

基于对比学习和田字格变换数据增强的小样本汉字字体生成方法

曾锦山, 汪叶飞, 熊康悦, 袁艺扬

(江西师范大学计算机信息工程学院 南昌 330022)
(jinshanzeng@jxnu.edu.cn)

摘要: 现有小样本汉字字体生成模型难以通过少量参考汉字有效实现字体风格与内容解耦, 导致字体生成效果欠佳. 为解决该问题, 提出一种基于对比学习和田字格变换数据增强的小样本汉字字体生成方法. 利用对比学习思想, 对输出图像分别进行风格对比和内容对比, 更好地实现字体风格和内容的自适应解耦; 为了进一步提高小样本字体生成性能, 提出一种基于田字格变化的数据增强方式, 通过对田字格区域的几何变换, 嵌入字体的全局结构信息, 提高字体的风格和内容解耦能力. 所提方法的有效性在公开的汉字字体数据集进行多次实验得到验证. 实验结果表明, 与当前先进的小样本字体生成模型相比, 所提方法在生成字体质量上有显著优势, 并且 FID, SSIM, PSNR, LPIPS 等性能指标也均优于对比方法模型. 特别地, 所提方法嵌入现有小样本字体生成模型, 仅使用较少训练数据集训练后, 便可显著地提升现有小样本汉字字体生成模型性能.

关键词: 小样本字体生成; 汉字字体生成; 生成对抗网络; 深度学习
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00733

Few-Shot Chinese Font Generation Based on Contrastive Learning and Square-Block Transformation Data Augmentation

Zeng Jinshan, Wang Yefei, Xiong Kangyue, and Yuan Yiyang

(School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022)

Abstract: Existing few-shot Chinese font generation models struggle to effectively decouple font style and content using a limited number of reference characters, leading to suboptimal font generation results. To address this issue, a novel few-shot Chinese font generation method based on contrastive learning and square-block transformation data augmentation is proposed. Utilizing the concept of contrastive learning, the output images undergo separate style and content comparisons, achieving better adaptive decoupling of font style and content. To further improve the performance of few-shot font generation, a data augmentation method based on square-block transformations is introduced, embedding the global structural information of the fonts through geometric transformations of the square-block areas, thereby enhancing the ability to decouple font style and content. The effectiveness of the proposed method is validated through extensive experiments on public Chinese font datasets. Experimental results demonstrate that, compared to current state-of-the-art few-shot font generation models, the proposed method significantly improves the quality of generated fonts and outperforms comparative models in performance metrics such as FID, SSIM, PSNR, and LPIPS. Notably, the proposed method can be integrated into existing few-shot font generation models and, with only a

收稿日期: 2023-11-13; 修回日期: 2024-07-10. 基金项目: 国家自然科学基金(62376110, 61977038); 江西省“双千计划”(jxsq2019201124); 江西省自然科学杰出青年基金(20224ACB212004). 曾锦山(1987—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为人工智能算法理论、计算机视觉、自然语言处理等; 汪叶飞(1997—), 男, 硕士研究生, 主要研究方向为计算机视觉、小样本学习等; 熊康悦(2003—), 男, 在校学生, 主要研究方向为计算机视觉、字体生成等; 袁艺扬(2003—), 女, 硕士研究生, 主要研究方向为计算机视觉、图像生成.

small training dataset, can significantly enhance the performance of these models.

Key words: few-shot font generation; Chinese font generation; generative adversarial networks; deep learning

小样本汉字字体生成指利用少量参考字符生成相应的汉字字体,近年来得到了研究者的持续关注. 现有的小样本汉字字体生成方法基于风格与内容解耦^[1-4],从汉字字库图像中提取内容信息,同时从少量参考汉字^[1]中提取风格先验信息;之后将风格先验与内容先验进行有效的融合,生成具有参考字体风格的汉字. 文献[1]改进了字体风格和内容表示,提出一种由风格编码器、内容编码器、混合器和解码器组成的广义风格传输网络;文献[2]在网络中设计不同模块约束字形和风格,最后通过2个解码器优化生成的汉字字形及风格;在DG-Font^[3]利用对抗损失分离汉字的风格与内容特征,并加入可变卷积模块提升汉字的生成质量. 文献[1-3]提出的小样本汉字字体生成方法主要是基于统计建模的方式,如通过估计相应的统计量分别表征风格和内容特征,实现风格与内容解耦;然而,由于汉字字体结构复杂和风格多样,基于统计建模的方式难以有效地表征汉字字体风格和内容信息,导致生成的字体效果欠佳. 因此,许多研究者将汉字的笔画、部首等局部信息引入模型中,并使用大量训练数据集来促使模型记忆汉字局部信息,提升小样本字体生成的性能^[5-10].

文献[5]首次提出将潜在特征分解为与内容相关和与风格相关的组件来捕捉汉字的不同特征;LF-Font^[6]将汉字字符分解为不同的部件,并通过模型对不同部件的风格与内容信息进行解耦;在此基础上,MX-Font^[7]进一步改进,通过固定数量的专家模块来对不同局部的风格与内容特征解耦;CG-GAN^[8]的判别器中利用注意力机制捕捉汉字部件信息,通过分类器对部件的风格与内容进行解耦;FS-Font^[9]利用交叉注意力从不同字符上拆解部件,并通过变形重组获得目标汉字,保证字符部件上风格与内容的一致性;文献[10]利用跨模态信息,通过预训练的模型捕捉字体的笔顺信息,提高汉字风格与内容的解耦性能;然而,这些基于汉字局部特征的小样本字体生成方法忽略了汉字字体结构信息,导致生成字符不完整,且整体风格表征效果不佳,同时,大量的训练数据集的收集也为小样本字体生成带来了困难.

为了进一步增强汉字的风格和内容信息,提升小样本汉字生成的性能,本文提出一种基于对比

学习模块(comparative learning module, CLM)和田字格数据增强(square-block transformation data augmentation, SDA)模块的小样本字体生成方法(comparative learning and square-block transformation data augmentation, CSDA). 所引入的 CLM 可有效克服现有基于统计建模的风格与内容解耦方式在模型表示上的不足,实现汉字字体风格与内容自适应解耦. 而 CDA 在利用较少训练数据集的同时,可有效捕捉汉字字体结构信息,以及汉字风格与内容先验,并提升对比学习的解耦效果. 值得注意的是,基于对比学习^[11-14]和田字格变换数据增强^[15-17]的风格内容与解耦模块无须改变现有小样本字体模型且不增加任何人工成本,从而可作为插件嵌入现有小样本字体生成模型中并进一步提升模型性能. 本文所提模块的有效性和通用性在多个现有小样本字体生成模型上得到验证.

