

# 基于注意力机制与三元组损失的少样本字体生成方法

闫肖月, 牟大中

北京印刷学院信息工程学院, 北京

收稿日期: 2025年12月25日; 录用日期: 2026年1月22日; 发布日期: 2026年1月29日

## 摘要

针对跨语言少样本字体生成任务中风格特征表达不足和字形结构迁移不够精准的问题, 本文提出一种融合注意力机制与三元组损失约束的少样本字体生成方法。该方法在对风格字体图像进行特征提取的基础上, 引入上下文增强注意力机制, 通过结合卷积操作与上下文感知特征变换模块, 实现对字体笔画细粒度局部特征与整体结构信息的联合建模。随后, 解码器对融合特征进行重建, 生成目标字体图像。此外, 本文引入三元组损失函数, 从内容一致性与风格判别性两个维度对生成结果进行约束, 以进一步增强字形结构一致性和风格表达能力。实验结果表明, 该方法在少样本及跨语言字体生成任务中能够有效提升字体的结构保真度和风格迁移质量。

## 关键词

少样本学习, 跨语言字体生成, 上下文增强注意力, 三元组损失

# Few-Shot Font Generation Method Based on Attention Mechanism and Triplet Loss

Xiaoyue Yan, Dazhong Mu

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: December 25, 2025; accepted: January 22, 2026; published: January 29, 2026

## Abstract

To address the problems of insufficient style feature representation and inaccurate glyph structure transfer in cross-lingual few-shot font generation tasks, this paper proposes a few-shot font generation method that integrates an attention mechanism with triplet loss constraints. Based on feature extraction from style font images, a context-enhanced attention mechanism is introduced, which

combines convolutional operations with a context-aware feature transformation module to jointly model fine-grained local stroke features and global structural information of glyphs. Subsequently, the decoder reconstructs the fused features to generate target font images. In addition, a triplet loss function is incorporated to constrain the generated results from both content consistency and style discriminability perspectives, thereby enhancing glyph structural consistency and style expression capability. Experimental results demonstrate that the proposed method effectively improves structural fidelity and style transfer quality in few-shot and cross-lingual font generation tasks.

## Keywords

Few-Shot Learning, Cross-Lingual Font Generation, Context-Enhanced Attention, Triplet Loss

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

字体作为视觉传达与信息表达的重要载体,在出版印刷、品牌设计、数字媒体、网页排版及人机交互等领域具有广泛应用[1]。除了传递文本信息外,字体还承载着文化符号、历史记忆和视觉美学特征,对品牌识别、用户体验以及信息可读性均起到关键作用。同时,不同字体风格能够强化情感表达、视觉层次和排版效果,是视觉设计与信息传达不可或缺的重要元素。随着数字化和全球化的推进,多语言文本在电子媒介中的使用日益增加,对字体生成、跨语言排版以及多语种字库构建提出了更高的需求[2]。然而,传统字体设计流程需依赖设计师逐字完成笔画构形、曲线光滑化、字重平衡以及字面度量调整等繁琐步骤,不仅劳动密集、耗时长,也难以在多语种大字符集上保持风格一致性和骨架结构稳定性[3]。

近年来,基于深度学习的少样本字体生成技术成为提升字体生产效率的重要途径。该类方法能够利用极少量的参考样本学习字体风格规律,重建完整字库,从而有效降低字体制作成本并提高生成效率。在单语言字体生成任务中,现有方法如 EasyFont [4]能够基于少量手写样本学习笔画形态的统计分布,并在大规模字符集上实现较为稳定的风格一致性生成。LSCG-FCST [5]则通过多层次相似度计算,进一步提升内容特征与风格特征在不同层级上的细粒度匹配能力。这些研究在单语言字体生成中取得了显著进展,为多语言字体生成提供了坚实基础。然而,在跨语言字体生成场景中,不同文字体系在几何结构、笔画组合方式、部件排列规则及视觉规范等方面存在显著差异,使单语言生成模型在结构保持、笔画骨架一致性和风格迁移上面临重大挑战。这些差异导致原有方法难以在跨语言任务中有效建模特定文字体系的结构特征,并限制了风格迁移的精细化与多样性表达。为解决上述问题,本文提出了一种基于注意力增强与三元组约束的少样本字体生成方法,主要体现在以下两个方面:

(1) 在风格特征提取阶段,引入一种基于上下文增强的注意力机制(Contextual Transformation Attention, CoT-Attn)。该模块首先利用局部卷积操作捕获笔画纹理、线条粗细等局部细节特征,随后通过自注意力结构建模字符整体布局及跨笔画的长程依赖关系,并结合上下文感知的特征变换与自适应融合策略,实现局部细节与全局结构信息的动态整合,从而增强模型对多尺度字体风格特征的表达能力。

