

# 基于字形增强和Transformer的损坏手写东巴字符图像修复

张 扬\*, 刘弈楠, 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20日; 录用日期: 2026年1月17日; 发布日期: 2026年1月26日

## 摘 要

受自然老化、环境侵蚀等因素影响, 古籍普遍存在文字损坏的现象, 现有通用图像修复算法因使用非专用的掩码且难以理解文字结构, 导致出现算法泛化性不足和字形结构修复错误问题。针对上述问题, 构建了一个模拟真实损坏的专用掩码数据集(其中包含九种损坏类型), 同时提出了一种融合字形增强和Transformer的东巴字符图像修复算法。该算法首先引入字形增强模块, 利用小波分解在低频域提取字形结构特征信息; 然后, 利用字形结构特征信息引导Transformer对缺失区域进行结构推断, 重建结构完整、纹理清晰的东巴字符图像。在损坏手写东巴字符图像数据集上的对比实验表明, 所提方法在不同掩码比例上在多项指标上均优于现有算法。实验结果验证了所提算法的有效性, 并为古籍文字及少数民族象形文字的修复提供了可行方案。

## 关键词

东巴文字, 图像修复, 小波变换, Transformer, 深度学习

# Text Image Inpainting for Damaged Handwritten Dongba Characters via Glyph Enhancement and Transformer

Yang Zhang\*, Yinan Liu, Dazhong Mu

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

Received: December 20, 2025; accepted: January 17, 2026; published: January 26, 2026

## Abstract

Influenced by factors such as natural aging and environmental erosion, ancient documents

\*通讯作者。

文章引用: 张扬, 刘弈楠, 牟大中. 基于字形增强和Transformer的损坏手写东巴字符图像修复[J]. 计算机科学与应用, 2026, 16(1): 317-327. DOI: 10.12677/csa.2026.161026

universally suffer from text degradation. Existing general inpainting algorithms, due to the use of non-specialized masks and a failure to capture character structures, result in weak generalization and structural errors in reconstructed glyphs. To address these issues, we construct a specialized mask dataset simulating real-world damage (encompassing nine damage types) and propose a Dongba character inpainting algorithm that integrates a glyph enhancement module with a Transformer. The algorithm first employs the glyph enhancement module, which utilizes wavelet decomposition to extract structural feature information from the low-frequency domain. Subsequently, these structural features guide the Transformer to perform structural inference and content generation for missing regions, thereby reconstructing Dongba character images with complete structures and clear textures. Comparative experiments on a damaged handwritten Dongba character dataset demonstrate that the proposed method significantly outperforms existing algorithms across multiple metrics at various mask ratios. Experimental results validate the effectiveness of the proposed algorithm and offer a viable solution for the inpainting of ancient texts and ethnic minority pictographic scripts.

## Keywords

Dongba Character, Image Inpainting, Wavelet Transform, Transformer, Deep Learning

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

古籍文献作为文化的重要载体，承载着丰富的历史记忆和文化价值。然而，受自然老化、环境侵蚀等因素影响，大量珍贵文献出现了不同程度的损坏，导致信息完整性和传承价值面临威胁。中国纳西族的东巴文字，作为世界上唯一仍在使用的象形文字，被誉为研究人类文字起源的活化石。作为兼具学术价值和文化意义的瑰宝，已被联合国教科文组织列入“世界记忆名录”[1]。在东巴古籍中，手写文字的损坏问题尤为显著(如图1所示)。现存古籍普遍存在笔画断裂、墨迹褪色及局部遮挡等问题，导致关键的字形拓扑结构严重缺失。这一现状不仅增加了古籍数字化处理的难度，也阻碍了东巴语言学的研究与文化遗产的传承。