1 相关工作

1.1 基于图像风格转换的汉字字体生成方法

作为一类特殊的图像,汉字字体生成问题通常可视为一类特殊的图像风格迁移问题,即将源风格字体映射到目标风格字体,如图1所示. 随着生成对抗网络的发展,基于生成对抗网络的汉字生成方法成为当前的主流方法. 文献[18]提出一种基于配对数据的汉字字体生成模型,虽然该模型的汉字生成效果不错,但配对数据集的获取通常耗时耗力. 为了解决配对数据集的问题,文献[19]将图像风格转换中的 CycleGAN 模型^[20]用于汉字字体生成. 然而,由于缺乏足够的信息引导,基于 CycleGAN 的汉字字体生成方法容易出现模式坍塌的现象^[21-22]. 为此,研究人员提出一系列基于汉字先验信息的字体生成方法^[15,23-27]. 文献[23]和文献[24]通过引入单字节笔画编码缓解 CycleGAN 的模式坍塌问题;文献[25]拓展了文献[23]的思想,并在模型中进一步采用少量配对数据加强信息引导;文献[15]通过引入田字格变换自监督方式引导网络捕捉汉字字体的结构信息;文献[26]利用骨架信息作为信息引导缓解 CycleGAN 的模式坍塌问题;文献[27]将汉字轮廓信息以及区域注意力用于汉字书法字体生成. 虽然这些方法生成的汉字效

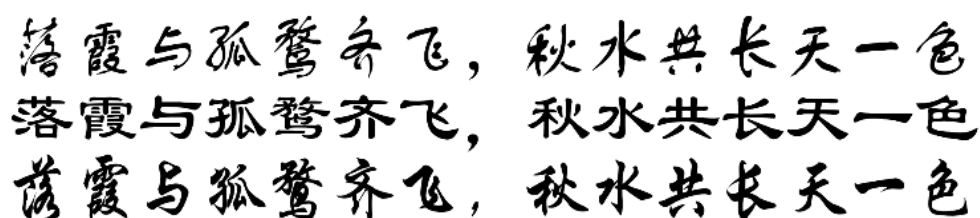


图 1 3 种不同汉字风格字体展示

果较好, 但需要大量的训练数据集, 通常难以将其应用到小样本字体生成.

1.2 小样本汉字字体生成方法

当前主流的小样本字体生成方法基于风格与内容解耦, 其中一些典型方法包括 EMD^[1], DG-Font^[3], SA-VAE^[5], MX-Font^[7], CG-GAN^[8], DC-Font^[28] 和 LF-Font^[29]. SA-VAE^[5] 首次提出采用风格与内容解耦的思想解决小样本字体生成; 之后, 该思想被文献[1]推广到未见风格字体生成. 为了更好地引导网络实现风格与内容解耦, LF-Font^[29] 训练模型学习组件级别的风格表征, 而不是通用风格; MX-Font^[7] 采用专家编码器对图像不同局部位置进行内容和风格解耦; FS-Font^[9] 从参考图像中获取细粒度的局部样式, 同时在内容和参考字形之间建立空间对应关系; CG-GAN^[8] 利用组件感知模块引导生成器在更复杂的层次上区分内容和样式; XMP-Font^[10] 提出一种基于跨模态转换器的编码器, 同时考虑了字形图像和相应的笔画标签: 这些

模型通过引入字符笔画和组件信息捕捉汉字的局部信息, 虽然可以通过引入汉字局部先验信息提高模型的生成效果, 但通常忽略了汉字的结构信息, 导致生成字体出现字体结构不完整的问题.

1.3 对比学习

小样本字体生成方法在训练中使用大量数据集, 使模型在大量数据中记忆字体特征, 进一步在测试中实现小样本生成. 然而, 由于大量数据集的获取本身就是一件困难的事情, 因此人们更加关注对比学习等无监督方法. SimCLR^[11] 通过总结对比学习的结构, 利用随机剪裁和颜色扰动等方式, 在提升对比学习的性能的同时增强了训练数据集. 文献[12]通过实例的 Softmax 嵌入方法, 提高了对比学习的学习速度和准确性; 文献[13]对同一样本的不同视角进行编码来学习样本的特征表示, 在无须标注数据集的情况下学习到样本的特征表示, 然而受到汉字字体结构的影响, 如图 2 所示, 这些方法往往表现不佳.