(2) 在模型训练阶段,引入三元组损失(Triplet Loss)构建内容-风格对比约束,将同一字符的同风格样本作为正样本、同字符的不同风格样本作为负样本,以缩小风格一致样本间的特征距离、扩大不同风

格样本间的差异, 从而促进内容特征与风格特征的有效解耦, 提升跨语言字体生成中的结构稳定性与风格一致性。

## 2. 研究现状

随着深度学习技术的快速发展, 字体风格迁移与生成领域迎来了显著的技术变革[6]。2016年, 田雨辰等人[7]首次将卷积神经网络引入字体风格迁移任务, 验证了深度学习方法在字体风格建模与迁移中的可行性, 为后续相关研究奠定了基础。此后, 基于深度神经网络的字体生成与风格迁移方法不断涌现, 逐步推动字体生成从传统规则驱动向数据驱动范式转变。

### 2.1. 少样本字体生成

在少样本字体生成方面, 研究者围绕风格特征建模、结构保持与生成稳定性等问题开展了大量探索。Dai 等[7]提出基于多尺度判别器的书法风格迁移方法, 通过在不同尺度上约束生成结果, 有效提升了字体整体轮廓与局部笔画细节的协调性。Zhang 等人[8]引入密集块与胶囊网络构建中文书法字符生成模型, 在一定程度上增强了模型对字符内部空间结构关系的表达能力。Liu 等人[9]提出基于风格引导判别器的字体生成方法, 通过构建字体样式空间, 实现了在少量样本条件下较为稳定的风格迁移效果。

针对内容与风格解耦问题, 刘宇等人[10]提出无监督的内容-风格分离模型, 通过密集语义对应与多尺度风格判别机制, 提高了生成字体在结构保持与风格一致性方面的表现。赵宏等人[11]基于 Res2Next 架构构建多尺度特征提取框架, 并引入对比学习与 Shuffle Attention 机制, 在手写字体生成任务中改善了细节还原效果。围绕名家书法风格扩展问题, 张嘉等人[12]结合 Zi2zi 与 CycleGAN, 构建了结构引导的风格生成模型, 实现了书法风格字库的有效扩展。此外, 周雅琴等人[13]通过文献计量分析指出, 深度学习、风格迁移与生成对抗网络已成为生成式字体设计领域的主要研究热点。

总体来看, 现有少样本字体生成方法在单语言场景下已能够较好地建模字体风格特征, 并实现一定程度的结构保持。然而, 这类方法普遍依赖于语言内部的结构一致性假设, 其在跨文字体系任务中的适应能力仍然有限。

### 2.2. 跨语言少样本字体生成

针对跨语言字体生成中不同文字体系在结构复杂度、笔画组合方式和空间布局规则上的显著差异, 研究者进一步提出了一系列跨语言少样本字体生成方法。FTransGAN [14]通过多层次注意力机制提取语言无关的高层风格特征, 实现了不同字体风格之间的基础对齐, 但在处理结构复杂或笔画密集字符时, 细粒度结构保持能力仍有不足。MF-Net [15]针对不同语言在结构复杂度上的差异, 引入语言复杂度感知的跨语言跳跃连接结构, 使模型能够根据字体特性自适应调整内容表征, 从而提升跨语言迁移的稳定性。MA-Font [16]构建了包含内容自适应模块、多头注意力模块和协同自适应模块的端到端框架, 通过多层次特征交互增强跨语种风格迁移效果, 在一定程度上提高了笔画保真度与风格一致性。FCA-GAN [17]引入全域卷积注意力机制, 对跨尺度笔画与纹理依赖关系进行建模, 改善了生成字体中局部细节连贯性与整体风格统一性。

总体而言, 现有跨语言字体生成方法多通过注意力机制、结构对齐或自适应特征建模策略缓解不同文字体系之间的结构差异, 在内容与风格融合方面取得了一定进展。然而, 在极少样本条件下, 这类方法仍面临结构细节保持不足、复杂字形泛化能力有限等问题, 跨语言字体生成的稳定性与一致性仍有待进一步提升。

### 3. 本文方法

#### 3.1. 生成对抗网络

生成对抗网络提供了一种无需依赖大量标注数据即可学习深度特征表示的方法, 其核心思想是通过生成器(Generator)和判别器(Discriminator)两个网络的对抗训练过程, 实现模型性能的优化[18]。具体而言, 生成器通过学习输入数据的分布, 生成与真实样本尽可能接近的输出; 而判别器的目标是区分生成器生成的输出与真实数据。生成器和判别器通过不断对抗优化, 最终使生成器能够输出高质量的结果。生成对抗网络的核心训练过程实际上是一种极小极大博弈, 其训练目标如下所示:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