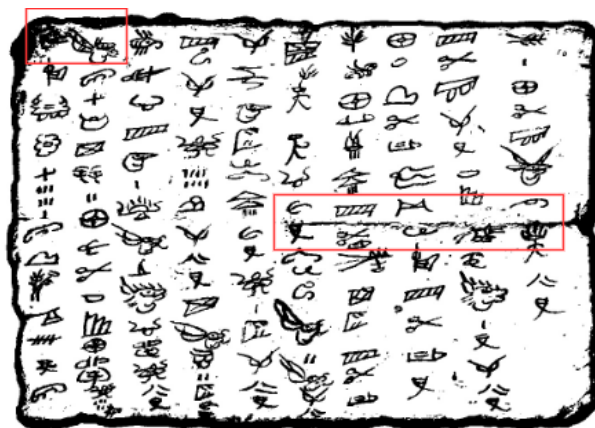


Figure 1. Damaged Dongba ancient books

图1. 损坏的东巴古籍

近年来,针对图像缺损问题,基于深度学习的通用算法取得显著进展,能够在受损区域重构出视觉上连贯且合理的图像内容[2]。然而,通用图像修复算法主要针对自然场景图像或人脸图像,侧重于纹理和色彩的还原,难以捕捉东巴文字的字形结构特征信息。因此,直接将现有图像修复算法应用于损坏的手写东巴字符图像,往往导致修复结果出现字形结构不完整、纹理不清晰和语义错误等问题。

这些问题具体表现为以下几个方面:第一,现有方法易生成伪字符或歧义字符。与自然图像不同,东巴字符图像修复不仅要求视觉连贯,更强调重建字符拓扑结构和语义信息。在全局结构信息缺失的情况下,算法过度依赖局部字形信息,极易生成错误的字符结构,导致语义失真。第二,手写风格的多样性增加了修复难度。手写的东巴文存在显著的类内差异[3]。这种差异不仅体现在不同书写者的工具使用和运笔习惯,还表现在同一字符存在多种结构迥异的书写变体。现有的通用算法难以捕捉字形结构的纹理特征,导致修复结果出现风格不一致与过度平滑的问题。第三,现有的多阶段修复算法存在局限性。为了提升模型的可解释性,研究者[4]-[8]常常采用多阶段修复算法,将修复任务分解为结构预测和纹理细化等子任务,实现渐进式图像修复。然而,多阶段修复算法的性能提升往往伴随着模型参数量和计算开销的显著增加;同时,多阶段修复不仅需要额外监督信息,还可能导致前一阶段的修复误差向后累积并放大,影响最终修复效果。第四,掩码构造方式与真实场景脱节。东巴文字拓扑复杂且像素分布稀疏,而现有算法在训练中常采用简单的矩形掩码来模拟损坏[9][10]。这些非真实掩码导致算法无法学习到足够的结构信息,在修复真实损坏古籍时泛化能力较弱,容易出现过拟合现象。

面对现有方法的局限性,手写东巴字符的损坏图像修复具有重要的研究意义与应用价值。本文首先对东巴数据集 DB1404 [11]进行数据清洗和筛选,在此基础上构建了高质量东巴字符图像数据集;其次,为了模拟古籍中真实的损坏,构建了一个与东巴字符图像相适配的专用掩码数据集,从而提升修复任务在真实场景下的泛化能力。为解决损坏东巴字符图像修复问题,本文提出一种融合字形增强和 Transformer 的字符图像修复算法。首先,引入字形增强模块,利用离散小波变换将图像映射至频域,并在低频子带中捕捉字形拓扑特征;然后,将字形拓扑特征作为结构先验信息,指导 Transformer 提取多尺度特征并建立长程依赖关系,从而在缺失区域生成结构完整、语义正确和纹理清晰的东巴图像。

## 2. 相关工作

随着生成式图像修复算法[4]-[10]不断发展,研究者聚焦于理解图像的高层语义,利用图像上下文信息,在缺损区域生成语义连贯的图像。然而,这些通用算法无法捕捉文字特有的拓扑结构和手写风格特征,导致修复结果出现笔画断裂和结构扭曲等问题,进而削弱下游识别任务的性能。因此,针对古籍文献图像的有效修复,越来越多研究者关注恢复文本图像的纹理和结构[8]。

基于多阶段修复算法的可解释性,研究者将图像修复任务分解为内容重建和纹理恢复。Liu 等[6]针对藏文古籍修复,提出骨架和边缘指导文本图像修复算法。利用 Transformer 重建字符骨架和边缘特征信息,并以此作为先验信息指导图像修复,在字形结构上取得较好的修复效果;但忽略了手写文本的风格多样性。Sun 等[7]和 Zhu 等[8]针对英文文本修复,提出前景-背景两阶段修复算法,通过恢复文本的二值化前景特征作为结构先验,指导模型进行背景补全和纹理细化。然而,训练过程中需要引入额外数据集。