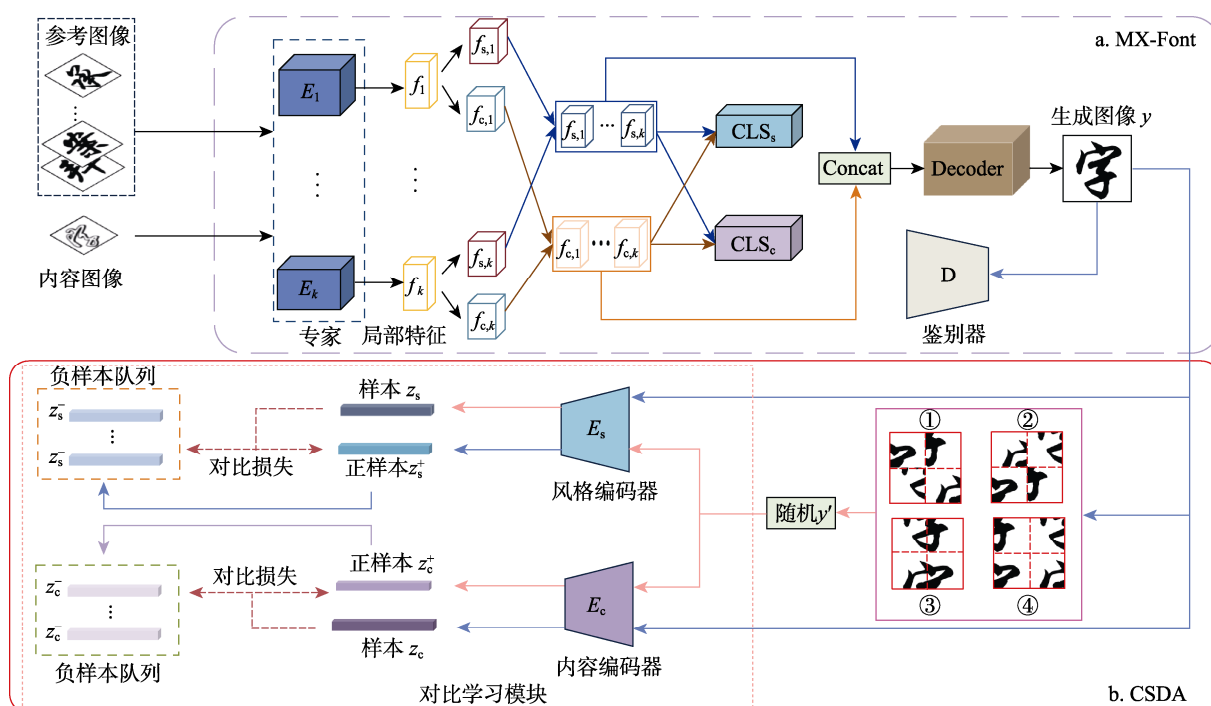


图 2 CSDA 的结构图

为了解决以上问题,本文提出一种基于对比学习和田字格变换数据增强的风格与内容解耦模块,其中,对比学习可有效地从数据中自适应地解耦风格与内容信息,而基于田字格变换数据增强方式在提供汉字字体结构信息的同时需要的训练数据集更少,并能达到较高解耦效果.该模块无须改变现有小样本字体生成模型,能够以插件的方式嵌入现有小样本字体生成模型,进一步提升模型的性能.

2 CSDA

CSDA 的结构如图 2 所示,其中,基线模型以当前最先进的小样本汉字生成模型 MX-Font^[7]为例,阐述 CSDA 的工作原理. CSDA 中,包含 CLM 模块和 SDA 模块.

2.1 MX-Font 方法

MX-Font 是当前具有代表性的小样本字体生成方法之一,其结构如图 2a 所示,由本地化专家 E_i 、风格分类器 Cls_s 、内容组件分类器 Cls_c 、解码器和鉴别器组成.图 2 中, MX-Font 的输入为提供内容特征的内容图像和提供风格特征的参考图像,其中,内容图像和参考图像会一起经过 k 个专家提取输入图像的局部特征 $f_i, \dots, f_k$;之后,分别与风格权重 $w_{i,s}$ 和内容权重 $w_{i,c}$ 相乘,提取出局部内容特征 $f_{c,i}$ 和局部风格特征 $f_{s,i}$.在这个过程中,为了使每个不同的本地化专家能够关注到不同的局部组件特征而不用在意组件所处位置, MX-Font 通过匈牙利算法^[30]将组件分配问题视为加权二分匹配问题,并借用希尔伯特-施密特独立性准则^[31]度量每个专家的独立性.第 i 个专家内部局部风格特征与内容特征的独立性损失 $L_{\text{intra-indp},i}$,以及与其他专家的独立性损失 $L_{\text{inter-indp},i}$ 分别定义为

$$L_{\text{intra-indp},i} = \text{HSIC}(f_{s,i}, f_{c,i}),$$

$$L_{\text{inter-indp},i} = \sum_{i' \neq i} \text{HSIC}(f_i, f_{i'}).$$

得到输入图像的局部内容特征和风格特征之后,分别组合这些输入的图像局部内容特征和风格特征,获得完整的内容特征和风格特征,再将内容字符的内容特征和参考字符的风格特征相结合,送入解码器中生成目标汉字.

为了更好地提取内容与风格特征, MX-Font 还引入一个 Cls_s 和 Cls_c 作为弱监督,分别引导字体

风格与内容特征的提取.对于多个本地化专家提取出来的 $f_{s,i}$ 和 $f_{c,i}$,将它们分别输入风格分类器和内容分类器来预测风格和-content,并且通过设计 2 个专门的损失函数实现 2 个分类器的一致性预测,定义为

$$L_{s,i}(f_{s,i}, y_s) = \text{CE}(\text{Cls}_s(f_{s,i}), y_s) - H(\text{Cls}_c(f_{s,i})),$$

$$L_{c,i}(f_{c,i}, U_c) = \sum_{j \in U_c} w_{ij} \text{CE}(\text{Cls}_c(f_{c,i}), j) - H(\text{Cls}_s(f_{c,i})).$$

其中, CE 表示交叉熵损失(cross entropy loss, CE); U_c 表示输入字符所对应的组件标签,即 $u_1^c, u_2^c, \dots, u_m^c$.对于所有组件 j ,如果 j 属于输入字符的组件标签 U_c ,则 $w_{ij} = 1$,否则, $w_{ij} = 0$, $i = 1, 2, \dots, k$, $j = 1, 2, \dots, m$, m 表示输入字符的组件数; H 表示分类器所得的最大预测概率.

MX-Font 借用生成对抗损失^[21]进行优化,其定义为

$$L_{\text{adv}}^D = E_{(\hat{y}, y_c, y_s)} \left\{ [1 - D_s(\hat{y}, y_s)]_+ + [1 - D_c(\hat{y}, y_c)]_+ \right\},$$

$$L_{\text{adv}}^G = -E_{(y, y_c, y_s)} [D_s(y, y_s) + D_c(y, y_c)].$$

其中, y 表示由 y_c 的内容特征和 y_s 的风格特征组合而成的生成图像, $\hat{y}$ 表示目标图像.为了提升模型生成性能及训练的稳定性, MX-Font 的鉴别器 D 除了对生成图像进行判别真伪,还对生成图像 y 和目标图像 $\hat{y}$ 在不同层之间的特征进行匹配,从特征层面对图像质量进行约束,特征匹配损失定义为

$$L_{\text{fm}} = E_{(\hat{y}, y)} \left[\sum_{l=1}^{L-1} \|D^l(\hat{y}) - D^l(y)\|_1 \right].$$

其中, L 表示判别器 D 的层数, D^l 表示判别器 D 的第 l 层输出.

MX-Font 还引入重构损失 L_{recon} 来约束模型训练,定义为 $L_{\text{recon}} = E_{(\hat{y}, y)} [\|\hat{y} - y\|_1]$.