其中  $x$  表示真实样本数据,  $z$  表示输入的噪声随机函数,  $D(x)$ 表示判别器判断是否为真实样本数据的概率,  $D(G(z))$ 表示判别器判断是否为生成样本的概率。

#### 3.2. 网络模型

本文提出的网络模型整体结构如图 1 所示, 由生成器与判别器两部分组成。生成器由内容编码器、风格编码器和解码器三个模块构成。内容编码器以内容字体图像作为输入, 用于提取字符的结构与语义特征; 风格编码器以目标风格字体图像为输入, 通过多样本联合建模的方式提取稳定且具有代表性的风格特征。通过将内容特征与风格特征进行拼接, 输入至解码器中进行重建, 从而生成兼具目标风格与内容结构的字体图像。

在对抗训练阶段, 生成图像分别与对应的内容图像和风格图像一同输入至内容判别器和风格判别器, 通过双判别器结构从结构保持与风格一致性两个维度对生成器施加约束。该对抗学习机制促使生成器在保持字符结构稳定性的同时, 更准确地迁移目标字体风格特征, 从而提升生成字体在结构保真度与风格一致性方面的整体性能。

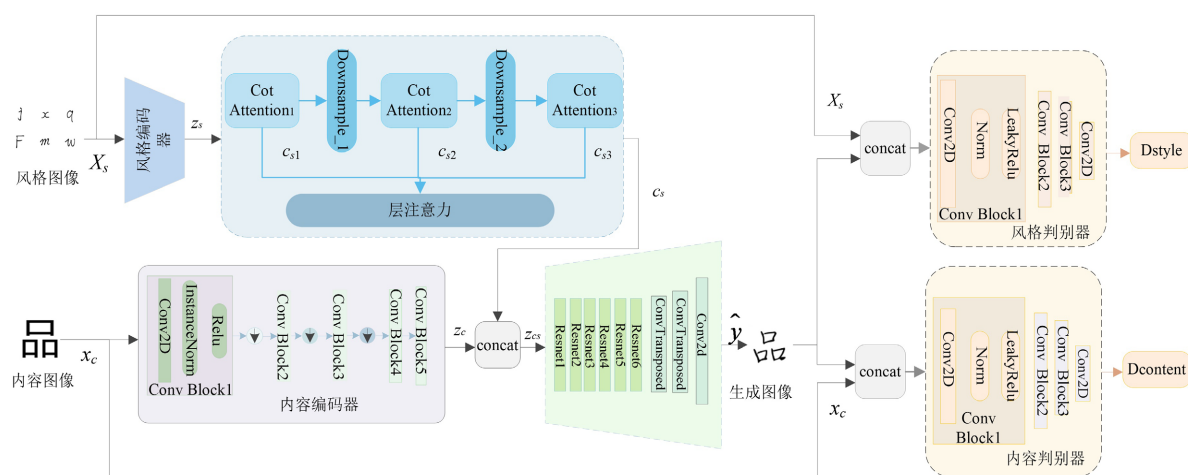


Figure1. Network structure diagram

图 1. 网络结构图

##### 3.2.1. 生成器

内容字体图像  $x_c$  经内容编码器处理后得到字体图像特征, 记为  $z_c$ 。风格字体图像  $X_s = \{x_{s_i}\}_{i=1}^6$  经与内

容编码器结构相同的风格编码器进行特征提取, 得到初始风格特征  $z_s$ 。为增强风格特征在空间结构与细节层面的表达能力, 本文在风格编码阶段引入层级注意力增强模块, 对风格特征  $z_s$  进行进一步优化处理。

