{H} 以此得到图像 B , 从而完成古彝文字符修复.

2 模型结构

2.1 古彝文字符判别器模型

古彝文字符判别器模型的输入为字符图像, 通过判别器判断该图像是否为古彝文字符. 模型包括 1 个输入层、4 个卷积层和 1 个输出层. 其模型结构如图 5(a) 所示, 其中层与层之间的黑色连接圆, 代表卷积的方法和激活函数, 如图 5(b) 所示.

古彝文字符判别器由 4 层 CNN (不包含输入层) 组成, 图 5(a) 中第 1 层为输入层, 输入原始数据, 该数据源于古彝文字符图像, 大小为 64×64 像素, 因为古彝文字符是灰度图像, 因此将图像的 3 通道修正为单通道. 最后为输出层, 只有 1 个节点. 卷积

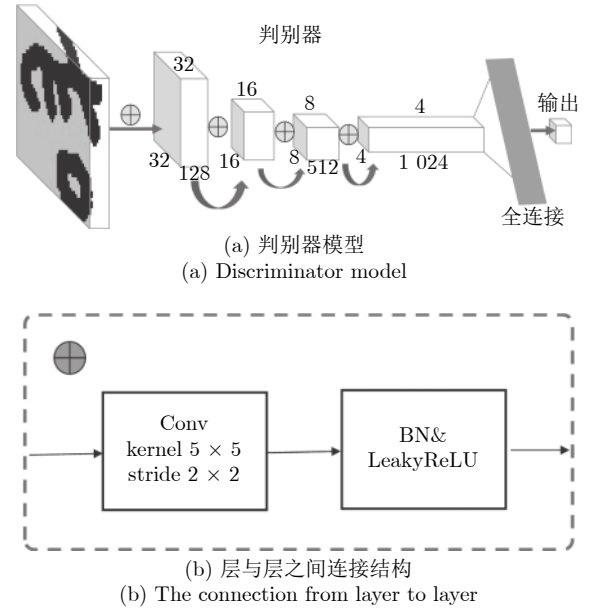


图 5 古彝文字符判别器模型详细结构

Fig.5 Detailed structure of the ancient Yi character discriminator model

层用 C 表示, 详细信息如表 1 所示. OUTPUT 层计算输入向量和权重向量之间的点积, 再加上一个偏置, 然后将其传递给 sigmoid 函数输出结果^[28]. 输出层的节点值表示是否为古彝文字符. 如果节点的值为 1, 则表示网络识别结果为古彝文字符, 0 则相反.

表 1 判别器模型参数表

层信息	卷积核个数	卷积核大小	步长	特征图大小	参数个数
C1层 (卷积层)	128	5×5	2	32×32	3328
C2层 (卷积层)	256	5×5	2	16×16	819456
C3层 (卷积层)	512	5×5	2	8×8	3277312
C4层 (卷积层)	1024	5×5	2	4×4	13108224
OUTPUT层 (输出层)				1×1	16385

2.2 古彝文字符生成器模型

古彝文字生成器模型由 4 层 CNN (不包含输入层) 组成, 图 6(a) 中第 1 层为输入层, 输入为服从均匀分布的 100 维向量的随机数. 最后为输出层, 输出数据为 $64 \times 64 \times 1$ 的矩阵, 希望通过训练, 该数据能表示古彝文字符图像, 其模型如图 6(a) 所示. 其中层与层之间的白色连接圆, 代表反卷积的方法和激活函数 (如图 6(b) 所示). 反卷积层用 DC 表示, 全连接层用 F 表示, 详细信息如表 2 所示.