为了增强模型对缺损区域的感知,一些算法提升掩码与图像特征在编码过程中的联系。Li 等[9]针对汉字古籍的虫蛀与墨迹扩散等问题,设计了一种模拟真实侵蚀的掩码,指导扩散模型精准定位修复区域;然而受限于掩码生成算法,无法修复非侵蚀类型损坏文本。Zhang 等[10]提出一种多尺度特征融合模块,通过增强掩码区域的上下文信息交互,提升了东巴文本图像的修复效果。

面对手写东巴文书写风格多变与结构复杂的挑战,直接将现有的算法[12]-[14]应用于手写东巴图像

修复时难以提取足够和有效的结构特征信息，导致修复后的图像存在风格不一致、内容不完整、结构残缺和纹理模糊等问题。因此，本文提出一种融合字形增强和 Transformer 的东巴字符图像修复算法，旨在增强对手写东巴字符图像字形特征信息获取。

### 3. 数据集

#### 3.1. 东巴字符数据集

尽管现存的东巴古籍文献中蕴含着丰富的文字资源，但其普遍遭受了不同程度的损坏。由于历史久远，获取同一东巴文字在破损与未破损两种状态下的成对图像几乎是不可能的，为基于监督学习的修复模型训练带来了严峻挑战。因此，本文基于数据集 DB1404 [11] 构建一个高质量的手写东巴字符数据集。

构建数据集的流程如下：首先，从数据集 DB1404 [11] 中筛选出 11830 张结构复杂的图像。然后，通过超分辨率、数据清洗和裁剪，将图像处理为  $256 \times 256$  像素大小的字符图像。最后，将数据集划分为训练集和测试集，其中训练集包含 10,830 张图像，测试集包含 1000 张图像，用于模型的训练与评估。

#### 3.2. 掩码数据集

为了模拟古籍中的损坏东巴图像，将掩码叠加到完好的东巴文字图像上。然而，现有的通用算法所使用的掩码遮挡率过高和形状单一，难以提取有效的结构特征，导致模型训练时容易产生过拟合，且泛化性能弱。为解决上述问题，本文构建一个与东巴图像相适配的专用掩码数据集。

本文通过物理损坏角度分析大量真实的东巴破损文字图像，将损坏归纳为 9 种类别：酸化、粘连、污渍、霉变、絮凝、鼠啮、虫蛀、侵蚀、焚烧。在此基础上总结出两条东巴字符损坏规律：

- (1) 损坏的蔓延遵循两条路径：a) 从图像边界向中心蔓延；b) 从非边缘区域开始并向周围扩散。
- (2) 尽管各种破损的成因不同，但在视觉上均表现相似。

基于上述损坏规律的分析，本文采用 3 种专用的掩码生成算法(凸包、快速绘画、不规则区域)生成掩码图像(如图 2 所示)。掩码生成的具体流程为：

- (1) 通过 3 种掩码生成方法，随机生成宽在 64 至 256 之间的基础掩码图像。
- (2) 采用随机几何变换对基础掩码扭曲和拉伸，其中几何变换包括仿射拉伸和非线性扭曲。随后，将掩码的尺寸裁剪为  $256 \times 256$  分辨率。
- (3) 将生成的二值掩码进行反色处理，并通过人工筛选，保留在形态上与真实损坏一致的掩码。