2.2 CSDA

MX-Font 中,局部专家主要关注汉字的局部特征而忽视了汉字的整体结构信息.为了更好地捕捉汉字的结构信息,受到文献[14-15]的启发,本文设计一种基于对比学习和田字格变换数据增强的风格与内容解耦模块,包括 SDA 和 CLM 这 2 个子模块.

2.2.1 SDA 模块

由于汉字结构复杂,因此获取其结构的信息的较为困难.受到文献[15]的启发,本文提出一种 SDA 模块,将汉字的结构信息融入模型中,使模型

学到更多的汉字信息。

如图 3 所示, 给定一个汉字, 首先按照“田字格”将汉字图像分为 4 个区域, 分别编号为 1~4, 然后定义 4 种几何变换, 其中, 这些变换对应于汉字独体结构、左右结构、上下结构和半包围结构 4 种基本结构。

变换 1. 受到汉字独体结构的启发, 每个区域的位置保持不变, 如日、太等;

变换 2. 受到汉字左右结构的启发, 区域 1 和区域 3 与区域 2 和区域 4 互换, 如涂、汪等;

变换 3. 受到汉字上下结构的启发, 区域 1 和区域 2 与区域 3 和区域 4 互换, 如杰、杏等;

变换 4. 受到汉字右上和左下包围结构的启发, 区域 1 与区域 3 互换, 如氛、连等。

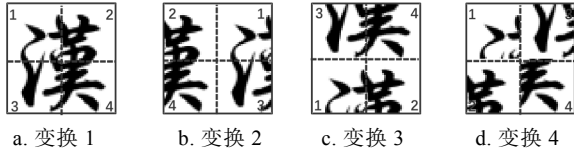


图 3 4 种田字格变换

在训练中, 将生成图像输入 SDA 随机获取一种结构变换结果, 并将其与生成图像分别输入风格与内容编码器。此过程中, SDA 模块对汉字图像实施随机变换, 进一步辅助 CLM 捕捉汉字结构信息。

2.2.2 CLM 模块

为了更有效地实现风格与内容解耦, 引入 CLM。如图 3 所示, CLM 包含风格对比学习和内容对比学习 2 部分。在风格对比学习部分, 对于 MX-Font 生成的图像 y , CSDA 首先将其输入 SDA 模块中, 随机获取一种结构变换的结果 y' ; 然后将 y 和 y' 分别输入风格编码器 E_s , 得到 2 个风格特征 Z_s 和 Z_s^+ 。由于 y' 是由 y 经过结构变换之后得到的汉字, 它们的风格特征通常具有一定的相似性, 因此把 y' 的风格特征作为正样本。同时, 定义一个包含 $K = 65\ 520$ 个风格特征的负样本队列 Z_s^- , 其在开始训练时随机生成, 每次执行对比损失后, 负样本队列执行出队操作, 而正样本 Z_s^+ 执行入队操作, 加入负样本队列。特别地, 负样本在训练中, 每个批次的特征与输出图像 y 编码得到的风格特征 Z_s 通常不相似。类似地, 将 y 和 y' 分别输入内容编码器 E_c 得到 2 个内容特征 Z_c 和 Z_c^+ , 以及定义一个包含 65 520 个风格特征的负样本队列 Z_c^- 。基于这些正负样本, 分别定义内容对比损失

L_{contr_c} 和风格对比损失 L_{contr_s} 为

$$L_{\text{contr}_c} = L_{\text{NCE}}(Z_c, Z_c^-, Z_c^+),$$

$$L_{\text{contr}_s} = L_{\text{NCE}}(Z_s, Z_s^+, Z_s^-).$$

其中,

$$L_{\text{NCE}}(Z, Z^+, Z^-) = -\text{lb} \frac{\exp\langle Z, Z^+ \rangle / \tau}{\exp\langle Z, Z^+ \rangle / \tau + \sum_{i=0}^K \exp\langle Z, Z^- \rangle / \tau}, \langle Z, Z^+ \rangle,$$

其表示 Z 与 Z^+ 的内积; τ 表示温度超参数, 默认设置为 0.7。基于上述定义的对比损失 L_{NCE} , CSDA 可以实现风格与内容的有效解耦。CSDA 的损失函数定义为

$$L_{\text{CSDA}} = L_{\text{adv}}^D + L_{\text{adv}}^G + L_{\text{recon}} + L_{\text{fm}} + \sum_{i=1}^k [L_{i,s} + L_{i,c} + L_{\text{intra-indp},i} + L_{\text{inter-indp},i}] + \lambda_{\text{contr}} (L_{\text{contr}_c} + L_{\text{contr}_s}),$$

其中, λ_{contr} 用于对比损失, 本文实验中设置为 0.1。

3 实验及结果分析

通过一系列实验, 验证 CSDA 方法的有效性。所有实验都是在 Linux 环境下进行的, 实验平台为 16 核和 24 线程的英特尔酷睿 i9-12900K CPU, 一个 GeForce RTX 3090 GPU。

3.1 实验设置

3.1.1 数据集

为了评估 CSDA, 从互联网上收集了印刷体和手写体共计 35 种字体, 其中, 30 种风格字体用于训练, 称为已见风格字体; 5 种字体用于测试, 称为未见字体。使用已见风格字体对模型进行训练, 其中, 每种字体使用 2988 个字符训练, 其余字符作为测试集。训练集使用的字符称为已见字符, 测试集使用的字符称为未见字符。

将未见风格任务划分为未见风格已见字符和未见风格未见字符 2 类子任务, 使用已见风格已见字符任务中训练过的模型直接在 2 类子任务上测试。对于未见风格的 5 种字体, 需要在每种字体中随机选取 4 幅图像作为参考汉字, 生成所有测试字符。为了得到更好的生成效果, 使用常见的宋体作为源字体提供内容特征。

此外, 准备了一个包含 100 种字体的训练数据集用于评估 CSDA 在少量训练数据集上的性能, 其中, 使用的已见字符与 30 种字体的训练数据集相同。

3.1.2 对比模型

为了验证 CSDA 的性能, 考虑 3 大类汉字生成模型: 基于生成对抗网络的汉字生成模型, 基于汉字特征先验的字体生成模型和基于汉字组件特征信息的小样本字体生成模型. 其中, 一些典型的生成模型包括基于生成对抗网络和非配对数据的典型模型 CycleGAN^[20]、基于田字格几何变换的自监督学习的 CycleGAN 改进模型 SQ-GAN^[15]、基于风格与内容解耦表征学习和非配对数据的图像迁移模型 FUNIT^[32]、基于组件因子分解和弱监督的小样本汉字字体生成模型 LF-Font^[6]、通过引入风格与内容编码器和分类器实现风格与内容解耦的小样本汉字生成模型 MX-Font^[7](它也是 CSDA 的基模型)、基于组件鉴别器的小样本汉字生成模型 CG-GAN^[8]、内容风格融合的小样本字体生成模型 CF-Font^[33]、以及多层相似度计算的汉字字体生成模型 LSG-FCST^[34].

