# 基于生成对抗网络的风格化书法图像生成

**王晓红** [1]，**卢辉** [2]，**麻祥才** [2]
（1.上海理工大学，上海 200093；2.上海出版印刷高等专科学校，上海 200093）

**摘要：目的** 为修复书法图像中的残缺字体，提出一种提取书法图像特征并自动生成风格化书法图像的方法。**方法** 首先针对书法作品的灰度图像利用变分自编码器提取字体的形状特征，同时将书法图像转换至 Lab 颜色空间中，通过四层卷积神经网络模型对 L 通道进行深度学习，提取书法字体中的风格特征；然后将风格特征作为条件输入，与形状特征一起输入生成对抗网络的生成器中进行联合训练，使生成的字体带有特定风格。研究过程中构建一个包含不同风格书法字体的中国书法字体生成数据集（CCGD-2019）用作模型训练。**结果** 提出了一种基于变分自编码与生成对抗网络的书法字体图像生成模型，能够从标准字体或随机噪声自动生成风格化书法字体图像。**结论** 人眼主观评价及 Fréchet 初始距离计算结果表明，生成字体的识别率和视觉质量均达到了令人满意的效果。

**关键词：**书法图像；生成对抗网络；形状特征；风格特征；颜色空间转换
**中图分类号：**TS206 **文献标识码：**A **文章编号：**1001-3563(2020)11-0246-08
**DOI：**10.19554/j.cnki.1001-3563.2020.11.036

## Generation of Stylized Calligraphic Image Based on Generative Adversarial Network

*WANG Xiao-hong*[1], *LU Hui*[2], *MA Xiang-cai*[2]
(1.University of Shanghai for Science and Technology, Shanghai 200093, China;
2.Shanghai Publishing and Printing College, Shanghai 200093, China)

**ABSTRACT:** The work aims to propose a method of extracting calligraphic image features and automatically generating stylized calligraphic image. Firstly, the variational auto-encoder was used to extract shape information of the character for the grayscale image of calligraphic works. At the same time, the calligraphic image was converted into the Lab color space and the stylistic feature was extracted through 4-layers convolutional neural network in L channel. Then, the stylistic feature was input to generator as conditions. Finally, the stylistic feature and shape information were used for joint training in generative adversarial network, which could generate Chinese characters with specific style. In addition, Calligraphy Character Generation Dataset (CCGD-2019) which contained several types of calligraphy character was constructed for model training in the process of experiment. A calligraphic character image generation model based on variational auto-encoding and generative adversarial network was proposed, which could automatically generate stylized calligraphic character images from standard characters or random noise. The results of subjective evaluation and Fréchet Inception Distance indicate that the recognition rate and the visual effect of generated characters reach the satisfactory effects.

**KEY WORDS:** calligraphic character; GAN; shape information; stylistic feature; color space conversion

**收稿日期：**2019-10-22
**基金项目：**上海市教育发展基金会和上海市教育委员会“晨光计划”（18CGB09）
**作者简介：**王晓红（1971—），博士，上海理工大学教授，主要研究方向为数字图像处理与色彩管理。

中国书法源远流长，博大精深，是中华文明的灿烂瑰宝，也是中华民族精神追求的象征。随着越来越多的书法作品数字化，借助计算机辅助技术对书法字体进行研究不仅有助于人们对于书法的理解与学习，对于传统书法作品的保护工作也具有十分重要的意义。

当前大多数关于书法字体的研究都集中于书法风格的识别，利用书法图像的结构性特点和笔画顺序，检测书法作品图像中文字区域，提取书法图像中的笔画或字体进行识别。雷国荣等[1]针对古代碑刻书法图像，提出基于 L0 范数引导滤波和自适应流行滤波器的清晰化方法，在去除噪声的同时保护了汉字笔画的完整性和结构性。许鹏飞[2]利用基于双阈值的引导滤波，准确地提取出书法作品中的印章信息。张福成等[3]提出基于神经网络的书法风格识别方法，通过对 LeNets-5 模型进行改进，大大提高了对楷书书法作品的识别率。章夏芬等[4]根据样本字特征动态改变剪枝范围的自适应匹配，利用轮廓形状相似性算法实现了书法字体的自适应匹配和检索。任晓文等[5]针对手写汉字图像识别率受随机噪声影响的问题，提出了一种基于深度学习与抑制噪声相结合的新算法，在有效处理手写汉字字符图片上随机噪声的同时又可以正确识别手写汉字。