2.3 古彝文字符筛选判别器模型

在筛选判别器训练及彝文字符修复阶段, 古彝

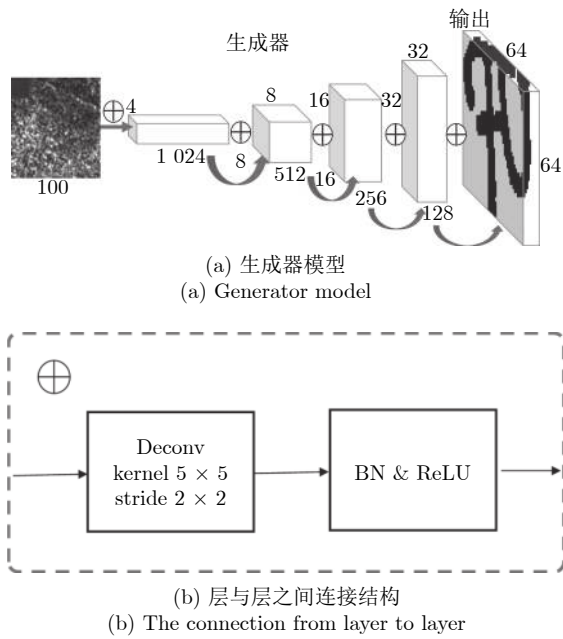


图 6 古彝文字符生成器模型详细结构

Fig.6 Detailed structure of the ancient Yi character generator model

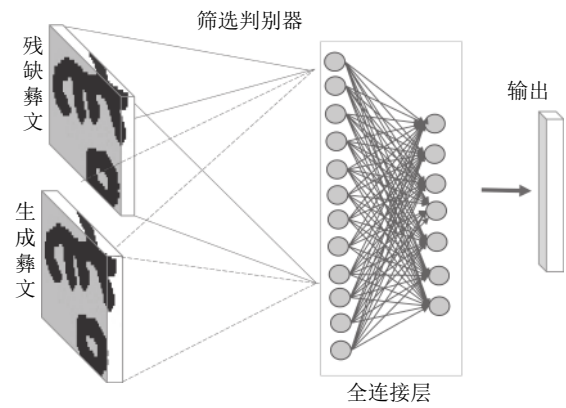
表 2 生成器模型参数表
Table 2 Parameter table of the generator model

层信息	卷积核个数	卷积核大小	步长	特征图大小	参数个数
C1层(卷积层)					1654784
C2层(卷积层)	512	5×5	2	8×8	13107712
C3层(卷积层)	256	5×5	2	16×16	3277056
C4层(卷积层)	128	5×5	2	32×32	819328
OUTPUT层(输出层)	1	5×5	2	64×64	3201

文生成器和古彝文判别器停止优化. 在没有约束的情况下, 生成器的生成效果, 只具有古彝文字符的分布特征, 而并不一定是真正的古彝文^[29-30]. 本文增加筛选判别器之后, 通过与待修复古彝文字残字图像做比较, 进一步对古彝文字符生成器进行约束, 使得生成的结果更符合古彝文的正常书写规范, 且更接近于待修复的古彝文. 其结构如图 7 所示.

筛选判别器的输入为残缺的字符和生成的字符, 仅有 1 个全连接层. 该全连接层神经网络的每层神经元权重的个数, $Param = (\text{输入数据维度} + 1) \times \text{神经元个数}$, 加 1 是因为每个神经元都有一个偏置值. 输入数据维度为 100, 该层使用了 $1 \times 64 \times 64$ 个节点, 所以参数数量为 $(100 + 1) \times 1 \times 100 = 10100$. 全连接层输出作为字符生成器的输入, 是一个 100 维的向量.

筛选判别器模型优化方法如下:

图 7 古彝文字符筛选判别器模型
Fig.7 Selecting discriminator model for ancient Yi character

1) 从古彝文数据集 B 中选取一个样本作为修复对象(如图 8). 因为该样本是完整的古彝文, 并无缺失. 为验证效果, 使用一个 20×20 的全 1 矩阵对图像的中间部分进行覆盖(如图 9). 为验证模型的鲁棒性, 除矩形外, 本文还使用了三角形和圆形等多种形状进行区域覆盖.



图 8 原始样本

Fig.8 The original sample



图 9 待修复古彝文

Fig.9 Ancient Yi character need to be restored

2) 将进行覆盖的字符(残缺字符)和生成器生成的字符作为输入, 用 1 个全连接网络构成筛选判别器的隐层, 同时, 把古彝文筛选判别器的输出 z , 作为古彝文生成器的输入, 通过正向传播可以得到输出 $G(z)$, 如图 10 所示.

3) 从图 8 ~ 10 可以看出, 此时古彝文字符生成器输出的 $G(z)$ 与待修复的古彝文残字(图 9)之间并无关联. 将 $G(z)$ 与残字进行对比, 根据式(6)得到损失值 $loss_z$.

4) 用 $loss_z$ 对筛选判别器进行优化.



图 10 通过生成器模型输出图像 $G(z)$

Fig.10 Output images $G(z)$ from the generator model

3 训练与实验

3.1 数据集与模型训练

实验样本来源于 37 万字的《西南彝志》中选取的 2142 个常用古彝文字符^[31], 并邀请彝族老师和学生进行临摹, 发放了 1200 份采集表 (如图 11 所示), 其中古彝文正体采集表 800 份、软笔风格采集表 200 份、硬笔风格采集表 200 份, 如图 12 所示, 共得到了 151200 个字体样本. 同时, 为了便于后期处理分析, 设计了相应的字体库 (如图 13 所示) 和古彝文输入法.