3.1.3 评估指标

为了有效地评估 CSDA 的性能, 采用 5 个常用评估指标: 峰值信噪比(peak signal-to-noise ratio, PSNR)用于评估生成图像质量的锐化程度; 结构相似性(structural similarity, SSIM)^[35]用于量化处理导致图像质量退化的感知程度; 学习感知图像块相似性(learning perceived image block similarity, LPIPS)^[36]用于量化生成图像和真实图像之间的感知相似性; 均方误差损失(mean square error, MSE)用来衡量生成样本与真实样本在像素意义下的损失; FID(Fréchet inception distance)^[37]用于度量生成图像和真实图像之间的距离. 通常, FID, LPIPS 和 MSE 值越小, 表示图像生成质量越好; 而 SSIM 和 PSNR 值越大, 表示图像生成质量越好. LPIPS 和 SSIM 的取值都在(0,1).

3.1.4 模型结构和优化器

CSDA 方法遵循生成对抗网络框架的主要思想, 同时引入了对比学习增强字体生成性能. 其风格与内容编码架构如表 1 所示, CSDA 中风格编码器和内容编码器的架构经过专门设计, 能更好地

表 1 CSDA 中风格与内容编码器架构

模块	CSDA				
	层	卷积核	通道	归一化	激活函数
内容编码器	Convolution	7	64	IN	ReLU
	Convolution	4	128	IN	ReLU
	Convolution	4	256	IN	ReLU
	Convolution	4	512	IN	ReLU
	Residual block	3	512	IN	ReLU
	Residual block	3	512	IN	ReLU
风格编码器	Convolution	7	64		
	Convolution	4	128		
	Convolution	4	256		
	Convolution	4	256		
	Convolution	4	256		
	AvgPooling		256		
	Convolution	1	512		

捕捉汉字风格与内容特征.

此外, 本文使用 Adam^[38]作为优化器对 CSDA 进行训练, 其中, 学习率取为 0.001, 2 个动量参数设置为默认值(0.900, 0.999), 优化器的迭代轮次设置为 20 万次.

3.2 消融实验

通过一系列消融实验, 本文验证了 CSDA 方法的有效性. 分别移除 SDA 和 CLM 模块, 观察其对模型性能的影响; 其次通过在现有模型中加入 CSDA 方法测试其可拓展性; 最后, 通过多种变换方式验证 SDA 模块的有效性.

3.2.1 CSDA 方法的有效性

CSDA 方法中包含 CLM 和 SDA 模块. 在已见风格未见字符, 以及未见风格未见字符 2 类任务上进行实验, 说明所提模块的有效性, 结果如表 2 所示. 在未见风格未见字符任务中的结果为 10 个字体生成任务的平均值, 而在未见风格未见字符任务中的结果为 5 个字体生成任务的平均值, w/o SDA 和 w/o (SDA, CLM)分别对应 CSDA 去除 SDA 模块和同时去除 SDA 与 CLM 模块.

表 2 消融实验结果 1

任务	模型	FID↓	LPIPS↓	MSE↓	SSIM↑	PSNR↑
已见风格未见字符	MX-Font w/ CSDA	47.60	0.305	0.205	0.443	7.07
	w/o SDA	49.16	0.308	0.211	0.442	7.01
	w/o (SDA, CLM)	58.35	0.316	0.215	0.428	6.86
未见风格未见字符	MX-Font w/ CSDA	40.97	0.281	0.196	0.475	7.16
	w/o SDA	41.92	0.286	0.202	0.459	7.03
	w/o (SDA, CLM)	45.24	0.291	0.205	0.455	6.97

注: 粗体表示最优值.

从表 2 可以看出, SDA 模块和 CLM 模块在 2 类任务中对 CSDA 的性能都有实质提升; 在 2 类字体生成任务中, 通过先后加入 CLM 和 SDA 模块, CSDA 在未见风格未见字符任务中的 FID 分别降低了 9.19 和 1.56, 而在未见风格未见字符中的 FID 分别降低了 3.32 和 0.95; 从 FID 的角度看, 加入 CLM 对 CSDA 的性能提升更为显著. 实验结果表

明, CLM 可以更加有效地实现风格与内容解耦, 同时, SDA 可以进一步提升 CLM 的解耦效果. 除了表 2 中的量化指标, 表 3 所示为先后加入 CLM 和 SDA 模块的汉字字体生成效果. 可以看出, 在加入 CLM 和 SDA 之后, 生成汉字的图像质量显著提升, 生成的汉字风格和内容都更优. 这些实验结果验证了 CSDA 的有效性.

表 3 消融实验结果 2

模型/数据	可见风格不可见内容	不可见风格不可见内容
参考字体	个乃喬且串京们乡	祆電躡豆頌逼輩迭
源字体	逞角霖鯨陞躡謬輕	靶趁齿触边钉订购
w/o (SDA, CLM)	逞角霖鯨陞躡謬輕	靶趁齿触边钉订购
w/o SDA	逞角霖鯨陞躡謬輕	靶趁齿触边钉订购
CSDA-Font	逞角霖鯨陞躡謬輕	靶趁齿触边钉订购
真实图像	逞角霖鯨陞躡謬輕	靶趁齿触边钉订购

在当前较先进的模型 CG-GAN^[8]和 CF-Font^[33]上, 验证 CSDA 的有效性. 分别采用 30, 60 和 100 个训练数据集对模型进行训练, CG-GAN w/CSDA 和 CF-Font w/CSDA 分别表示在 CG-GAN^[8]和 CF-Font^[33]上加入 CSDA. 在未见字体上验证 CSDA 在少量数据集上的优异性能, 结果如表 4 所示. 其中, 未见字体的结果为随机选取未参与训练的 5 个字体生成结果的平均值, w/ CSDA 表示基模型加入

CSDA 后的结果. 可以看出, 加入 CSDA CG-GAN^[8]和 CF-Font^[33]模块后, CG-GAN^[8]和 CF-Font^[33]的性能均显著提升, 表明 CSDA 有效地提升了模型的生成性能; 当训练数据集为 30 时, 加入 CSDA 的 CG-GAN^[8]和 CF-Font^[33]的性能在多数指标上能超越训练数据集为 100 的基模型 CG-GAN^[8]和 CF-Font^[33], 在其他指标上性能取得次优. 表明 CSDA 能够在训练数据集更少情况下, 达到较好的解耦效果.