通过在训练中引入专用掩码，增强了算法在真实任务中的泛化能力，也显著提升了其面对复杂损坏情况下的鲁棒性。

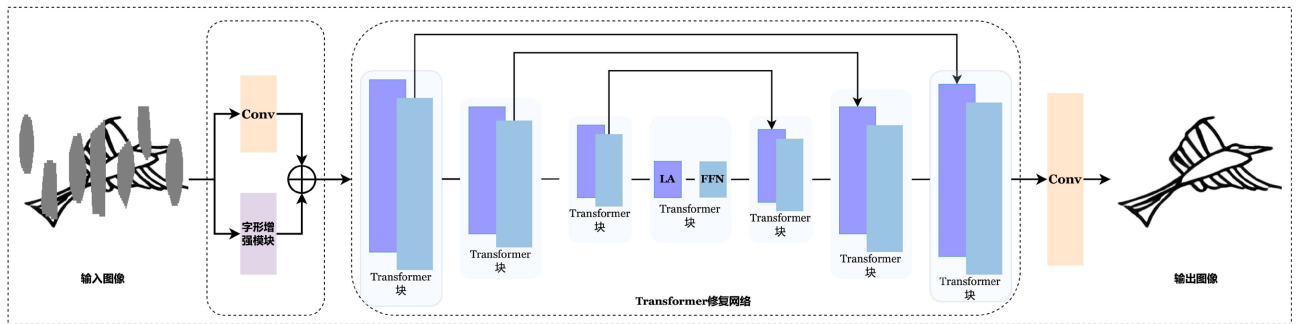


Figure 2. Masked image  
图 2. 掩码图像

## 4. 本文算法

### 4.1. 网络总体结构

针对东巴字符结构复杂且易像素分布稀疏问题,本文提出一种融合字形增强和 Transformer 的字符图像修复算法,整体算法架构如图 3 所示。首先,引入字形增强模块,利用离散小波变换将图像映射至频域,并在低频子带中提取字符整体轮廓的拓扑结构信息,从而增强字形的拓扑结构特征。然后,将字形拓扑特征作为结构先验信息,指导 Transformer 提取多尺度特征并建立长程依赖关系,从而在缺失区域生成结构完整和纹理清晰的东巴图像。



**Figure 3.** Overall architecture text image inpainting for damaged handwritten Dongba characters via glyph enhancement and transformer

**图 3.** 基于字形增强与 Transformer 的东巴字符图像修复网络总体结构图

### 4.2. 字形增强模块

为了实现图像结构信息与细节纹理的有效分解,本文采用离散小波变换将图像映射至频域进行多尺度分解。图像的低频分量主要编码了字符的整体骨架与拓扑形状,而高频分量则对应噪声与纹理。因此,通过增强低频特征子带,能够显著提升模型对字形结构特征提取的鲁棒性[15]。公式(1)表示一组固定卷积核  $W$ :

$$f_{LL} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, f_{LH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, f_{HL} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, f_{HH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (1)$$

小波变换(WT)通过公式(1)中固定卷积核  $W$  实现,如公式(2)所示:

$$[X_{LL}, X_{LH}, X_{HL}, X_{HH}] = \text{Conv}(W, X) \quad (2)$$

其中,  $X_{LL}$  为低频特征,而  $X_{LH}, X_{HL}, X_{HH}$  分别垂直、水平和对角线方向的高频特征,在空间维度上的分辨率都是  $X$  的一半。

由于公式(1)是一组正交基,可以通过转置卷积实现小波逆变换(IWT),公式为:

$$X = \text{Transposed-Conv}(\text{WT}(X), W) \quad (3)$$

其中, Transposed-Conv 为转置卷积,  $W$  表示公式(1)中的一组固定卷积核, WT 表示小波变换。

为了增强字符的形状特征,本文采用一种级联的小波变换的方法(如图 4 所示)。首先,对输入图像执行两级离散小波分解,如公式(4)所示,将特征映射至深层频域空间。然后,在低频子带引入  $1 \times 1$  卷积层,提取并增强字符的骨架结构信息。最后,利用离散小波逆变换将增强后的频域特征重建至空间域,如公式(5)所示。



$$\text{WT}^{(2)}(X) = \text{WT}(\text{WT}(X)) \quad (4)$$

$$X_1 = \text{IWT}\left(\text{Conv}\left(\text{IWT}\left(\text{Conv}\left(\text{WT}^{(2)}(X)\right)\right)\right)\right) \quad (5)$$

其中, *Conv* 表示卷积操作, WT 表示小波变换, IWT 表示小波逆变换。在特征融合阶段, 输入图像通过  $1 \times 1$  卷积, 得到初始特征。将字形结构特征作为指导信息, 并与初始特征融合, 指导 Transformer 模块在损坏区域内容进行修复。