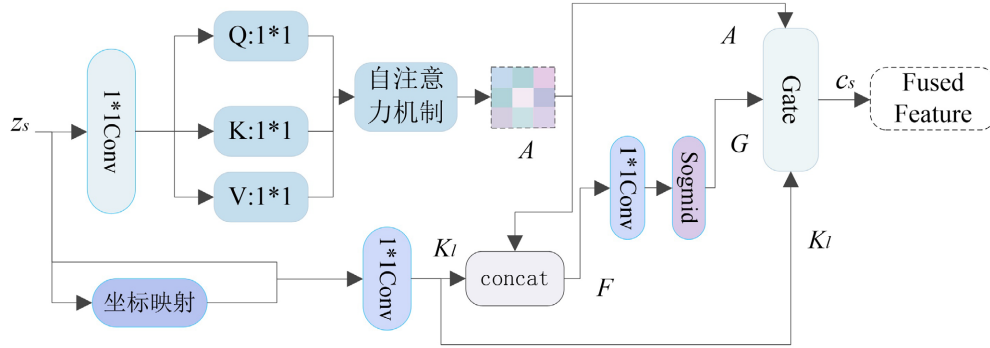


Figure 2. CoT-Attn network architecture diagram

图 2. CoT-Attn 网络结构图

为提升模型对字形空间结构的建模能力, 在风格特征提取阶段引入了 CoT-Attn 模块, 其网络结构图如图 2 所示。该模块通过显式引入二维归一化坐标映射, 以对字体的空间位置信息进行建模。显式引入二维坐标映射能够为字体风格特征提供连续的空间参考, 使模型在特征学习过程中感知笔触纹理、粗细变化等风格属性在字形空间中的分布关系, 从而在风格迁移过程中提升空间一致性。该坐标根据特征图空间尺寸自动生成, 其中  $X, Y \in [-1, 1]$  分别表示水平与垂直方向的归一化位置编码。随后, 将位置坐标与风格特征  $z_s$  在通道维度拼接, 并通过  $1 \times 1$  卷积完成通道映射, 用于生成具备位置感知能力的局部上下文风格特征表示  $K_l$ 。

$$K_l = \text{Conv}(z_s, X, Y) \quad (2)$$

同时, 为建模风格特征内部的长程依赖关系, 模块并行引入自注意力机制, 对输入特征  $z_s$  分别进行 Query、Key 与 Value 映射, 并通过注意力权重计算获得全局风格特征表示  $A$ 。

$$Q = \text{Conv}_Q(z_s), K = \text{Conv}_K(z_s), V = \text{Conv}_V(z_s) \quad (3)$$

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

为实现局部与全局信息的自适应融合, CoT-Attn 模块引入门控机制对两类特征进行动态融合。具体而言, 首先将特征  $K_l$  与特征  $A$  在通道维度拼接得到特征表示  $F$ 。随后通过卷积操作及 Sigmoid 激活函数生成空间自适应的门控权重。该权重用于调控不同空间位置上局部风格特征与全局风格特征的贡献, 在不破坏内容结构的前提下强化整体风格一致性。通过上述方式, 模块在保证全局风格协调性的同时, 有效保留字体的局部结构细节。

$$G = \text{Sigmoid}(\text{Conv}(F)) \quad (5)$$

$$c_{s1} = G \odot A + (1 - G) \odot K_l \quad (6)$$

其中  $\odot$  表示逐元素加权。

在获得具备空间感知能力的风格特征表示后, 为进一步增强风格在不同尺度下的稳定表达能力, 本文引入多尺度风格表示。具体而言, 通过下采样操作获得更低空间分辨率的层级特征, 并分别输入特征增强模块, 得到对应的增强特征表示  $c_{s2}$ 、 $c_{s3}$ 。在此基础上, 引入层注意力机制对各层风格特征进行自适应



加权融合, 动态调节不同尺度特征在最终风格表达中的贡献, 从而获得融合后的风格特征表示  $c_s$ 。

$$c_s = w_1 \cdot c_{s1} + w_2 \cdot c_{s2} + w_3 \cdot c_{s3} \quad (7)$$

其中  $w_1$ 、 $w_2$  和  $w_3$  为权重系数。

解码器将内容编码器提取的内容特征  $z_c$  与融合后的风格特征  $c_s$  进行特征拼接, 并通过多尺度卷积与逐级上采样操作, 逐步恢复字形的空间分辨率与细节结构, 最终生成在内容结构与风格属性上均与目标字体保持一致的字形图像  $\hat{y}$ 。

### 3.2.2. 判别器

本文采用双分支判别器结构, 由内容判别器和风格判别器组成, 分别从结构一致性和风格一致性两个角度对生成字体进行评价。两个判别器均基于 PatchGAN 架构, 由多层卷积块堆叠构成, 每个卷积块包含卷积、BatchNorm 和 LeakyReLU 激活, 最终通过局部判别输出评估图像的真实性[5]。通过双判别器的协同训练, 生成器能够同时保持字形结构与风格特征的一致性, 从而提升生成字体的整体质量。

### 3.3. 损失函数

本文的损失函数由对抗损失、L1 损失和三元组损失三部分组成。其中, 对抗损失用于提升生成字体的整体真实性; L1 损失在像素层面约束生成结果与目标之间的差异; 三元组损失则通过特征距离约束, 进一步强化生成字体在风格层面的判别性与一致性。

