

基于 OCR 与改进 CycleGAN 的汉字识别与字体风格生成系统

罗雪姣 李鲁英 许树洲

武汉设计工程学院智能工程学院 湖北 武汉 430000

【摘要】：传统文字处理方式主要依赖人工输入与字体编辑，不仅效率较低，而且难以实现复杂字体风格的自动生成。针对上述问题，本文设计并实现了一套基于深度学习的汉字字体识别与风格生成系统。系统集成 PaddleOCR 文字识别模块以及基于改进 CycleGAN 的书法字体生成模块，并通过 PyQt5 构建可视化交互平台，实现文字识别与字体转换、书法风格生成的统一处理流程。实验结果表明，系统字符级识别准确率达到 97%，字段级识别准确率达到 92%。字体风格生成方面，在保持汉字结构稳定的同时生成具有明显书法风格特征的字体图像。

【关键词】：PaddleOCR；CycleGAN；字体生成；PyQt5

DOI:10.12417/3041-0630.26.07.062

1 引言

汉字作为中华文化的重要载体，其字体形态具有丰富的艺术表现形式。在数字化时代，大量文本信息需要进行自动识别、转换与排版处理，因此研究高效的汉字识别与字体生成技术具有重要意义。

随着深度学习的发展，基于卷积神经网络的 OCR 算法显著提升了图像中的文字识别精度，使复杂场景中的文字识别成为可能。另一方面，生成对抗网络（GAN）在图像生成领域取得了突破性进展。通过对抗训练机制，生成器能够学习数据分布并生成具有真实感的图像。GAN 技术在图像风格迁移、图像修复和字体生成等领域得到了广泛应用。

然而，汉字结构复杂、笔画数量多、风格变化丰富，使得字体风格生成任务具有较高难度。因此，设计一种结合文字识别、字体转换与书法生成的综合系统具有重要研究价值。

2 系统方案设计

本文提出的汉字字体处理系统主要由四个核心模块组成：

- （1）OCR 文字识别与字体转换模块
- （2）书法字体生成模块
- （3）可视化交互界面模块

系统采用模块化设计，各模块之间通过统一接口进行数据交互，从而提高系统扩展性。

3 文字识别与字体转换模块设计与实现

文字识别与字体转换模块主要任务是从输入图像中准确

提取文本信息，并将图像形式的文字转换为可编辑的文本数据，并实现不同字体样式之间的转换与展示。

3.1 OCR 识别算法

本文采用 PaddleOCR 框架实现上述功能，该框架在中文文本识别任务中具有较高的识别精度和稳定性。该框架主要由文本检测模型、方向分类模型和文本识别模型组成。文本检测模型 ch_PP-OCRv3_det_infer 用于检测图像中的文字区域；文本方向分类模型 ch_ppocr_mobile_v2.0_cls_infer 用于判断文字方向并进行校正；文本识别模型 ch_PP-OCRv3_rec_infer 用于将检测到的文字区域转换为文本内容。

在 OCR 识别之前，对输入图像进行预处理，能显著提升低对比度、光照不均、模糊文档图片的文字检测和识别准确率。系统采用以下两种预处理方法：（1）CLAHE 对比度增强，通过局部直方图均衡化增强图像细节，同时限制对比度过度放大，从而减少噪声影响。（2）自适应阈值分割，根据局部区域亮度自动计算阈值，使得文字区域能够从背景中有效分离。

3.2 文字识别与转换实验结果

在 OCR 识别实验中，选取 100 张人民日报图像进行测试，并统计字符识别准确率和字段识别准确率。实验结果表明，系统在字符级别上的平均识别准确率达到约 97%，而字段级识别准确率约为 92%。在大多数测试图像中，系统能够正确识别新闻标题和正文内容，满足系统应用需求。

为了实现不同字体样式之间的转换，系统利用操作系统字体库提供的字体资源，并通过图形界面实现实时字体切换。图 1 为文字识别并转换为黑体后的结果。

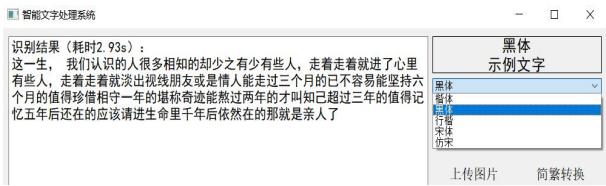


图1 字体样式转换结果

4 基于改进 CycleGAN 的书法字体生成模型

4.1 数据集构建与预处理

在字体风格生成任务中，训练数据的质量对模型性能具有重要影响。为了实现印刷体汉字向书法字体的风格迁移，本文构建了一个双域字体数据集，包括印刷字体域和书法字体域两个部分。其中，源域数据为宋体汉字图像，目标域数据为颜真卿书法字体图像。两类字体图像均经过统一处理，以保证模型训练的稳定性。