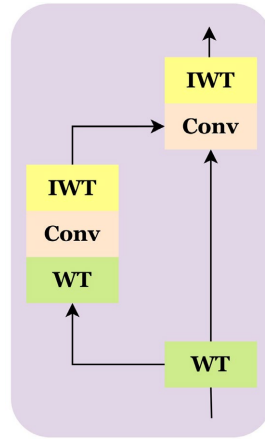


Figure 4. Glyph enhancement module  
图 4. 字形增强模块

### 4.3. U-Net 与 Transformer

为了获得多尺度特征并建立像素间长程依赖关系, 本文引入了 Transformer 的图像修复网络模块。该网络融合 U-Net 架构和 CNN 特征金字塔原理, 采用多尺度编码器-解码器结构。编码器通过多尺度下采样与通道扩展, 提取东巴文字的深层结构特征; 解码器利用上采样与通道压缩, 重构东巴图像的字形结构和纹理细节。其中, Transformer 块由两个级联的子层构成: 线性注意力(linear attention, LA)模块[16]和前馈网络(feed-forward network, FFN)模块[17]。在第  $k$  个 Transformer 块的计算过程如公式(6)所示。

$$\begin{cases} Y^k = \text{LA}(X^{k-1}) + X^{k-1} \\ X^k = \text{FFN}(Y^k) + Y^k \end{cases} \quad (6)$$

其中,  $X^{k-1}$  为第  $k-1$  个 Transformer 块的输出特征,  $Y^k$  为第  $k$  个 LA 模块的输出特征,  $X^k$  为第  $k$  个 Transformer 块的输出特征。

LA 模块如图 5 所示, 为了增强特征的上下文依赖关系, 采用多头线性注意力模块[18]和门控机制[19]级联结构。在第  $k$  个 Transformer 块中, 输入特征信息首先被送入双分支并行结构。主分支利用多头线性注意力[18]捕获全局上下文以生成注意力特征图; 门控分支则通过  $1 \times 1$  卷积与 ReLU 激活函数[20]生成自适应的门控权重。随后, 利用逐元素点乘机制, 将门控权重对注意力特征进行调制, 并通过  $3 \times 3$  卷积处理特征。最后, 引入残差连接[19]将输入特征与增强后的门控注意力特征融合。在第  $k$  个 Transformer 块中, LA 模块的计算过程如公式(7)所示:

$$Y_k = X_{k-1} + \text{Conv}\left(\text{Attention}(X_{k-1}) \odot \text{ReLU}\left(\text{Conv}(X_{k-1})\right)\right) \quad (7)$$

其中,  $X_{k-1}$  是第  $k$  个 Transformer 块的输入特征,  $Y_k$  是第  $k$  个 Transformer 块中 LG 模块输出特征。

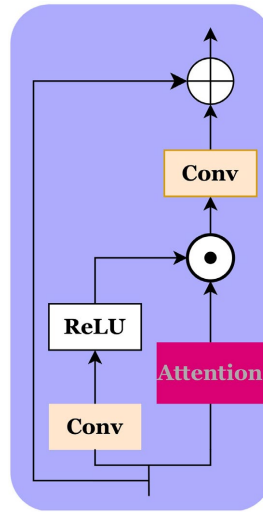


Figure 5. Linear attention structure diagram  
图 5. 线性注意力结构图

FFN 模块如图 6 所示。在第  $k$  个 Transformer 块中，输入特征被分流至两条并行路径。扩展路径利用由 2 个  $1 \times 1$  级联卷积构成的通道扩展模块，将特征通道维度从  $C$  扩展到  $rC$ ，其中， $r$  为扩展比；门控路径则通过两个级联的  $3 \times 3$  深度卷积配合  $ReLU$  激活函数[20]，生成门控权重。随后，利用逐元素点乘机制，将门控权重与高维扩展特征进行融合，实现对特征的空间自适应调制。最后，通过一个  $1 \times 1$  卷积将通道数压缩回  $C$ ，并引入残差连接与输入特征融合，完成特征的重构与更新。在第  $k$  个 Transformer 块中，FFN 的计算过程如公式(8)所示：

$$X_k = Y_k + \text{Conv}(\text{Convs}(Y_k) \odot \text{ReLU}(\text{Convs}(Y_k))) \quad (8)$$