#### 3.3.1. 对抗损失

对抗损失由内容对抗损失和风格对抗损失两部分组成, 分别用于约束生成结果在字形细节与风格上与目标字体保持一致, 其定义如下:

$$L_{\text{GAN}} = L_{\text{adv}_c} + L_{\text{adv}_s} \quad (8)$$

内容对抗损失由内容生成损失和内容判别损失两部分组成。其中, 内容生成损失用于约束生成图像在结构层面逼近内容图像, 而内容判别损失则用于引导判别器区分生成图像与真实内容图像之间的结构差异, 其损失函数定义如下:

$$L_{\text{adv}_c} = L_{\text{content}_G} + L_{\text{content}_D} \quad (9)$$

$$L_{\text{content}_G} = -E_{\hat{y}, x_c \sim P(\hat{y}, x_c)} [D_{\text{content}}(\hat{y}, x_c)] \quad (10)$$

$$L_{\text{content}_D} = -E_{y, x_c \sim P(y, x_c)} [\min(0, D_{\text{content}}(y, x_c) - 1)] - E_{\hat{y}, x_c \sim P(\hat{y}, x_c)} [\min(0, -D_{\text{content}}(\hat{y}, x_c) - 1)] \quad (11)$$

其中  $y$  为目标图像,  $\hat{y}$  为生成图像,  $x_c$  为内容图像,  $D_{\text{content}}$  为内容判别器。

风格对抗损失由风格生成损失和风格判别损失两部分组成。风格生成损失促使生成图像学习风格字体图像的特征, 风格判别损失则用于判断生成图像是否具备与目标图像一致的风格特征, 其损失函数定义如下:

$$L_{\text{adv}_s} = L_{\text{style}_G} + L_{\text{style}_D} \quad (12)$$

$$L_{\text{style}_G} = -E_{\hat{y}, x_s \sim P(\hat{y}, x_s)} [D_{\text{style}}(\hat{y}, x_s)] \quad (13)$$

$$L_{\text{style}_D} = -E_{y, x_s \sim P(y, x_s)} [\min(0, D_{\text{style}}(y, x_s) - 1)] - E_{\hat{y}, x_s \sim P(\hat{y}, x_s)} [\min(0, -D_{\text{style}}(\hat{y}, x_s) - 1)] \quad (14)$$

其中为  $y$  目标图像,  $\hat{y}$  为生成图像,  $x_s$  为风格图像,  $D_{\text{style}}$  为风格判别器。

### 3.3.2. L1 损失函数

为了提高字体生成的稳定性, 本文引入 L1 损失来衡量生成图像和参考图像之间的像素级误差, 从而确保生成图像在局部特征上与目标图像保持一致。其定义如下:

$$L_{L1} = E_{\hat{y}, y \sim P(\hat{y}, y)} [\|\hat{y} - y\|_1] \quad (15)$$

其中  $y$  为目标图像,  $\hat{y}$  为生成图像。

### 3.3.3. 三元组损失函数

三元组损失函数最初用于度量学习任务中, 通过比较锚点图像、同类正样本图像和不同类别负样本图像的特征距离, 提升模型在特征空间中对类别差异的判别能力。在字体生成任务中, 为了进一步增强生成字体在特征空间对目标风格的判别能力, 本文将三元组损失应用于生成字体  $\hat{y}$  的特征约束, 使其更接近于目标字体  $y$ , 同时远离内容字体  $x_c$ 。在具体计算过程中, 生成字体、目标字体和内容字体的特征均通过 ResNet 网络进行提取, 得到对应的图形特征。基于欧氏距离的三元组损失定义如下:

$$L_{\text{triplet}} = \max \{d[f(\hat{y}), f(y)] - d[f(\hat{y}), f(x_c)] + \text{margin}, 0\} \quad (16)$$

其中,  $f(\cdot)$  表示通过预训练的 ResNet 网络提取的字体特征。该网络在训练过程中参数保持不更新, 仅作为特征提取器用于构建三元组损失, 以约束生成字体与目标字体之间的风格相似性,  $d$  为欧氏距离,  $\text{margin}$  为间隔超参数, 用于控制正负样本之间的最小距离差。

### 3.3.4. 整体损失函数

综上所述, 根据式(8)~(16), 本文模型的整体损失函数定义如下:

$$L_{\text{total}} = \min_G \max_D \lambda_{\text{adv}} (L_{\text{GAN}}) + \lambda_{L1} L_{L1} + \lambda_{\text{triplet}} L_{\text{triplet}} \quad (17)$$