上述研究在书法字体识别方面取得了良好的效果，但是都将书法作品看作自然图像，结合边缘提取和图像分割等先验知识进行图像增强，忽略了书法字体中的潜在特征。ZHENG 等[6]通过颜色空间转换和两阶段引导滤波策略相结合的方式，模拟人眼视觉感知，在提取汉字书法形状特征的同时发掘隐藏在灰度变化中的情感信息。张巍等[7]利用条件生成对抗网络(CGAN)的对抗学习机制直接分割笔画，有效提取出精确的笔画并得到高层语义特征，便于后期质量评价。在书法图像的生成方面，Lyu 等[8]将书法字体的生成看作端到端图像翻译问题，将对抗损失函数添加到自编码网络中，实现了从标准字体生成特定的书法字体，但生成的图像较为模糊，且无法适用于形变较大或较为复杂的字体生成。任健[9]使用基于 GAN 的 pix-to-pix 网络模型，成对输入 2 种不同风格的书法文字进行训练，实现了一对多的书法文字风格迁移，该方法针对书法字体特点进行了模型框架的调整和网络超参数优化，但为了消除冗余噪声而增加均方根误差作为目标函数，导致生成图像比较模糊，对于书法图像的颜色和纹理信息也缺少足够的重视。总之，书法字体生成是一个充满挑战性的问题，书法字体不仅包含各种不同的风格，更蕴含着书写者自身的情感。不同的书法字体，不同的书写者，甚至是同一书写者在不同时期所写的文字风格都有所不同，隐藏特征提取较为困难，书法字体的生成不仅要求需要重现字体风格，也要考虑到汉字书法浓墨重彩之间隐藏的情感信息，做到“形神兼备”。

文中通过对中国古代书法名家作品的电子稿进行分割整理，建立中国书法字体生成数据库，以此作为数据基础，提出一种基于变分自编码和生成对抗网络相结合（VAEGAN）的中国书法汉字图像生成模型。在生成书法汉字形状的同时，通过颜色空间转换，学习书法图像的情感特征，实现风格化的书法汉字图像生成。经过人眼主观评价实验证明，生成汉字图像不仅能够被准确识别，而且具有较好的视觉质量。

# 1　书法图像处理与生成对抗网络的基本原理

## 1.1　书法情感信息与颜色空间转换

书法的艺术美感由形状特征和情感特征等方面组成。其中，形状特征可以通过边缘识别、轮廓提取等图像处理方式快速获取，情感特征的提取则相对较为困难。由于书法图像以“白纸黑字”等灰度信息为主，彩色信息较少，书法图像的情感特征往往通过灰度变化来进行反映。如“浓墨重彩”和“轻描淡写”这 2 种不同的书写风格，就反映了书写者不同的性格与情感。

数字图像一般都处于 RGB 颜色模式下，RGB 三通道之间存在强相关性，灰度提取时较为不便。针对这种情况，文中在学习汉字外形特征的同时，将图像从 RGB 颜色空间转换到 Lab 颜色空间。Lab 颜色空间是一种设备无关的均匀颜色空间，由 $L$（亮度）、$a$（从绿色到红色的分量）、$b$（从蓝色到黄色的分量）3 个分量组成，具有感知均匀的特点，不仅比 RGB 更符合人类视觉特点，而且也更加适用于书法图像的特征提取。将书法图像转换至 Lab 颜色空间后，书法图像信息集中在 $L$ 通道内，用 $L$ 通道的信息代替图像灰度信息。$L$ 值较高的部分对应“轻描淡写”的书写方式，书写时情感一般较为轻松欢快，反之，$L$ 值较低则意味着书写时较为缓慢用力，代表情感较为激烈沉重。通过在 $L$ 通道内进行训练学习书法图像中的情感特征，在保证学习效果的同时加快了运行效率。

## 1.2　变分自编码器

自编码器（AE）起初是一种用于数据压缩的神经网络模型，其基本结构是一个多层感知机的神经网络，从输入到输出有多个隐层，输入层与输出层拥有相同的节点数量，中间编码层的节点小于输入和输出层的节点数量以实现数据压缩，输入层到编码层的网络部分成为编码器，编码层到输出层的网络部分成为解码器。训练时将编码器和解码器绑定在一起，将数

据本身作为其标签进行无监督学习。

自编码器能在输出层很好地还原输入层的数据，但对于训练集中未出现的数据效果很差。为增强网络的适用性，变分自编码器[10]（VAE）在普通自编码器的基础上添加了限制条件，要求编码器产生的隐含向量遵循某种特定分布（如高斯分布），使编码器能够真正读懂训练数据的潜在规律。变分自编码器的编码阶段输出结果从一个变成 2 个（均值向量和标准差向量），共同组成一个隐含变量模型。这样模型从训练数据中学习的不再是一个确定的函数，而是一个概率分布，这样便能够对任务的输入编码产生有相关意义的输出。