图 11 原始样本

Fig.11 The original sample

将古彝文字体库中的样本转换为 64×64 个像素点构成的图片, 每个像素点用 $0 \sim 255$ 灰度值表示, 对采集到的样本, 每个字按 7:3 的比例分为训练集 A 和测试集 B , 部分样本如图 14 所示.

实验环境: CPU Intel(R) core(7M)i7-7700, 3.6 GHz; 内存 DDR4, 8.00 GB; GPU NVIDIA GeForce RTX 2080 SUPER, 基础频率 1650 MHz, 加速频率 1815 MHz, 显存: GDDR6, 8 GB, 显存位宽 256 bit, 显存频率 15.5 GHz, 显存带宽 496 GB/s.

3.2 彝文字符生成器和彝文字符判别器训练过程

实验使用训练集 A 对模型进行训练, 每次训练以 32 个样本为单位, 进行批量训练, 过程为:

1) 从训练集 A 中抽取 32 个彝文数据, 将其作为彝文判别器的输入, 通过彝文判别器的正向传播,

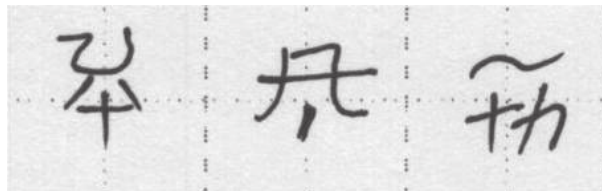


图 12 古彝文硬笔 (上) 和软笔 (下)

Fig.12 Ancient Yi hard pen (upper) and soft pen (down)



图 13 待修复古彝文

Fig.13 Ancient Yi character need to be restored



图 14 古彝文手写数据集样例

Fig.14 The handwritten sample of ancient Yi

得到 $d(x, \theta^D)$, 即彝文是否为“真”的概率, 将其代入式 (8), 得到损失值 $d1$

$$d1 = \sum \lg(d(x, \theta^D)) \tag{8}$$

2) 随机产生 32 个服从均匀分布的 100 维向量, 将其作为彝文生成模型的输入, 通过正向传播获得 32 个 $64 \times 64 \times 1$ 的矩阵, 将这批矩阵作为彝文判

别模型的输入, 通过正向传播也得到 $d(\tilde{x}, \theta^D)$, 将其代入式 (9), 得到损失值 $d2$

$$d2 = \sum \lg(d(\tilde{x}, \theta^D)) \quad (9)$$

3) 将 $d1$ 和 $d2$ 代入式 (5), 得到彝文判别器的损失值 $loss_d$, 通过 Adam^[32-33] 算法, 对彝文判别的参数进行优化.

4) 彝文字符生成器输出的 32 个图像通过彝文字符判别器得到 $d(\tilde{x}, \theta^D)$, 在作为彝文字符判别器的损失值 $d2$ 的同时也作为彝文字符生成器的损失值. 因为对彝文字符生成器的期望是所有生成出的数据, 彝文字符判别器都将其判别为真, 所以将 $d(\tilde{x}, \theta^D)$ 代入式 (2) 得到彝文生成器的损失值 $loss_g$, 用 Adam 算法对彝文字符生成模型进行优化.

重复一次上述的过程, 称为完成一次训练. 通过不断重复训练, 对彝文字符判别器和彝文字符生成器同时进行优化.

模型采用 tensorflow 框架提供的 Adam 随机梯度下降优化算法进行训练, 在训练时, 需要设置学习率 (Learning rate) 控制参数的更新速度. 该参数会极大地影响模型收敛速度, 过小会收敛太慢, 增加训练成本; 过大会导致参数在最优解附近振荡, 无法获取最优解^[34]. 本文对多个学习率下损失值的变化进行对比, 当训练次数为 2000 时, 学习率分别为 0.2, 0.02, 0.002 的判别器损失函数曲线如图 15 所示. 横轴代表训练次数, 纵轴代表损失值大小, 单位为 px (像素).