表 4 消融实验结果 3

任务	模型	Dataset	FID↓	MSE↓	LPIPS↓	PSNR↑	SSIM↑
未见字体已见字符	CG-GAN	30	39.51	0.199	0.283	7.17	0.478
		60	40.54	0.195	0.281	7.21	0.480
		100	35.66	0.189	0.273	7.30	0.496
	CG-GAN w/ CSDA	30	35.44	0.186	0.271	7.26	0.482
	CF-Font	30	40.99	0.188	0.287	7.22	0.498
		60	38.64	0.182	0.276	7.31	0.502
		100	31.66	0.181	0.268	7.40	0.508
	CF-Font w/ CSDA	30	30.22	0.185	0.265	7.41	0.510
未见字体未见字符	CG-GAN	30	42.28	0.208	0.282	6.92	0.451
		60	40.69	0.205	0.283	6.99	0.455
		100	35.33	0.199	0.279	7.10	0.455
	CG-GAN w/ CSDA	30	37.91	0.200	0.269	7.01	0.453
	CF-Font	30	52.53	0.196	0.278	7.27	0.491
		60	43.22	0.191	0.280	7.31	0.492
		100	34.61	0.191	0.278	7.40	0.500
	CF-Font w/ CSDA	30	36.01	0.189	0.266	7.38	0.499

注: 粗体表示最优值.

3.2.2 SDA 模块的有效性

进一步,与现有 2 种典型的图像数据增强方式,即对汉字图像进行水平翻转和垂直翻转的数据增强方式进行实验,验证本文数据增强方式的有效性.与第 3.2.1 节的实验类似,也同时考虑了已见风格未见字符与未见风格未见字符 2 类字体生成任务,结果如表 5 所示.可以看出,本文所提 SDA 模块在大多数指标上显著优于现有的水平翻转和垂直翻转的数据增强方式,而垂直翻转数据增强方式的效果通常优于水平翻转数据增强效果.其中潜在原因在于本文所提的 SDA 模块不仅可以提供更多的风格和内容信息,而且可以提供更多的汉字结构信息,从而更加有效地提升模型性能.实验结果验证了本文所提 SDA 模块的优越性.

表 5 SDA 与现有常用数据增强方式对比

任务	模型	FID↓	LPIPS↓	MSE↓	SSIM↑	PSNR↑
已见风格 未见字符	水平翻转	60.83	0.310	0.206	0.443	7.02
	垂直翻转	52.47	0.307	0.206	0.441	6.99
	田字格	47.60	0.305	0.205	0.443	7.07
未见风格 未见字符	水平翻转	47.39	0.285	0.199	0.486	7.11
	垂直翻转	42.88	0.285	0.201	0.460	7.06
	田字格	40.97	0.281	0.196	0.471	7.16

注. 粗体表示最优值.

3.3 与现有模型对比实验

3.3.1 定量比较

与先进的模型在已见风格未见字符任务上的实验结果如表 6 所示.其中, w/CSDA 表示嵌入 CSDA 后相应的改进模型.可以看出,加入现有模型后,与基模型相比, CSDA 在 FID, MSE, LPIPS, PSNR 和 SSIM 指标上均有较大提升,与现有基于图像风格转换的汉字字体生成方法 CycleGAN^[20], SQ-GAN^[15]和 FUNIT^[32]相比,加入 CSDA 的小样本字体生成方法显著优于这些方法,充分地说明风格与内容解耦性能的提升对字体生成性能的提升具有重要作用;与较先进的小样本字体生成模型 LF-Font^[6]相比,加入 CSDA 后,本文所提基模型在 FID, MSE, LPIPS, PSNR 和 SSIM 这 5 个指标上显著优于 LF-Font^[6];与基模型 MX-Font, CG-GAN, LSG-FCST^[34]和 CF-Font^[33]相比,引入 CSDA 后,性能全面优于上述模型.实验结果验证了 CSDA 在已见风格未见字符任务上的优越性.

在未见风格字体生成任务上的性能对比如表 7 所示.由于 CycleGAN^[20]和 SQ-GAN^[15]等基于图像风格转换的汉字生成模型难以处理该情形,因

此在该部分实验中未将这 2 种模型作为基线模型进行对比.

表 6 与当前主流方法在已见风格未见字符任务上对比

模型	FID↓	MSE↓	LPIPS↓	PSNR↑	SSIM↑
CycleGAN ^[20]	73.73	0.234	0.328	6.42	0.401
SQ-GAN ^[15]	80.23	0.239	0.356	6.30	0.380
FUNIT ^[32]	137.87	0.261	0.440	5.91	0.329
LF-Font ^[6]	67.17	0.211	0.346	7.16	0.463
MX-Font ^[7]	58.35	0.211	0.316	6.86	0.428
CG-GAN ^[8]	49.89	0.211	0.306	6.88	0.428
LSG-FCST ^[34]	54.12	0.198	0.288	7.22	0.463
CF-Font ^[33]	31.26	0.173	0.257	7.54	0.501
MX-Font w/ CSDA	47.60	0.205	0.305	7.03	0.443
CG-GAN w/ CSDA	39.61	0.200	0.287	7.01	0.440
LSG-FCST w/ CSDA	37.61	0.191	0.271	7.33	0.485
CF-Font w/ CSDA	29.55	0.168	0.241	7.67	0.512

注. 粗体表示最优值.