### 1.3 生成对抗网络

生成对抗网络模型[11]（Generative Adversarial Network，GAN）由 2 个子模型生成器 $G(\theta|z)$和判别器 $D(x,\theta)$组成，其中 $z$ 是随机噪声，$x$ 是真实数据，生成器和判别器均为定义好的多层结构模型。生成器的训练目标是生成判别器无法分辨真伪的数据，判别器的训练目标是能够分辨真实数据和来自生成器的生成数据。

GAN 自 Goodfellow 于 2014 年提出后，成为深度学习领域最受关注的热点之一，至今已产生上百种变体，其中以 cGAN[12]，DCGAN[13]，WGAN[14]等几种优化方式最为经典。cGAN 通过给 GAN 添加限制条件指导数据的生成过程，使得生成的数据具有特定性质的样本，从而解决了 GAN 生成样本不受控制的问题。DCGAN 对 GAN 的网络结构和参数进行了优化和调整，极大地提升了 GAN 训练时的稳定性及生成结果质量。WGAN 针对 GAN 中交叉熵的损失函数不适合衡量具有不相交部分的分布之间的距离的缺陷，使用 Wassertein 距离衡量生成数据和真实数据分布之间的距离，使得 WGAN 有较好的线性梯度，解决了训练不稳定的问题。

## 2 基于生成对抗网络的书法字体生成

### 2.1 数据集的建立与预处理

目前还没有关于书法字体通用的数据集，在手写汉字的识别领域，大多数研究采用的数据集是中科院自动化所构建的联机手写汉字样本库[15]（CASIA-HWDB1.1），该数据集样本由 1020 位书写者手写，包含大约 390 万个 7356 个类的样本（7185 个汉字和 171 个符号）。由于书写者汉字书写水平参差不齐，所以该数据集主要应用于手写汉字的识别领域，不适用于书法字体的高质量生成。在文献[8]中，研究者建立了自己的数据集（CCS-4）用于书法字体的生成，包含中国古代 4 位书法名家的手写字体共计 7000 幅图像。该数据集包含的书法类型和样本量相对较少，而且暂时并没有开源，文中在进行研究时，以汉字标签为分类依据，建立了中国书法字体生成数据集（CCGD-2019），应用于书法图像的生成，建立流程见图 1。

为了更好地进行书法字体的高质量自动生成，文中在实验前通过网络获取了从秦汉时期到晚清明国以来书法名家的 7227 幅书法作品电子图像。在书法图像中，无论是碑帖书法还是纸质书法，在保存过程中不可避免地容易被损坏，出现斑驳的印记，同时在书法图像数字化的过程中，也会对书法图像质量产生影响，此外，书法作品中还包含着书写者和品鉴者的印章，对书法字体的提取产生干扰。由此，在对书法图像进行分割之前，需要对图像进行预处理，按照文献[2]中的方法提取并去除了书法图像中的印章信息，然后对书法图像中的随机噪声进行了去除，获得了更加清晰和完整的书法图像。

书法图像的文字排列较为规则和整齐，文中采用投影法获取图像内文字分布信息，对图像中的文字进行自适应分割，获取大量单个汉字书法字体图像。然后对单个书法字体图像进行裁剪和规范化处理，最终处理后的图像尺寸为 256×256。对于少数自动化处理出现偏差的字体，人工进行修正。通过对处理后的标准单个字体图像进行归纳整理，形成了标准字体图片数据集，最后对数据集进行图片标注，得到最终中国书法字体生成数据集。数据集中包含 40 000 张书法图像（100 个汉字，每个汉字 400 张）。

### 2.2 模型介绍

如图 2 所示，字体生成框架基于 VAE-GAN[16]模型，将 VAE 中的解码器作为 GAN 中的生成器，从而将 VAE 和 GAN 联系起来。情感提取框架通过颜色空间转换和卷积神经网络，对书法图像中的灰度信息进行深度训练，输出结果作为风格特征指标添加到字体生成框架中编码器生成的潜在变量中，实现特定风格的字体生成，也有助于生成的书法字体图像更加符合人眼感知。