其中  $\lambda_{\text{adv}}$ 、 $\lambda_{L1}$  和  $L_{\text{triplet}}$  分别为损失项的可学习权重, 用于在训练中实现自适应平衡。

## 4. 实验

### 4.1. 数据集

本研究采用 FTransGAN 数据集作为实验数据来源[14]。所有字体图像均统一标准化为  $64 \times 64$  的灰度图像, 以保证不同字体风格之间的分辨率一致性。训练集包含超过 800 种中文字体, 每种字体由 1000 个中文字符及 52 个英文字母构成。为降低字符结构差异对模型训练的干扰, 本文选取 Microsoft YaHei 作为默认内容字体, 并从英文字母中随机抽取 6 个字母作为风格参考样本。

在测试阶段, 本文沿用 FTransGAN 数据集中的两种标准评测设置。未见字体 - 已见字符测试集 (Unseen Font Seen Character, UFSC) 包含 29 种训练阶段未出现的字体样式, 每种字体包含 1000 个中文字符, 这些字符类别在训练集中均已出现, 用于评估模型在未知字体条件下的风格迁移能力。已见字体 - 未见字符测试集 (Seen Font Unseen Character, SFUC) 由训练集中出现的 818 种字体构成, 但在字符层面随机剔除 29 个中文字符, 用以检验模型在已知字体风格条件下生成未见字符的能力。上述两个测试集在字体样式和字符组成上均与训练集严格区分, 从而保证实验评估的客观性与公平性。

### 4.2. 评价指标

生成字体的质量从像素级重建精度、结构相似性以及感知真实度三个方面进行评估。像素级重建误差采用平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE), 其数值越

低表示生成图像与参考图像之间的像素差异越小、重建精度越高[19]。结构相似性采用结构相似性指数 (Structural Similarity Index, SSIM) 及多尺度结构相似性指数 (Multi-Scale SSIM, MS-SSIM), 用于衡量生成图像与参考图像在亮度、对比度及结构纹理等方面的一致性[20]。感知真实度采用学习的感知图像块相似性 (Learned Perceptual Image Patch Similarity, LPIPS) 指标, 在深度特征空间中评估生成字体的视觉自然度, 其数值越低表示生成结果在感知层面上越接近真实字体[21]。

4.3. 对比试验

为进一步验证所提出方法的有效性, 本文将其与 EMD [22]、FTransGAN [14] 和 MF-Net [15] 方法进行了对比实验。其中, EMD 采用独立的内容编码器与风格编码器结构, 分别提取字符的内容特征与风格特征; FTransGAN 则通过多尺度注意力机制实现局部与全局风格特征的融合; MF-Net 结合语言复杂度感知的跳跃连接以适应跨语言结构差异。

实验结果如图 3 所示。在 UFSC 数据集上, EMD 在字体生成过程中容易丢失字符的整体结构信息, 部分样例甚至无法完整保持字符轮廓; FTransGAN 虽然能够较好地维持字符的整体结构, 但在风格细节表达方面仍存在不足, 生成结果与目标风格之间存在明显差异; MF-Net 更适用于纤细笔画字体的生成, 对于笔画较粗的字体, 容易丢失部分结构特征。相比之下, 本文方法在保持字符结构稳定性的同时, 能够更准确地复现目标字体的风格特征, 使生成字体在结构一致性与风格相似性方面均表现出更优的效果。

在 SFUC 数据集中, 纤细笔画及复杂交叉结构更能体现模型对字符结构的建模能力。EMD 与 FTransGAN 在处理细线条时均不同程度地出现笔画断裂、节点缺失等问题, 例如在“丹”字中, 交叉笔画处存在明显的结构丢失现象。MF-Net 虽然能够保持字符的整体结构, 但在风格细节刻画方面仍存在不足, 导致生成结果与目标字体之间差异较大。相比之下, 本文方法能够更有效地捕获细粒度笔画特征, 在保证字符可辨识性的同时, 实现更加清晰且一致的风格表达。

	Unseen Font Seen Character (UFSC)						Seen Font Unseen Character (SFUC)					
Style Image	9	Q	□	□	ρ	A	A	m	N	y	z	∅
Content Image	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井
EMD	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井
FTransGAN	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井
MF-Net	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井
Ours	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井
Target	离	毛	突	鲜	当	到	斑	丹	叉	私	尺	井