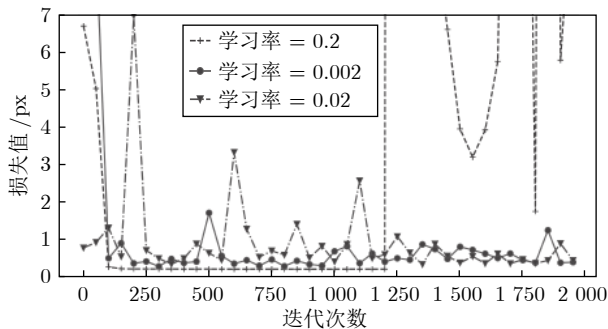


图 15 学习率 0.2, 0.02, 0.002 的损失值变化曲线
Fig. 15 The loss variation of the learning rate involving 0.2, 0.02 and 0.002

通过观察图 15 可以发现, 学习率在 0.2 时, 损失值在 1240 次之后就剧烈振荡, 这是明显的学习率过高, 导致无法收敛到最优点的情况. 将学习率 0.002 和 0.02 的损失值进行比较, 学习率 0.002 的损失值更接近于在 1 附近进行波动, 其预测值会更接近于 1, 因此本文以 0.002 为基础再次进行实验. 当训练次数为 2000 时, 学习率分别为

0.0002, 0.001, 0.002 的判别器损失函数曲线如图 16 所示. 通过观察图 16, 可以判断学习率为 0.001 时, 损失值的波动范围最接近于 1, 因此本文设置学习率为 0.001.

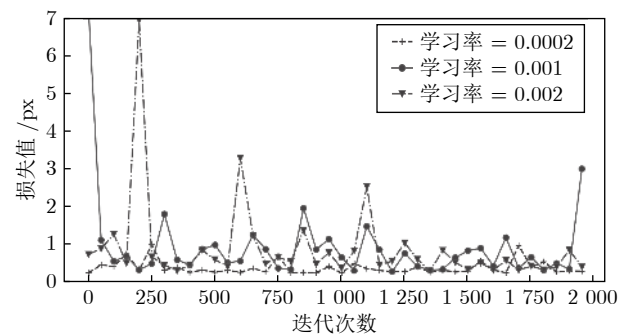


图 16 学习率为 0.0002, 0.001, 0.002 的损失值变化曲线
Fig. 16 The loss variation of the learning rate involving 0.0002, 0.001 and 0.002

训练过程中, 设置训练次数为 10000 次, 每次训练输出 25 幅 64×64 像素的生成图像. 在第 1 次、第 50 次、第 100 次、第 500 次、第 1000 次、第 2000 次、第 5000 次、第 10000 次、第 15000 次、第 20000 次训练完成后, 用生成器生成图像, 每次训练完成生成 25 幅图像, 共 250 幅图像, 如图 17 所示.

经彝文专家判断, 在训练超过 10000 次之后, 生成器生成的图像趋于稳定. 故采用经过 10000 次训练的生成器网络.

在训练过程中, 每 100 次记录一次彝文判别器和彝文生成器的损失值, 经过 10000 次训练之后, 损失值变化曲线如图 18 所示. 从图中可以看出, 彝文生成器模型和彝文判别器模型在训练过程中不断进行博弈. 当彝文判别器模型的损失值减小时, 判



图 17 不同训练次数下生成器生成图像
Fig. 17 The generator generates image under different training times

别器判定生成的彝文为假的概率增加. 而彝文生成器模型损失值减小时, 判别器能较大概率判定生成的彝文为真. 当两个模型的损失值交替上升时, 表示这两个模型在不断地进行博弈, 并交替地对模型参数进行优化.

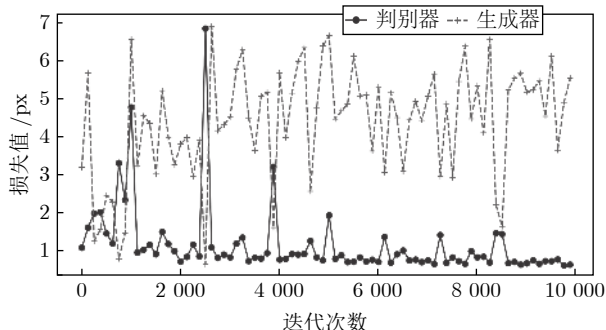


图 18 10000 次训练后损失值变化曲线
Fig.18 Loss curve after 10000 training

本文用随机产生的 1000 个 100 维向量作为完成 10000 次训练的生成器模型的输入, 通过模型正向传播, 得到得到 1000 个图像, 经彝文专家判断生成结果与真实彝文字符的形态非常接近. 部分数据如图 19 所示.