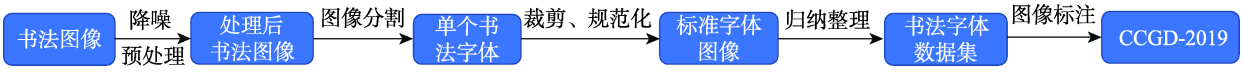


图 1 CCGD-2019 建立流程

Fig.1 Construction flowchart of CCGD-2019

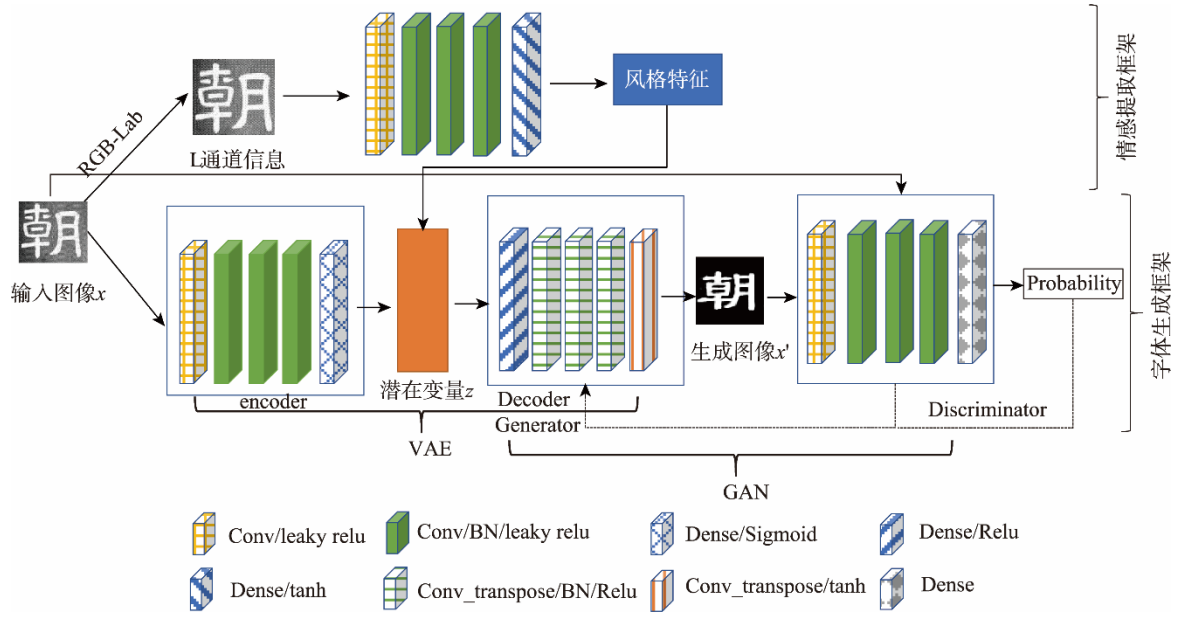


图 2　模型训练网络框架

Fig.2 Framework of model training network

### 2.2.1　字体生成框架

字体生成框架由编码器、解码器（生成器）、判别器等 3 个子网络构成，编码器将输入图像 $x$ 生成一个由均值向量和标准差向量构成的 100 维隐含变量，与情感提取框架生成的向量特征融合成一个整体的潜在变量 $z$，作为解码器的输入。解码器接收 $z$ 后，通过四层反卷积网络生成书法图像 $x'$，判别器接受真实图像与输入图像进行联合训练，输出生成图像为真的概率，用于指导下一次生成器的迭代训练过程。

1）编码器。图像编码器将 256×256×3 的图像作为输入，输出一个 100 维的向量作为字体图像的形状特征。编码器网络由四层卷积层和一层全连接层共同组成，卷积层图像通道数分别为{32，64，128，256}，核尺寸分别为{5，5，5，5}，步长分别为{2，2，2，2}，边缘部分通过等值填充。卷积层后除第 1 层外全部添加块归一化（BN 层）进行规范，使用 Leaky Relu 函数（$\alpha$ 取 0.2）进行激活，再利用 dropout 函数随机选取 80%的数据进入下一层训练以防止过拟合。全连接层使用 sigmoid 函数进行激活，最后输出一个 50 维的均值向量和 50 维的标准差向量组成的一个隐含变量。

2）解码器（生成器）。生成器网络由一层全连接层和四层反卷积网络组成，卷积层通道数分别为{256，128，64，3}，核尺寸分别为{5，5，5，5}，步长分别为{2，2，2，2}，边缘部分通过等值填充。全连接层使用 tanh 函数进行激活，卷积层除最后一层外全部使用 BN 层将特征层的输出归一化到一起，使用 Relu 函数进行激活，加速训练，用 dropout 函数随机选取 80%的数据进入下一层训练。生成器接受的随机变量来自编码器输出的隐藏变量和情感提取网络输出的特征信息，输出 256×25×3 的书法图像。

3）判别器。判别器网络与编码器网络具有相似的框架结构，也是由四层卷积层和一层全连接层构成。但在输入和输出方面有所区别，判别器网络中输入真实图像 $x$ 和生成器输出的生成图像 $x'$进行联合训练，经过 4 层卷积层处理后，最终经过全连接层输出判别结果，输出的概率值反馈至生成器网络中指导下一次训练。

### 2.2.2　情感提取框架

在情感提取框架中，首先将图像从 RGB 颜色空间转换至 Lab 颜色空间中，提取 $L$ 通道的信息作为卷积神经网络的输入。卷积神经网络的框架与判别器网络结构类似，但其输入仅为图像 Lab 颜色空间中的 $L$ 通道信息。经过卷积处理后，通过全连接层并用 tanh 函数进行激活，输出一个 20 维的潜在向量，作为书法图像的风格特征，输入到字体生成框架的生成器中作为条件输入，用以指导特定风格的书法字体生成。