Figure 3. Comparison of experimental results display  
图 3. 对比实验结果展示图

如表 1 所示, 在 UFSC 数据集中, 与 EMD 方法相比, 本文方法在 MAE、RMSE、SSIM 及 MS-SSIM 等指标上均具有显著优势, 充分证明其在字体结构保真度和风格一致性方面的有效性。FTransGAN 的整体性能优于 EMD, 但在 SSIM 与 MS-SSIM 等指标上仍不及本文方法, 表明其在字符结构精细保持方面存在一定不足。MF-Net 在部分结构指标上表现相对稳定, 但其在像素误差与感知一致性方面仍存在一定



差距，整体性能略逊于本文方法。在 SFUC 测试集上，本文方法在 MAE、SSIM 与 MS-SSIM 指标上优于对比模型，进一步验证了其在生成复杂笔画未见字符时的泛化能力和鲁棒性。

**Table 1.** Summary table of indicators from comparative experiments

**表 1.** 对比实验各项指标汇总表

	LPIPS↓	SSIM↑	MAE↓	RMSE↓	MS-SSIM↑
Unseen Font Seen Character (UFSC)					
EMD	0.2723	0.2891	0.4857	0.6870	0.2368
FTransGAN	<b>0.1389</b>	0.3171	0.3624	0.5774	0.3361
MF-Net	0.2045	0.3206	0.4302	0.5742	0.3105
Ours	0.1391	<b>0.3216</b>	<b>0.3601</b>	<b>0.5759</b>	<b>0.3383</b>
Seen Font Unseen Character (SFUC)					
EMD	0.2052	0.3945	0.2815	0.4558	0.3383
FTransGAN	<b>0.1131</b>	0.4447	0.2459	<b>0.3829</b>	0.4418
MF-Net	0.2102	0.4172	0.3313	0.4426	0.3567
Ours	0.1149	<b>0.4452</b>	<b>0.2375</b>	0.3831	<b>0.4429</b>

#### 4.4. 消融实验

为评估 CoT-Attn 模块在风格特征提取与迁移过程中的具体作用，本文在 UFSC 数据集上对其进行了消融实验。实验旨在验证不同层数的 CoT-Attn 对生成字体的风格表达、笔画连贯性及整体风格一致性的影响。如图 4 所示，单层 CoT-Attn (w/1) 虽能引入部分风格信息，但整体风格表达仍不稳定，局部笔画存在断裂与噪声。随着 CoT-Attn 层数增加 (w/2、w/3)，生成字体在笔画连贯性、粗细一致性以及整体风格完整性方面均得到明显提升。结果表明，多层 CoT-Attn 有助于更充分地建模和融合风格特征，使生成结果在 UFSC 场景下更接近目标字体风格，从而提升风格迁移的稳定性与一致性。



**Figure 4.** Ablation experiment results

**图 4.** 消融实验结果

## 5. 结论

本文提出了一种融合多层次上下文注意力机制与三元组损失约束的少样本字体生成方法，针对跨语

言字体生成场景中风格特征表达不足、结构迁移不稳定等问题进行了有效改进。通过在风格编码阶段引入多级注意力模块, 模型能够协同建模字体的局部笔画细节与整体结构布局, 从而增强生成结果的风格表达能力; 同时, 三元组损失从特征空间层面对内容与风格进行约束, 有效提升了生成字体在结构完整性与风格一致性方面的表现。

实验结果表明, 所提出方法在字体结构保真度和风格迁移质量等多项评价指标上均优于现有主流方法, 验证了其在跨语言少样本字体生成任务中的有效性与鲁棒性。未来工作将进一步围绕注意力建模效率与跨语言特征泛化能力展开研究, 以支持更复杂字体结构和更丰富风格条件下的字体生成应用。