数据集中每类字体包含约 5000 张单字图像，所有图像均经过灰度化处理并统一缩放至 128×128 像素，以减少不同分辨率对模型训练造成的影响。为了提高模型的泛化能力，在训练阶段对数据进行随机增强处理，包括随机裁剪、轻微旋转和亮度扰动等操作，从而扩充数据分布范围。训练集与测试集的划分比例为：训练集：90%，测试集：10%。

由于不同字体之间通常不存在严格的一一对应关系，因此本文采用非配对数据训练策略，即从两个域中随机抽取样本进行训练，使模型能够学习字体风格之间的映射关系。

4.2 改进 CycleGAN 模型结构

CycleGAN 是一种典型的无配对图像到图像转换模型，其核心思想是通过循环一致性约束，使模型在缺乏成对训练数据的情况下仍然能够学习稳定的映射关系。本文在 CycleGAN 基础上进行结构改进，以提高汉字结构保持能力和书法风格表达能力。

模型整体结构由两个生成器和两个判别器组成：

生成器 $G: X \rightarrow Y$

生成器 $F: Y \rightarrow X$

判别器 D_X

判别器 D_Y

生成器 G 的作用是将印刷体汉字转换为书法字体，而生成器 F 则负责将书法字体重新映射回印刷体域，从而形成循环一致性结构。

该结构可表示为：

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$

$$y \rightarrow F(y) \rightarrow G(F(y)) \approx y$$

通过该循环约束，模型可以在缺乏配对数据的情况下稳定学习两个域之间的映射关系。

4.3 增强型 U-Net 生成器

在原始 CycleGAN 模型中，生成器通常采用残差网络结构。然而，在汉字字体生成任务中，汉字笔画结构较为复杂，如果仅依赖深层残差网络容易导致结构信息丢失。为了解决这一问题，本文采用增强型 U-Net 生成器结构。

U-Net 结构由编码器和解码器两部分组成，并通过跳跃连接实现低层特征与高层特征之间的信息传递。编码器部分通过多层卷积和下采样操作提取图像特征，而解码器则通过反卷积逐步恢复图像分辨率。为了保持汉字笔画结构信息，在编码器与解码器之间引入跳跃连接，通过跨层特征融合，可以有效减少汉字笔画结构丢失问题，从而提高字体生成质量。

4.4 多尺度 PatchGAN 判别器

在传统 GAN 结构中，判别器通常对整幅图像进行真假判断。然而，对于字体生成任务而言，局部笔画细节同样非常重要。为此，本文采用 PatchGAN 判别器结构，通过对图像局部区域进行判别，提高模型对细节的识别能力。

PatchGAN 判别器将输入图像划分为多个局部图像块，并分别对每个图像块进行真假判断。本文设计了三尺度判别结构，即在不同分辨率下同时进行判别：原始分辨率：128×128；中等分辨率：64×64；低分辨率：32×32。

通过多尺度判别机制，模型能够同时关注整体结构和局部细节，从而生成更加真实的书法字体图像。

4.5 损失函数设计

为了保证生成图像既具有书法风格特征，又能够保持原始汉字结构，本文在 CycleGAN 框架基础上设计了多目标损失函数，包括对抗损失、循环一致性损失和身份保持损失。

为了保证生成图像既具有书法风格特征，又能够保持原始汉字结构，本文在 CycleGAN 框架基础上设计了多目标损失函数，包括对抗损失、循环一致性损失和身份保持损失和感知损失。

(1) 对抗损失：对抗损失用于约束生成器生成的图像能够欺骗判别器，其表达式为：

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log (1 - D_Y(G(x)))]$$

该损失函数使生成图像逐渐接近真实书法字体分布。

(2) 循环一致性损失：为了保证生成图像在结构上保持原始汉字特征，引入循环一致性损失：

$$L_{\text{cycle}}(G, F) = E_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]$$

该损失函数能够有效防止生成器出现结构扭曲问题。

(3) 身份保持损失：在字体转换任务中，如果输入图像本身已经属于目标域，则生成器应尽可能保持其不变。因此，引入身份保持损失：

$$L_{\text{identity}}(G, F) = E_{y \sim p_{\text{data}}(y)}[\|G(y) - y\|_1] + E_{x \sim p_{\text{data}}(x)}[\|F(x) - x\|_1]$$