其中， $X_k$  为第  $k$  个 Transformer 块中 FFN 模块的输出特征， $\text{Convs}$  表示由  $1 \times 1$  卷积层和  $3 \times 3$  卷积层串联构成的卷积结构，其中  $1 \times 1$  卷积层用于调整通道维度， $3 \times 3$  卷积层提取空间特征信息。

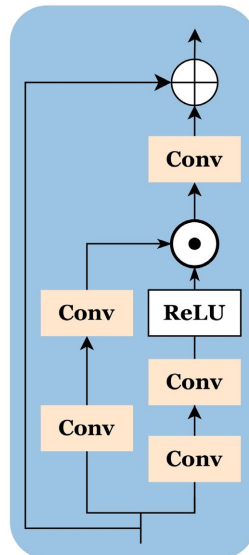


Figure 6. Feedforward network structure diagram  
图 6. 前馈网络结构图

#### 4.4. 损失函数

为了生成内容完整、结构连贯且纹理清晰的东巴图像，本文采用由重建损失、对抗性损失、感知损失、风格损失组成的混合损失函数指导模型训练。总的损失函数由 4 个损失函数加权组成，公式为：

$$L = \lambda_r L_{re} + \lambda_a L_{adv} + \lambda_p L_{perc} + \lambda_s L_{style} \quad (9)$$

其中， $L_{re}$  表示重建损失， $L_{adv}$  表示对抗性损失， $L_{perc}$  表示感知损失，而  $L_{style}$  表示风格损失。 $\lambda_r, \lambda_a, \lambda_p, \lambda_s$  权重参数设置为[10, 0.1, 1, 250]。

重建损失计算为：

$$L_{re} = \|I_{out} - I_{gt}\|_1 \quad (10)$$

其中， $I_{out}$  表示算法修复图像， $I_{gt}$  表示真实图像。

对抗性损失通过生成网络与判别网络之间的博弈进行联合训练，促使生成器输出分布逼近真实数据分布，从而生成视觉上更加真实的图像，计算为：

$$L_{adv} = E_{I_{gt}} [\log D(I_{gt})] + E_{I_{out}} [1 - \log D(I_{out})] \quad (11)$$

其中， $D$  表示谱归一化的 Patch-GAN 判别器。

感知损失利用高层语义特征来衡量生成图像与真实图像之间的差异，计算为：

$$L_{perc} = \sum_i \|\varphi_i(I_{out}) - \varphi_i(I_{gt})\|_1 \quad (12)$$

其中， $\varphi_i$  表示模型 VGG-19 第  $i$  层的特征图。

风格损失计算图像特征之间的特征相关性，可以有效抑制生成图像中的棋盘格伪影[6]，计算为：

$$L_{style} = \sum_i \|\varphi_i(I_{out})\varphi_i^T(I_{out}) - \varphi_i(I_{gt})\varphi_i^T(I_{gt})\|_1 \quad (13)$$

其中， $\varphi_i(I_{out})\varphi_i^T(I_{out})$  表示由预训练 VGG-19 模型第  $i$  层的特征图所构建的格拉姆矩阵。

### 5. 实验

#### 5.1. 实验环境及重要参数设置

本文实验均在 Ubuntu22.04 操作系统上进行，使用 PyTorch 深度学习框架实现，并在单张 NVIDIA RTX 3090 GPU 上完成训练与测试。所提出模型的 U-Net 结构各阶段的 Transformer 块数量设置为[1, 2, 3, 4, 3, 2, 1]。在训练过程中，使用 AdamW 优化器寻求最优解，动态参数设置为  $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ；学习率设置为 0.0001，batchsize 设置 4。模型共训练 40 个 epoch。

#### 5.2. 定量比较

为了验证本文算法的有效性，选取了 ATO-GAN[21]，MADF[22]，SCCGAN[23]，T-FORMER[16] 等代表性修复算法进行对比实验。由于单一指标无法有效评价图像修复质量，本文使用峰值信噪比(peak signal to noise ratio, PSNR)，结构相似性指数(structural similarity index, SSIM)，平均绝对误差(mean absolute error, MAE) 3 种指标评价不同算法模型的性能。不同算法的定量对比结果如表 1 所示，其中黑体为效果最优。数据表明，本文算法在各项指标上均取得了最优结果，证明了其在东巴古籍修复任务中的有效性。