## 2.3　损失函数

在字体生成框架中，图像编码器与解码器、生成器与判别器都是镜像对称的网络架构，而情感提取框架中只存在一个单一的卷积神经网络。基于这一特点，文中针对情感提取框架和字体生成框架设计了不同的损失函数。

在情感提取框架中，为了兼顾像素的重建精度和像素间的联系，文中将损失函数表示为总变分损失（Tv Loss）和感知损失（Perceptual Loss）[17]之和，见式（1）。研究发现，受噪声污染图像的总变分明

显比无噪图像的总变分大，通过最小化总变分损失可达到最小化噪声的目的，在情感提取框架中加入总变分损失，有助于保持字体边缘的平滑性，从而保留书法图像浓淡变化之间的情感特征信息。总变分损失用于衡量单一像素的重建精度，感知损失用于衡量图像的整体特征。

$$L_{\text{spirit}} = \beta_1 * L_{\text{tv}} + \beta_2 * L_{\text{perceptual}} \tag{1}$$

式中：$L_{\text{spirit}}$，$L_{\text{tv}}$，$L_{\text{perceptual}}$ 分别为情感提取框架的整体损失、总变分损失和感知损失；$\beta_1$，$\beta_2$ 分别为总变分损失和感知损失的权重。

在字体生成框架中，损失函数由对抗损失和 1 个 L2 正则化约束共同构成[18]，见式（2）。对抗损失使用 Wassertein 距离作为衡量生成数据与真实数据分布之间距离的方式，增强训练的稳定性。

$$L_{\text{form}} = L_{\text{wgan}} + \alpha * L_2 \tag{2}$$

式中：$L_{\text{form}}$，$L_{\text{WGAN}}$，$L_2$ 分别为字体生成框架的整体损失、对抗损失和 L2 正则项；$\alpha$ 为正则化系数。

原始 GAN 在训练过程中单纯采用最小化生成器 loss 函数，最小化生成分布与真实分布之间的 KL 散度的同时又要最大化两者的 JS 散度，导致梯度不稳定，对于完全不相交的 2 个样本之间的距离甚至还会存在梯度消失的情况。损失函数的缺陷导致 GAN 训练难度增大，判别器性能很好时，生成器梯度消失严重，判别器性能不足时，生成器梯度不准导致训练不稳定。Wassertein 距离能够计算所有可能的联合分布中期望的下界，对于完全不相交的数据也能计算其距离的最优解，而且 WGAN 对应的梯度几乎成线性变化，解决了梯度消失的问题，WGAN 的计算见式（3）。

$$W\left(P_{\text{r}}, P_g\right) = \inf_{\gamma \in \prod\left(P_{\text{r}}, P_{\text{g}}\right)} \left| E_{(x-y)\sim\gamma}\left[\|x-y\|\right] \right| \tag{3}$$

式中：$\prod\left(P_{\text{r}}, P_{\text{g}}\right)$ 指真实数据与生成数据的联合概率分布，即 $\sum_x \gamma(x,y) = P_r(x)$，$\sum_y \gamma(x,y) = P_g(y)$，$\gamma \in \prod\left(P_{\text{r}}, P_{\text{g}}\right)$。

# 3 实验

文中以 CCGD-2019 数据集为基础，在 tensorflow 框架下实现了书法字体的自动生成，实验硬件设备为 Intel Core i3 4790k 4.0GHz NVIDIA Quard T4000 GPU。将 VAE，GAN，infoGAN[18]，WGAN 等算法模型与文中所提出的模型生成效果进行了比较，对网络结构参数等训练细节进行调节，以识别率（Distinguish Rate, DR）、主观评价分数（Mean object score, MOS）和 Frechet 初始距离（Fréchet Inception Distance, FID）[19]作为评估指标对生成的书法图像进行评价。

## 3.1 训练过程

在图像编码和情感提取框架中的深度学习网络中，根据文献[13]中的研究结果，使用 Adam 函数进行优化（bata 值取 0.4），学习率取 0.0002。在图像生成器和判别器网络中，采用 RMSProp 优化器，训练时使用 Xavier 方式进行初始化。生成对抗网络训练过程中损失函数随迭代次数的变化情况见图 3，图 3 中横坐标表示训练迭代次数，纵坐标表示损失函数。实验结果表明，书法字体越复杂，风格化特征越明显，潜在特征越难捕捉，所需训练迭代次数越多。一般情况下，迭代训练 650 次左右即可达到较好的效果。