该损失有助于稳定模型训练过程。

(4) 感知损失：感知损失利用 VGG 网络提取特征进行比较：

$$L_{\text{perc}} = \sum_i \|\phi_i(x) - \phi_i(G(x))\|_2$$

其中 ϕ_i 表示 VGG 网络第 i 层特征。

最终损失函数为：

$$L = L_{\text{GAN}} + \lambda_1 L_{\text{cycle}} + \lambda_2 L_{\text{identity}} + \lambda_3 L_{\text{perc}}$$

$$\lambda_1 = 20, \lambda_2 = 1, \lambda_3 = 0.1$$

通过调节不同损失项的权重，可以在字体风格迁移效果与结构保持之间取得良好的平衡。

4.6 模型训练策略

在模型训练阶段，本文采用交替优化策略对生成器和判别器进行更新。训练过程中，首先利用生成器生成书法字体图像，然后通过判别器判断生成图像与真实书法字体之间的差异，并根据损失函数反向传播更新模型参数。

训练过程中使用 Adam 优化器，其参数设置为，学习率：0.0002， β_1 : 0.5， β_2 : 0.999

模型训练共进行 400 个 Epoch，在训练后期逐渐降低学习率以提高模型收敛稳定性。

4.7 字体风格生成结果与分析

实验结果表明，经过充分训练后，模型能够较好地学习书

法字体的笔画结构特征，并在保持汉字结构稳定的同时生成具有明显书法风格的字体图像。

点击书法生成功能按键，运行外部子进程，选择要转换的印刷体宋体单字图片，并点击开始转换，子进程 GUI 界面会展示出生成结果，并提示已将生成结果图保存到了指定目录中，实现效果如图 2 所示。

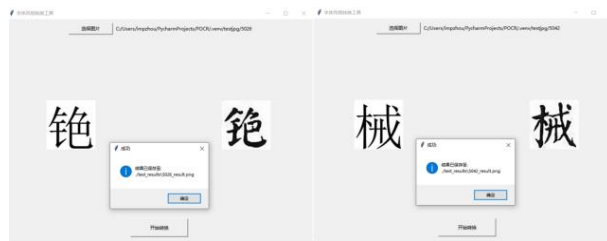


图 2 宋体生成颜真卿体图

SSIM（结构相似性），通过计算两张图片在结构特征、亮度及对比度三个维度的相似性来评估图像质量，既可用于传统图像处理中的失真分析，也可作为生成式模型输出图像的客观评价标准。SSIM 数值在 0 到 1 之间，其中越接近 1，就代表生成质量越好，反之越差。使用其中 47 张测试集中生成图片和对应的颜体书法图片进行 SSIM 评估，得出平均 SSIM 值为：0.40814665，其中最高为 0.7826522，最低 0.15312505。

实验结果表明：模型能够学习书法风格特征，大部分字体生成质量较好，个别样本存在笔画缺失问题。

5 总结

本文提出了一种基于深度学习的汉字字体识别与风格生成系统。在文字识别方面，系统采用 PaddleOCRv3 模型，并结合 CLAHE 对比度增强与自适应阈值预处理，提高复杂图像环境下的识别准确率。在字体风格生成方面，本文基于 CycleGAN 框架提出增强型 U-Net 生成器与多尺度判别器结构，并引入四元损失函数，模型在印刷体到颜真卿书法体的转换任务中获得较好的生成效果。

未来研究可通过扩大训练数据集和引入更先进的生成模型进一步提升字体生成质量。

参考文献：

[1] 吴建磊,杨慧炯.基于改进 CycleGAN 网络的图像风格迁移技术研究[J].信息技术与信息化,2025,(02):79-82.
 [2] 李云红,段姣姣,苏雪平,等.基于改进生成对抗网络的书法字生成算法[J].浙江大学学报(工学版),2023,57(07):1326-1334+1459.
 [3] 曾悦,马明栋.基于 Tesseract_OCR 文字识别的研究[J].计算机技术与发展,2021,31(11):76-80.
 [4] 赵明,王存睿,战国栋.基于改进条件生成对抗网络的字体风格迁移算法[J].大连民族大学学报,2024,26(01):57-61.
 [5] 谭宏卫.基于生成式对抗网络的无监督图像生成研究[D].西南大学,2020.