## 参考文献

- [1] 李少波, 张泠泠. 传统、技术、市场——当代汉字字体艺术的创新路径[J]. 美术观察, 2024(2): 23-25.
- [2] 杨艳芳. 重建数字字体设计的秩序[J]. 南京艺术学院学报(美术与设计版), 2025(1): 171-175.
- [3] 吕文锐, 普园媛, 赵征鹏, 等. 基于字形约束和注意力的艺术字体风格迁移[J]. 计算机工程, 2024, 50(12): 306-317.
- [4] Lian, Z., Zhao, B., Chen, X. and Xiao, J. (2018) EasyFont: A Style Learning-Based System to Easily Build Your Large-Scale Handwriting Fonts. *ACM Transactions on Graphics*, **38**, 1-18. <https://doi.org/10.1145/3213767>
- [5] Li, Y., Lin, G., He, M., Yuan, D. and Liao, K. (2023) Layer Similarity Guiding Few-Shot Chinese Style Transfer. *The Visual Computer*, **40**, 2265-2278. <https://doi.org/10.1007/s00371-023-02915-w>
- [6] 马启星, 李成城, 郭友, 等. 粉笔字字体的多风格生成与迁移方法研究[J/OL]. 计算机工程与应用: 1-11. <https://link.cnki.net/urlid/11.2127.TP.20251023.1840.011>, 2025-10-24.
- [7] Dai, J. (2024) Chinese Calligraphy Style Transfer with Generative Adversarial Network. *Proceedings of the 2024 International Conference on Advanced Robotics, Automation Engineering and Machine Learning*, Hangzhou, 28-30 June 2024, 104-108. <https://doi.org/10.1145/3677454.3677471>
- [8] Zhang, W., Sun, Z. and Wu, X. (2024) An End-to-End Generation Model for Chinese Calligraphy Characters Based on Dense Blocks and Capsule Network. *Electronics*, **13**, Article No. 2983. <https://doi.org/10.3390/electronics13152983>
- [9] Liu, Y., Khalid, F.b., Wang, C., Mustaffa, M.R.b. and Azman, A.b. (2024) An End-to-End Chinese Font Generation Network with Stroke Semantics and Deformable Attention Skip-Connection. *Expert Systems with Applications*, **237**, Article ID: 121407. <https://doi.org/10.1016/j.eswa.2023.121407>
- [10] 刘宇, 丁阳, Fatimah Binti Khalid, 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.
- [11] 赵宏, 黄金海, 常兆斌. 对比学习和多尺度特征提取改进手写字体生成[J/OL]. 华中科技大学学报(自然科学版): 1-10. <https://doi.org/10.13245/j.hust.250582>, 2024-10-29.
- [12] 张嘉, 王志星. 基于风格迁移与骨架化的书法字库自动化生成方法[J]. 电脑知识与技术, 2024, 20(28): 30-33.
- [13] 周雅琴, 车大爽. 基于知识图谱的生成式字体设计研究可视化分析[J]. 包装工程, 2024, 45(S1): 8-21.
- [14] Li, C., Taniguchi, Y., Lu, M. and Konomi, S. (2021) Few-Shot Font Style Transfer between Different Languages. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2021, 433-442. <https://doi.org/10.1109/wacv48630.2021.00048>
- [15] Zhang, Y., Man, J. and Sun, P. (2022) MF-Net: A Novel Few-Shot Stylized Multilingual Font Generation Method. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 2088-2096. <https://doi.org/10.1145/3503161.3548414>
- [16] Qiu, Y., Chu, K., Zhang, J. and Feng, C. (2024) Ma-Font: Few-Shot Font Generation by Multi-Adaptation Method. *IEEE Access*, **12**, 60765-60781. <https://doi.org/10.1109/access.2024.3394441>
- [17] Zhao, H., Ji, T., Rosin, P.L., Lai, Y., Meng, W. and Wang, Y. (2024) Cross-Lingual Font Style Transfer with Full-Domain Convolutional Attention. *Pattern Recognition*, **155**, Article ID: 110709. <https://doi.org/10.1016/j.patcog.2024.110709>
- [18] Pan, W., Zhu, A., Zhou, X., Iwana, B.K. and Li, S. (2023) Few Shot Font Generation via Transferring Similarity Guided Global Style and Quantization Local Style. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 19506-19516. <https://doi.org/10.1109/iccv51070.2023.01787>
- [19] Chen, W., Liu, C. and Ji, Y. (2022) Chinese Character Style Transfer Model Based on Convolutional Neural Network. In: Pimenidis, E., et al., Eds., *Artificial Neural Networks and Machine Learning—ICANN 2022*, Springer, 558-569.

- 
- [https://doi.org/10.1007/978-3-031-15937-4\\_47](https://doi.org/10.1007/978-3-031-15937-4_47)
- [20] Yao, M., Zhang, Y., Lin, X., Li, X. and Zuo, W. (2024) VQ-Font: Few-Shot Font Generation with Structure-Aware Enhancement and Quantization. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 16407-16415. <https://doi.org/10.1609/aaai.v38i15.29577>
  - [21] Wang, Y., Xiong, K., Yuan, Y. and Zeng, J. (2025) EdgeFont: Enhancing Style and Content Representations in Few-Shot Font Generation with Multi-Scale Edge Self-Supervision. *Expert Systems with Applications*, **262**, Article ID: 125547. <https://doi.org/10.1016/j.eswa.2024.125547>
  - [22] Zhang, Y., Zhang, Y. and Cai, W. (2018) Separating Style and Content for Generalized Style Transfer. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8447-8455. <https://doi.org/10.1109/cvpr.2018.00881>