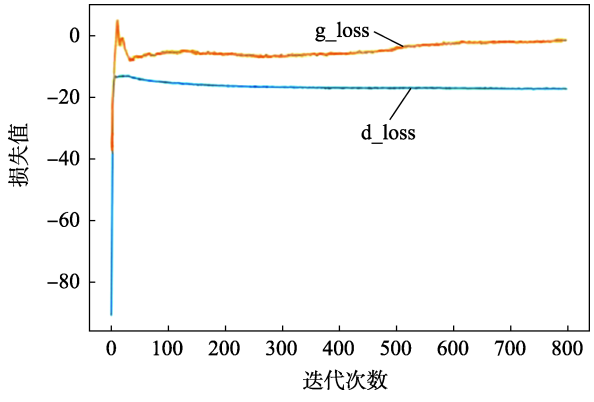


图 3 生成对抗网络训练过程

Fig.3 Training process of GAN

## 3.2 生成结果

### 3.2.1 字体汉字视觉效果

文中基于 tensorflow 框架，实现了魏碑、楷体、行书、行楷、草书、舒体、隶书、柳体（柳公权书法）等 7 种不同风格书法字体的自动化生成。部分汉字生成的案例见图 4，从图 4 中可以看出，生成的汉字字体结构清晰，字迹工整，可辨识性强，具有明显的书法风格，说明文中框架不仅能够实现书法字体的自动生成，且能够根据条件生成特定的风格。在不同风格的书法生成方面，草书、舒体等风格的字体个性化及艺术风格更为强烈，特征学习较为困难，生成效果有待进一步提高；魏碑、楷书、行书、行楷等风格的书法字体结构性强，生成效果更好，隶书、柳体的生成结果最好，还能够学习繁体字的书写特征。

### 3.2.2 与其他生成模型结果比较

将文中所提模型与 VAE，GAN，InfoGAN[20]，WGAN 等图像生成算法进行对比，比较结果见图 5。从图 5 中可以看出，VAE 所生成的结果不仅书法字体难以辨认，图像整体也比较模糊；原始 GAN 能够初步生成书法字体，但训练过程十分不稳定，生成的书法字体难以辨认，图像中包含较多噪声点；InfoGAN 在抑制噪声方面有了很大改善，但生成的

字体结构较为松散；WGAN 生成的图像质量较高，但边缘部分出现了较多“锯齿”甚至笔画缺失。与上述方法相比，文中所提出的算法模型所生成的汉字图像在结构性和图像质量方面都有明显提升，笔画结构清晰，背景无噪声点，并带有鲜明的风格特征。

图 4　文中模型生成结果

Fig.4 Generated results of this paper

图 5　不同方法生成结果比较

Fig.5 Comparison of generated results by different methods

## 3.3　图像质量评价

目前针对书法图像和生成对抗模型结果都缺少通用的评价方法，为对实验效果进行定量评价，文中采用 DR，MOS，Fréchet 初始距离（FID）作为衡量指标，对文中所提模型生成的结果进行评价。

DR 值表示生成的书法字体能被正确辨识的概率，用于衡量生成字体的准确性，MOS 值表示 5 种模型生成结果的人眼主观评价分数，用于衡量生成图像的视觉质量。通过将生成的 8 种风格的书法字体各选取 10 幅图像，随机打乱后在普通光环境下进行人眼主观评价实验，每位观察者需要对图像字体进行辨认并对图像质量进行打分（1~10 分，分值越高代表人眼视觉质量越好）。文中最终选取 30 位观察者的实验结果的平均值作为最终结果，分别见图 6—7。

FID 通过提取中间层的特征，然后用一个正态分布模拟这些特征的分布，其计算方式见式（4），用来衡量 GAN 网络的 2 个指标：生成图片的质量和多样性，较低的 FID 意味着较高图片的质量和多样性。相比较 IS 来说，FID 对噪声有更好的鲁棒性，FID 计算结果见图 8。