图 19 生成器输出图像
Fig.19 Output images by generator

3.3 筛选判别器的训练及彝文修复

根据第 2.3 节所述流程, 在对筛选判别器的训练过程中, 我们设置学习率为 0.001, 对筛选判别器模型进行 2000 次优化. 过程中 $loss_z$ 的变化曲线如图 20 所示, 横轴代表训练次数, 纵轴代表损失值大小, 单位为 px (像素单位). 可以看出在训练 750 次后, 损失值已经逐步下降到一个范围内, 即表示当前生成的字符与待修复的字符之间的差异也下降到一定范围. 在 2000 次优化后, 筛选判别器生成的 z' , 通过在古彝文字符生成器中正向传播得到 $G(z')$, 如图 21 所示, 从图中可以看出 $G(z')$ 与残字接近, 却又保留有残字所残缺的那部分. 将待修复古彝文

中缺失的部分, 用 $G(z')$ 进行填充, 得到修复后的图像, 如图 22 所示.

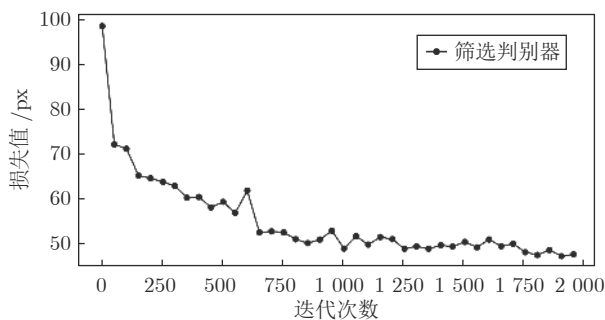


图 20 筛选判别器的训练过程中损失值变化曲线
Fig.20 The loss curve in process of the training of the selecting discriminator



图 21 训练得到 z' , 然后输入 z' 到生成器得到的图像
Fig.21 After training, z' is generated, and then input z' to the generator to get the image



图 22 修复后的图像
Fig.22 The restored image

上述实验重复 1000 次, 每一次随机从古彝文测试集 B 中抽取一个样本制作成待修复的古彝文, 然后按本文的方法进行修复, 对古彝文字符的修复率为 77.3%, 如表 3 所示.

表 3 中的评价标准如下: 修复后的图像与原图像字体形态一致为完全修复; 能够知道图像的古彝文字符是哪个古彝文字符, 但存在偏旁或部首的缺损为部分修复; 通过观察修复后的图像, 不能判断出是哪个古彝文字符为未完成修复. 图 23 为 1000 次实验中的部分数据, 从左至右分别为古彝文原始样本图像、生成出的待修复的古彝文图像、生成器生成出来的图像以及修复之后的效果图. 图 24 为多种残缺形状的修复效果.

本文使用上述模型对从彝文古籍中选取出的 10 个残缺的古彝文字符进行处理和修复, 其效果如图 25 所示. 从左至右分别为残缺的古彝文字符、通

表 3 古彝文字符修复比例

Table 3 Restoration proportion of ancient Yi characters

	个数	占比 (%)	时间消耗 (ms)
完全修复	523	52	52
部分修复	246	25	61
未完成修复	231	23	56
合计	1000	100	

过模型生成的用于修复的字符以及修复结果. 其具体修复方法为, 取模型生成的字符图像中某一区域的黑色像素, 叠加到待修复的图像中. 区域的位置, 根据原始古彝文字符的缺损位置来确定. 然而, 通过专家判别, 仍有 22.7% 是修复失败的. 部分样本如图 26 所示. 其原因在于, 残字缺失的部分, 其笔画较为复杂, 且该字符异体字较多, 而生成器网络结构对复杂笔画的适应性较弱, 因此对残缺部分的

重构效果较差.

4 结束语

本文采用深度卷积生成式对抗网络来获取古彝文字符的分布概率, 并且在网络结构之上增加一个筛选判别器, 形成一个双判别器对抗生成网络. 该模型实现从彝文古籍文献中获取古彝文字符图像的概率分布, 通过已获得的概率分布去预测待修复古彝文字符图像, 完成古彝文字符的修复. 这也是对古文字修复的积极尝试, 对保存和发扬民族文化进行了有益探索. 然而, 本文对筛选判别器的优化函数是图像间像素的差值, 其并不能很好地代表字符图像间概率分布的差异, 后续研究会尝试使用多种度量标准作为损失函数. 同时, 文字修复判断的可量化指标没有一个统一标准, 因此建立统一的修复标准也是未来研究的方向.