表 7 与当前主流方法在未见风格任务上对比

任务	模型	FID↓	MSE↓	LPIPS↓	PSNR↑	SSIM↑
未见风格 已见字符	FUNIT ^[32]	111.42	0.268	0.455	5.79	0.353
	LF-Font ^[6]	64.58	0.186	0.334	7.34	0.492
	MX-Font ^[7]	44.62	0.197	0.294	7.19	0.477
	CG-GAN ^[8]	39.51	0.199	0.283	7.17	0.478
	LSG-FCST ^[34]	37.29	0.199	0.276	7.28	0.497
	CF-Font ^[33]	40.99	0.188	0.287	7.22	0.498
未见风格 未见字符	MX-Font w/ CSDA	38.01	0.190	0.289	7.35	0.492
	CG-GAN w/ CSDA	35.44	0.186	0.271	7.26	0.482
	LSG-FCST w/ CSDA	30.22	0.185	0.265	7.41	0.510
	CF-Font w/ CSDA	35.78	0.177	0.274	7.40	0.508
	FUNIT ^[32]	116.06	0.250	0.420	6.07	0.383
	LF-Font ^[6]	73.56	0.198	0.346	7.21	0.487
未见风格 未见字符	MX-Font ^[7]	45.24	0.205	0.291	6.97	0.455
	CG-GAN ^[8]	42.28	0.208	0.282	6.92	0.451
	LSG-FCST ^[34]	40.77	0.198	0.281	7.31	0.488
	CF-Font ^[33]	52.53	0.196	0.278	7.27	0.491
	MX-Font w/ CSDA	40.97	0.196	0.281	7.16	0.471
	CG-GAN w/ CSDA	37.91	0.200	0.269	7.01	0.453
未见风格 未见字符	LSG-FCST w/ CSDA	28.21	0.191	0.273	7.49	0.500
	CF-Font w/ CSDA	36.01	0.189	0.266	7.38	0.499

注. 粗体表示最优值.

表 7 中, 从未见风格已见字符任务上可以看出, 加入 CSDA 后, LSG-FCST^[34]在 FID, LPIPS, PSNR 和 SSIM 指标上取得最优的性能, 而在 MSE 指标上取得了次优的性能; 与 FUNIT 和 MX-Font 相比, CSDA 在 5 个指标上都优于这 2 种对比模型; 与 LF-Font^[6]相比, LSG-FCST w/ CSDA 在给出的 5 个指标上的性能显著优于 LF-Font, 特别是在基模型对比中, LF-Font^[6]在 PSNR 上取得了最好的结果; 而加入 CSDA 后, 均能实现较大程度反超 MX-Font^[7], LSG-FCST^[34]和 CF-Font^[33], 验证了 CSDA 的有效性; 与 CG-GAN^[8]相比, CSDA 在 FID, MSE, PSNR 和 SSIM 这 5 个指标上的性能显著优于 CG-GAN^[8], 而在 LPIPS 指标上取得了与

CG-GAN^[8]相近的性能. 实验结果表明, CSDA 在未见风格已见字符任务上具有优越性.

3.3.2 定性比较

表 7 中性能最佳的拓展模型 LSG-FCST w/ CSDA, 与对比方法中最先进的小样本汉字生成方法进行对比实验, 使用多种颜色的矩形框强调不同模型生成的字符的缺陷. 图 4a 所示为 CSDA 与对比模型在已见风格未见字符任务上的生成效果. 可以看出, CSDA 在该任务上的生成字体质量优于对比方法, 尤其是在保持汉字图像整体结构准确性和风格一致性方面表现出色; 相比之下, LF-Font^[6]和 MX-Font 在内容和风格的解耦上效果较好, 但存在部分局部细节缺失的情形; CG-GAN^[8]在

LF-Font ^[6]	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
MX-Font ^[7]	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
CG-GAN ^[8]	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
LSG-FCST ^[34]	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
CF-Font ^[33]	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
LSG-FCST w/CSDA	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链
GT	跋 论 透 邴	谤 褒 财 踌	贬 钵 辟 郑	辨 钵 铬 靛	饱 表 顿 链

a. 已见字体未见字符

LF-Font ^[6]	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
MX-Font ^[7]	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
CG-GAN ^[8]	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
LSG-FCST ^[34]	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
CF-Font ^[33]	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
LSG-FCST w/CSDA	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿
GT	辞 略 铅 设	蹠 踪 那 郅	郭 街 邱 迪	睬 郊 邦 避	裕 雁 逝 鹿

b. 未见字体未见字符

LF-Font ^[6]	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
MX-Font ^[7]	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
CG-GAN ^[8]	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
LSG-FCST ^[34]	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
CF-Font ^[33]	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
LSG-FCST w/CSDA	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但
GT	凹 程 掉 吩	岸 胺 耙 捕	安 棒 愁 蛾	啊 芭 拔 棒	藹 剥 茶 但

c. 未见字体已见字符

图 4 CSDA 加入 LSG-FCST 后与对比方法在 3 个任务上的字体生成效果比较

一些字体生成任务上风格转换不明显; LSG-FCST^[34]生成的字符在细节部分缠绕(如蓝框所示);与对比方法中最先进的模型 CF-Font^[33]相比, CSDA-Font 在汉字字体结构保持上更加出色. CSDA 及其对比模型在未见风格未见字体任务上的生成效果如图 4c 所示. 可以看出, 在大多数字体生成任务中, CSDA-Font 生成的字体质量更高, 结构更加完整. 类似的现象在未见风格未见字体任务上也可以观察到, 如图 4b 所示. 实验结果表明了 CSDA 的优越性.

4 结 语

实现字体风格和内容的有效解耦是小样本汉字字体生成的核心问题. 针对该问题, 本文提出一种基于 CLM 模块和 SDA 模块的风格与内容解耦方法 CSDA, 用于自适应地解耦汉字字体的风格与内容特征, 并可以嵌入当前最新的模型. 对所提模块及其相应模型进行大量实验的结果表明, CSDA 可以有效地实现汉字字体风格与内容解耦, 并显著地提升模型性能, 其在多类小样本字体生成任务中都优于文中对比的最先进的模型; 在降低训练数据集的情况下, CSDA 可以以插件的形式嵌入现有许多基模型并进一步提升其性能; CSDA 的通用性也在一系列的实验中得到验证. 未来, 将把 CSDA 应用于其他语言, 如韩语和日文字体生成, 以及研究更加高效的字体风格与内容解耦方式.