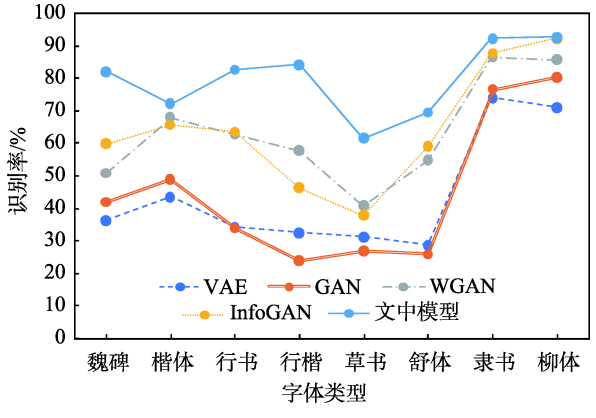


图 6　生成字体识别率结果比较

Fig.6 Comparison of recognition rate of generated characters

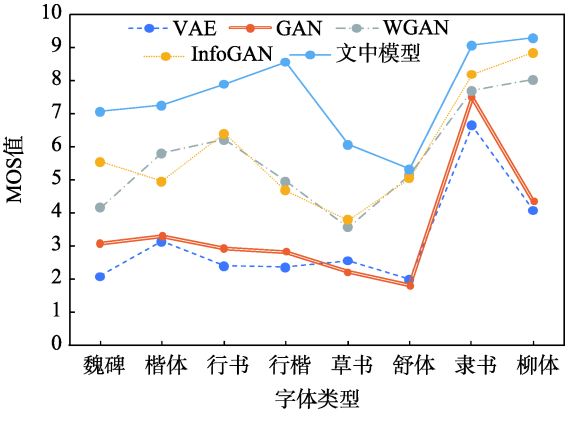


图 7　生成字体 MOS 值打分结果比较

Fig.7 Comparison of mean object score for generated characters

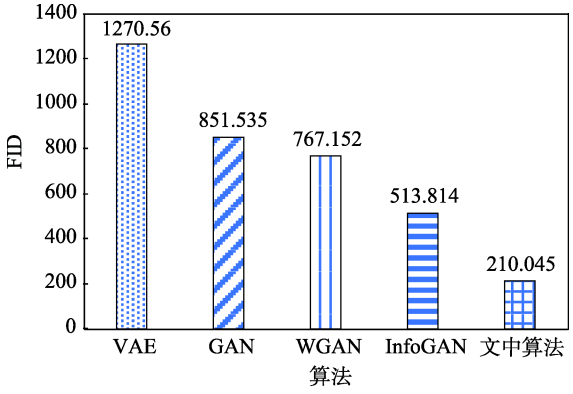


图 8 FID 计算结果比较

Fig.8 Comparison of FID results

$$F=\left\|\mu_{\mathrm{r}}-\mu_{\mathrm{g}}\right\|^{2}+T_{\mathrm{r}}\left(\Sigma_{\mathrm{r}}+\Sigma_{\mathrm{g}}-2\left(\Sigma_{\mathrm{r}}\Sigma_{g}\right)^{1/2}\right) \tag{4}$$

式中：$F$ 为 FID 计算结果；$T_{\mathrm{r}}$ 为矩阵对角线上元素的总和。

从图 6 可以看出，算法生成的各种书法字体识别率均高于其他比较生成模型，除草书、舒体等字体本身结构复杂、个性化较强导致识别率较低外，其他字体的正确识别率均达到 70%以上，对于字体框架清晰、结构工整的隶书、柳体等书法字体，正确识别率超过 90%。在视觉质量方面，从图 7 可以看出，文中算法结果的 MOS 值打分结果集中在 7 分以上，远高于其他比较模型，表明生成图像符合人眼对于书法图像的观察特性。图 8 中文中算法 FID 计算结果远低于其他模型，说明文中算法在生成书法字体多样性和视觉质量方面都有了显著提升。总体而言，文中算法不仅能够生成多种不同风格的书法图像，而且识别率高，图像质量好，尤其适用于隶书、柳体等结构性强的字体生成领域。

同时发现，图像质量与识别率并不完全对应，隶书字体的识别率最高，但柳体图像的视觉评价结果最好，草书字体的识别率最低，但舒体图像的评价质量却最差，表明人眼对于书法图像的观赏并不完全取决于能否正确识别书法作品中的字体，也进一步表明了书法中的情感风格在书法图像生成与研究中的重要性。

## 4 结语

文中以生成对抗网络为基础，提出一种基于生成对抗网络的书法字体生成模型，实现了特定书法风格的汉字字体生成。模型由情感提取框架和字体生成框架 2 部分组成，情感提取框架通过颜色空间转换和卷积神经网络，提取隐藏于灰度信息中的风格特征，字体生成框架由一个变分自编码器和生成对抗网络组合而成，在实现模型的过程中还建立了书法图像数据库为网络模型提供训练样本。实验表明，文中模型能够自动生成特定风格的书法字体且视觉效果良好。目前算法对于草书、舒体等个性化风格比较浓烈的字体生成效果尚待提高，部分生成图像中包含噪声干扰，网络结构的稳定性和普适性有待进一步加强。

**参考文献：**

[1] 雷国荣. 古代碑刻书法图像清晰化方法及系统研究[D]. 西安: 西安理工大学, 2018: 7—33.
LEI Guo-rong. Image Clear Method and System of Ancient Inscriptions Calligraphy[D]. Xi'an: Xi'an University of Technology, 2018: 7—33.

[2] 许鹏飞. 基于 Guide-filter 的中国书法作品中印章信息的提取[J]. 传感器与微系统, 2016, 35(12): 125—128.
XU Peng-fei. Extraction of Seals From Chinese Calligraphy Works Based on Guide-filter[J]. Transducer and Microsystem Technologies, 2016, 35(12): 125—128.