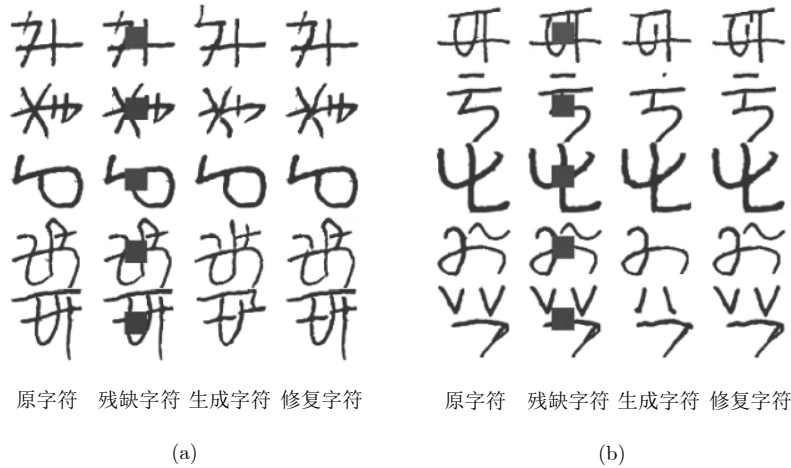


图 23 部分古彝文修复结果
Fig.23 The repair effect of some ancient Yi

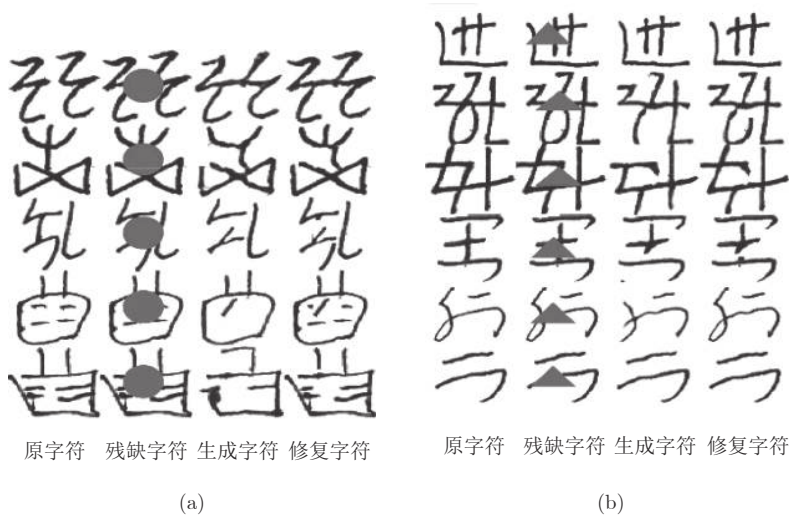


图 24 多形状残缺修复结果
Fig.24 The repair effect of ancient Yi character of multiple shape occlusion