参考文献(References):

- [1] Zhang Y X, Zhang Y, Cai W B. Separating style and content for generalized style transfer[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8447-8455
- [2] Gao Y, Guo Y, Lian Z H, *et al.* Artistic glyph image synthesis via one-stage few-shot learning[J]. ACM Transactions on Graphics, 2019, 38(6): Article No.185
- [3] Xie Y C, Chen X Y, Sun L, *et al.* DG-font: deformable generative networks for unsupervised font generation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 5126-5136
- [4] Azadi S, Fisher M, Kim V, *et al.* Multi-content GAN for few-shot font style transfer[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7564-7573
- [5] Sun D Y, Ren T Z, Li C X, *et al.* Learning to write stylized Chinese characters by reading a handful of examples[C] //Proceedings of the 27th International Joint Conference on Artificial Intelligence. San Fransisco: Morgan Kaufmann, 2018: 920-927
- [6] Park S, Chun S, Cha J, *et al.* Few-shot font generation with localized style representations and factorization[C] //Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021: 2393-2402
- [7] Park S, Chun S, Cha J, *et al.* Multiple heads are better than one: few-shot font generation with multiple localized experts[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 13880-13889
- [8] Kong Y X, Luo C J, Ma W H, *et al.* Look closer to supervise better: one-shot font generation via component-based discriminator[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 13472-13481
- [9] Tang L C, Cai Y Y, Liu J M, *et al.* Few-shot font generation by learning fine-grained local styles[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 7885-7894
- [10] Liu W, Liu F Y, Ding F, *et al.* XMP-font: self-supervised cross-modality pre-training for few-shot font generation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 7895-7904
- [11] Chen T, Kornblith S, Norouzi M, *et al.* A simple framework for contrastive learning of visual representations[C] //Proceedings of the 37th International Conference on Machine Learning. New York: PMLR, 2020: 1597-1607
- [12] Ye M, Zhang X, Yuen P C, *et al.* Unsupervised embedding learning via invariant and spreading instance feature[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 6203-6212
- [13] Tian Y L, Krishnan D, Isola P. Contrastive multiview coding[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 776-794
- [14] He K M, Fan H Q, Wu Y X, *et al.* Momentum contrast for unsupervised visual representation learning[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 9726-9735
- [15] Zeng Jinshan, Chen Qi, Wang Mingwen. Self-supervised Chinese font generation based on square-block transformation[J]. Scientia Sinica Informationis, 2022, 52(1): 145-159(in Chinese)
(曾锦山, 陈琪, 王明文. 基于田字格变换的自监督汉字字体生成[J]. 中国科学: 信息科学, 2022, 52(1): 145-159)
- [16] Li B H, Hou Y T, Che W X. Data augmentation approaches in natural language processing: a survey[J]. AI Open, 2022, 3: 71-90
- [17] Bartoli A. Groupwise geometric and photometric direct image registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(12): 2098-2108
- [18] Tian Y. Master Chinese calligraphy with conditional adversarial networks[OL]. [2023-11-13]. <https://github.com/kaonashi-tyc/zi2zi>
- [19] Chang B, Zhang Q, Pan S Y, *et al.* Generating handwritten

- Chinese characters using CycleGAN[C] //Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2018: 199-207
- [20] Zhu J Y, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2242-2251
- [21] 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: MIT Press, 2014: 2672-2680
- [22] Liu Yu, Ding Yang, Khalid F B, *et al.* Unsupervised font generation network integrating content and style representation[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(5): 865-876(in Chinese)
(刘宇, 丁阳, Khalid F B, 等. 联合内容和风格表示的无监督字体生成网络[J]. 计算机辅助设计与图形学学报, 2025, 37(5): 865-876)
- [23] Zeng J S, Chen Q, Liu Y X, *et al.* StrokeGAN: reducing mode collapse in Chinese font generation via stroke encoding[C] //Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021: 3270-3277
- [24] Wang Cunrui, Ding Yang, Liu Yu, *et al.* Chinese font generation from stroke semantic and attention mechanism[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(8): 1229-1237(in Chinese)
(王存睿, 丁阳, 刘宇, 等. 融合笔画语义和注意力机制的汉字字体生成算法[J]. 计算机辅助设计与图形学学报, 2022, 34(8): 1229-1237)
- [25] Zeng J S, Wang Y F, Chen Q, *et al.* StrokeGAN+: few-shot semi-supervised Chinese font generation with stroke encoding[OL]. [2023-11-13]. <https://arxiv.org/abs/2211.06198>
- [26] Zhou J, Wang Y F, Yuan Y Y, *et al.* SGCE-font: skeleton guided channel expansion for Chinese font generation[OL]. [2023-11-13]. <https://arxiv.org/abs/2211.14475>
- [27] Zeng J S, Tu L, Zhou J, *et al.* Enhancing Chinese calligraphy generation with contour and region-aware attention[C] //Proceedings of the International Joint Conference on Neural Networks. Los Alamitos: IEEE Computer Society Press, 2023: 1-9
- [28] Jiang Y, Lian Z H, Tang Y M, *et al.* DCFont: an end-to-end deep Chinese font generation system[C] //Proceedings of SIGGR-APH Asia 2017 Technical Briefs. New York: ACM Press, 2017: 1-4
- [29] Park S, Chun S, Cha J, *et al.* Few-shot font generation with weakly supervised localized representations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(3): 1479-1495
- [30] Kuhn H W. The Hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955, 2(1-2): 83-97
- [31] Gretton A, Bousquet O, Smola A, *et al.* Measuring statistical dependence with Hilbert-Schmidt norms[C] //Proceedings of the 16th International Conference on Algorithmic Learning Theory. Heidelberg: Springer, 2005: 63-77
- [32] Liu M Y, Huang X, Mallya A, *et al.* Few-shot unsupervised image-to-image translation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 10550-10559
- [33] Wang C, Zhou M, Ge T Z, *et al.* CF-font: content fusion for few-shot font generation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 1858-1867
- [34] Li Y M, Lin G F, He M L, *et al.* Layer similarity guiding few-shot Chinese style transfer[J]. The Visual Computer, 2024, 40(4): 2265-2278
- [35] Wang Z, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612
- [36] Zhang R, Isola P, Efros A A, *et al.* The unreasonable effectiveness of deep features as a perceptual metric[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 586-595
- [37] Heusel M, Ramsauer H, Unterthiner T, *et al.* GANs trained by a two time-scale update rule converge to a local nash equilibrium[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2017: 6629-6640
- [38] Kingma D P, Ba J. Adam: a method for stochastic optimization[OL]. [2023-11-13]. <https://arxiv.org/pdf/1412.6980>