[3] 张福成. 基于卷积神经网络的书法风格识别的研究[D]. 西安: 西安理工大学, 2018: 36—49.
ZHANG Fu-cheng. Calligraphy Style Recognition Based on Convolution Neural Network[D]. Xi'an: Xi'an University of Technology, 2018: 36—49.

[4] 章夏芬, 张龙海, 韩德志, 等. 自适应书法字图像匹配和检索[J]. 浙江大学学报(工学版), 2016, 50(4): 766—776.
ZHANG Xia-fen, ZHANG Long-hai, HAN De-zhi, et al. Adaptive Matching and Retrieval for Calligraphic Character[J]. Journal of Zhejiang University (Engineering Science), 2016, 50(4): 766—776.

[5] 任晓文, 王涛, 李健宇, 等. 基于深度学习的异噪声下手写汉字识别的研究[J/OL]. 计算机应用研究: 1—5[2019-05-30]. https://doi.org/10.19734/j.issn.1001-3695.2018.06.0579.
REN Xiao-wen, WANG Tao, LI Jian-yu, et al. Research on Handwritten Chinese Character Recognition Based on Deep Learning with Different Noise[J/OL]. Application Research of Computers: 1—5[2019-05-30]. https://doi.org/10.19734/j.issn.1001-3695.2018.06.0579.

[6] ZHENG X, MIA Q, SHI Z, et al. A New Artistic Information Extraction Method with Multi Channels and Guided Filters for Calligraphy Works[J]. Multimedia Tools and Applications, 2016, 75(14): 8719—8744.

[7] 张巍, 张筱, 万永菁. 基于条件生成对抗网络的书法字笔画分割[J/OL]. 自动化学报: 1—8[2019-07-22]. https://doi.org /10.16383/j.aas.c190141.
ZHANG Wei, ZHANG Xiao, WAN Yong-jing. Stroke Segmentation of Calligraphy Based on Conditional Generative Adversarial Network[J]. Acta Automatica Sinica: 1—8[2019-07-22]. https://doi.org/10.16383/j.aas.c190141.

[8] LYU P, BAI X, YAO C, et al. Auto-Encoder Guided GAN for Chinese Calligraphy Synthesis[J]. International Conference on Document Analysis and Recognition, 2017(1): 1095—1100.

[9] 任健. 面向文物修复的中国书法风格迁移研究与应

用[D]. 西安: 西北大学, 2019: 31—45.
REN Jian. Research and Application of Chinese Calligraphy Style Transfer for Cultural Relics Restoration[D]. Xi'an: Northwest University, 2019: 31—45.

[10] DIEDERIK P K, WELLING M. Auto-Encoding Variational Bayes[C]// International Conference on Learning Representations, 2014.

[11] GOODFELLOW I, POUGET J, MIRZA M, et al. Generative Adversarial Nets[C]// Conference on Advances in Neural Information Processing Systems. Montreal, Curran Associates, 2014: 2672—2680.

[12] ZHAO Z, FEI Y, ZHANG X Y, et al. Handwritten Chinese Character Blind Inpainting with Conditional Generative Adversarial Nets[C]// IEEE Asian Conference on Pattern Recognition, 2017: 804—809

[13] RADFORD A, METZ L, CHINTALA S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[C]// International Conference on Learning Representations, 2016.

[14] WU Ji-qing, HUANG Zhi-wu, THOMA J. Wasserstein Divergence for GANs[C]// European Conference on Computer Vision, 2018: 673—688.

[15] LIU Cheng-lin, YIN Fei, WANG Da-han, et al. CASIA Online and Offline Chinese Handwriting Databases[C]// International Conference on Document Analysis and Recognition, 2011: 1520—5363.

[16] ANDERS B L L, SONDERBY S K, LAROCHELLE H, et al. Autoencoding Beyond Pixels Using a Learned Similarity Metric[C]// International Conference on Machine Learning, 2016: 1558—1566.

[17] LIU Gui-lin, FITSUM A, KEVIN J, et al. Image Inpainting for Irregular Holes Using Partial Convolutions[C]// European Conference on Computer Vision, 2018: 89—105.

[18] OREST K, VOLODYMYR B, MYKOLA M, et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks[C]// Computer Vision and Pattern Recognition, 2018: 8183—8192.

[19] SALIMANS T, GOODFELLO I, ZAREMBA W, et al. Improved Techniques for Training GANs[C]// Neural Information Processing Systems, 2016: 2234—2242.

[20] XI Chen, YAN Duan, HOUTHOOFT R, et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[C]// Conference and Workshop on Neural Information Processing Systems, 2016: 2172—2180.