图 25 彝文古籍文献中残缺字符修复效果

Fig. 25 The repair effect of incomplete characters in ancient Yi literature



图 26 古彝文残缺字符修复失败效果

Fig. 26 The failed repair effect of ancient Yi incomplete characters

References

- Wang Zheng-Xian. Brief introduction to Yiwen and gold stone language. *Guizhou Ethnic Studies*, 2002, **20**(2): 156-163 (王正贤. 彝文金石述略. 贵州民族研究, 2002, **20**(2): 156-163)
- Wu Xie, Lu Yu-Ping, Wang Ming-Gui. On the construction of ancient Yi coded character sets in Guizhou. *Journal of Chinese Information*, 2014, **28**(4): 153-158 (吴颢, 禄玉萍, 王明贵. 论贵州古彝文编码字符集构建. 中文信息学报, 2014, **28**(4): 153-158)
- Shi Jin-Bo. On the collation and research of historical literature of contemporary minority characters. *Qinghai Nationalities Research*, 2018, **29**(3): 108-115 (史金波. 当代少数民族文字历史文献整理与研究刍议. 青海民族研究, 2018, **29**(3): 108-115)
- Yunnan Sichuan Guizhou Guangxi Yi Language Cooperation Group. *Yunnan Sichuan Guizhou Guangxi Yi Dictionary*. Yunnan Nationalities Publishing House, 2001, **1**(1): 278-286 (滇川黔桂彝文协作组, 滇川黔桂彝文字典. 云南民族出版社, 2001, **1**(1): 278-286)
- Chen Shan-Xiong, Wang Xiao-Long, Han Xu, Liu Yun, Wang Ming-Gui. A recognition method of ancient Yi based on deep learning. *Journal of Zhejiang University (Science Edition)*, 2019, **46**(3): 261-269 (陈善雄, 王小龙, 韩旭, 刘云, 王明贵. 一种基于深度学习的古彝文识别方法. 浙江大学学报(理学版), 2019, **46**(3): 261-269)
- Bouillon M, Ingold R, Liwicki M. Grayification: A meaningful grayscale conversion to improve handwritten historical documents analysis. *Pattern Recognition Letters*, 2018, **121**(15): 46-51
- Cao X Y, Zhou F, Lin X, Meng D Y, Xu Z B, Paisley J. Hyperspectral image classification with markov random fields and a convolutional neural network. *IEEE Transactions on Image Processing*, 2018, **27**(5): 2354-2367
- Danilla C, Persello C, Tolpekin V, Bergado J R. Classification of multitemporal sar images using convolutional neural networks and Markov random fields. In: Proceedings of the 2017 Geoscience and Remote Sensing Symposium (IGARSS), IEEE International Symposium. Fort Worth, TX, USA: IEEE, 2017. 2231-2234
- Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017(1): 1510-1519
- Cireřan D C, Meier U, Schmidhuber J. Transfer learning for Latin and Chinese characters with deep neural networks. In: Proceedings of the 2012 International Joint Conference on Neural Networks. Brisbane, QLD, Australia: IEEE, 2012. 1-6
- Goodfellow L, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Kuching, Malaysia: MIT Press, 2014, 2672-2680
- Chen S M, Wang W J, Xia B H, You X G, Peng Q M, Cao Z H, et al. CDE-GAN: Cooperative dual evolution-based generative adversarial network. *IEEE Transactions on Evolutionary Computation*, 2021, **25**(5): 986-1000
- Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the 4th International Conference on Learning Representations (ICLR). San Juan, Puerto Rico: OpenReview.net, 2016.
- Zhao Shu-Yang, Li Jian-Wu. Generative adversarial network for generating low-rank images. *Acta Automatica Sinica*, 2018, **44**(5): 829-839 (赵树阳, 李建武. 基于生成对抗网络的低秩图像生成方法. 自动化学报, 2018, **44**(5): 829-839)
- Wei P, Li H, Hu P. Inverse discriminative networks for handwritten signature verification. In: Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 5764-5772
- Li X W, Shen H Q, Yu M, Wei X, Han J, Zhu J L, et al. Blind image inpainting using pyramid GAN on thyroid ultrasound images. In: Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). San Diego, CA, USA: IEEE, 2019. 678-683
- Chang J, Gu Y J, Zhang Y, Wang Y F. Chinese handwriting imitation with hierarchical generative adversarial network. In: Proceedings of the 2018 British Machine Vision Conference (BMVC). Newcastle, UK: BMVA, 2018.
- Chang B, Zhang Q, Pan S Y, Meng L L. Generating handwritten Chinese characters using cyclegan. In: Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV, USA: IEEE, 2018. 199-207
- Yeh R A, Chen C, Lim T Y, Schwing A G, Hasegawa-Johnson M, Do M N. Semantic image inpainting with deep generative models. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 6882-6890
- Fu Xiao, Shen Yuan-Tong, Li Hong-Wei, Cheng Xiao-Mei. A semi-supervised encoder generative adversarial networks model for image classification. *Acta Automatica Sinica*, 2020, **46**(3): 531-539 (付晓, 沈远彤, 李宏伟, 程晓梅. 基于半监督编码器生成对抗网络的图像分类模型. 自动化学报, 2020, **46**(3): 531-539)
- Gao Y P, Gao L, Li X Y. A generative adversarial network based deep learning method for low-quality defect image reconstruction and recognition. *IEEE Transactions on Industrial Informatics*, 2021, **17**(5): 3231-3240
- You C Y, Cong W X, Vannier M W, Saha P K, Hoffman E A, Wang G, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging*, 2020, **39**(1): 188-203
- Liu M, Ding Y K, Xia M, Liu X, Ding E R, Zuo W M, Weng S L. Stgan: A unified selective transfer network for arbitrary image attribute editing. In: Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 3668-3677
- Zhong Z L, Li J, Clausi D A, Wong A. Generative adversarial networks and conditional random fields for hyperspectral image classification. *IEEE Transactions on Cybernetics*, 2020, **50**(7): 3318-3329
- Nguyen T D, Le T, Vu H T, Phung D. Dual discriminator generative adversarial nets. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). Long Beach, CA, USA: Curran Associates, 2017. 2671-2681

- 26 Memisevic R. Gradient-based learning of the 2011 higher-order image features. In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011, 1591–1598
- 27 Zhang S, Wang M, Xiong J, Liu S, Chen P Y. Improved linear convergence of training CNNs with generalizability guarantees: A one-hidden-layer case. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(6): 2622–2635
- 28 Shi Y, Davaslioglu K, Sagduyu Y E. Generative adversarial network in the air: deep adversarial learning for wireless signal spoofing. *IEEE Transactions on Cognitive Communications and Networking*, 2021, **7**(1): 294–303
- 29 Chang J, Scherer S. Learning representations of emotional speech with deep convolutional generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Acoustic, Speech and Signal Processing. New Orleans, LA USA: IEEE, 2017. 2746–2750
- 30 Meng Y Y, Kong D Q, Zhu Z F, Zhao Y. From night to day: GANs based low quality image enhancement. *Neural Processing Letters*, 2019, **50**(1): 799–814
- 31 Ma Jin-Wei. On the Origin and Development of Yi Language [Ph.D. dissertation]. Southwest University, China, 2010. (马锦卫, 彝文起源及其发展考论 [博士学位论文], 西南大学, 中国, 2010.)
- 32 Arcos-García, álvaro, Álvarez-García, Juan A, Soria-Morillo L M. Deep neural network for traffic sign recognition systems: an analysis of spatial transformers and stochastic optimisation methods. *Neural Networks*, 2018, 99: 158–165
- 33 Shin H C, Roth H R, Gao M C, Lu L, Xu Z J, Nogues I, Yao J H, Mollura D, Summers R M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 2016, **35**(5): 1285–1298
- 34 Chen S X, Han X, Gao W Z, Liu X X, Mo B F. A classification method of oracle materials based on local convolutional neural network framework. *IEEE Computer Graphics and Applications*, 2020, **40**(3): 32–44



陈善雄 西南大学计算机与信息科学学院教授。2013 年获得重庆大学计算机科学与技术博士学位。主要研究方向为数据挖掘, 模式识别, 压缩感知。
E-mail: csxpml@163.com

(**CHEN Shan-Xiong** Professor at the College of Computer and Information Science, Southwest University. He received his Ph.D. degree in computer science from Chongqing University in 2013. His research interest covers data mining, pattern recognition, and compressed sensing.)



朱世宇 重庆工程学院计算机与物联网学院博士研究生。2012 年获得重庆大学电子与通信工程专业硕士学位。主要研究方向为计算机视觉和自然语言处理。本文通信作者。
E-mail: herb_711@icloud.com

(**ZHU Shi-Yu** Ph.D. candidate at the College of Computer and Internet of Things, Chongqing Institute of Engineering College. He received his master degree in electronics and communication engineering from Chongqing University in 2012. His research interest covers machine vision and natural language processing (NLP). Corresponding author of this paper.)



熊海灵 西南大学商学院教授。2007 年获得西南大学土壤学博士学位。主要研究方向为农业信息化与数字农业, 农业生态与环境过程计算机模拟。
E-mail: xionghl@swu.edu.cn

(**XIONG Hai-Ling** Professor at the Business College, Southwest University. He received his Ph.D. degree in soil science from Southwest University in 2007. His research interest covers agricultural informatization and digital agriculture, computer simulation of agricultural ecology and environmental process.)



赵富佳 西南大学计算机与信息科学学院硕士研究生。2017 年获得盐城工学院计算机科学与技术学士学位。主要研究方向为模式识别, 图像处理, 深度学习。
E-mail: zfj_ai_python@163.com

(**ZHAO Fu-Jia** Master student at the School of Computer and Information Science, Southwest University. He received his bachelor degree in computer science and technology from Yancheng Institute of Technology in 2017. His research interest covers pattern recognition, image processing, and deep learning.)



王定旺 西南大学计算机与信息科学学院软件学院硕士研究生。2017 年获得重庆理工大学计算机科学与工程学院学士学位。主要研究方向为模式识别, 图像处理, 深度学习。
E-mail: 18996632565@163.com

(**WANG Ding-Wang** Master student at the School of Computer and Information Science, Southwest University. He received his bachelor degree in computer science and engineering from Chongqing University of Technology in 2017. His research interest covers pattern recognition, image processing, and deep learning.)



刘云 贵州工程应用技术学院彝学研究院副研究员。主要研究方向为彝文古籍整理, 彝文信息化。
E-mail: ly_3692022@163.com

(**LIU Yun** Associate researcher at the Resrarch Institute of Yi Nationality Studies, Guizhou University of Engineering Science. His research interest covers arrangement of Yi ancient books, and Yi language informatization.)