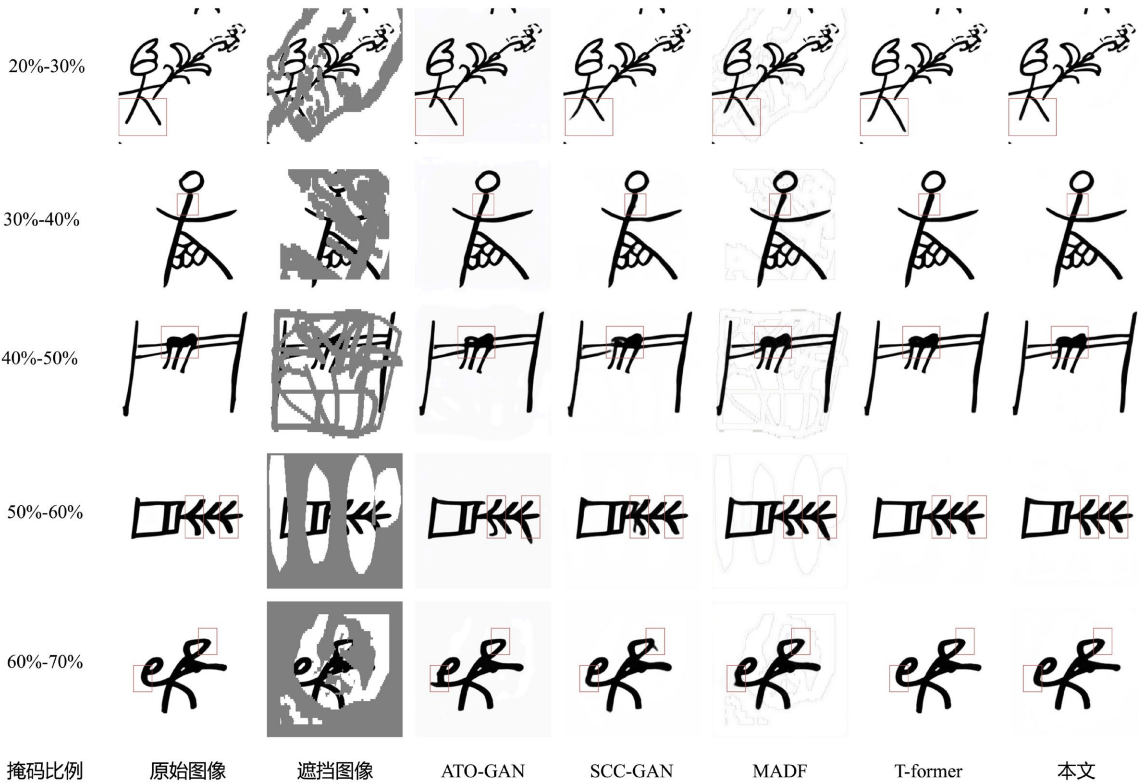
#### 5.3. 定性比较

除了定量指标外，为了更直观比较 5 种算法的修复效果，本文还采用了定性比较方法。图 7 表示在东巴数据集上，针对不同损坏程度手写东巴文字的修复效果。



**Table 1.** Comparison of image inpainting effects of five algorithms  
**表 1.** 对比 5 种算法图像修复效果

评价指标	掩码比例	ATO-GAN	SCCGAN	MADF	T-former	本文
MAE	20%~30%	0.0082	0.0066	0.0074	0.0054	<b>0.0053</b>
	30%~40%	0.0114	0.0119	0.0118	0.0097	<b>0.0095</b>
	40%~50%	0.0184	0.0181	0.0180	0.0159	<b>0.0155</b>
	50%~60%	0.0158	0.0152	0.0156	0.0129	<b>0.0127</b>
	60%~70%	0.0191	0.0186	0.0193	0.0161	<b>0.0159</b>
PSNR	20%~30%	23.95	22.70	23.87	24.04	<b>24.17</b>
	30%~40%	20.99	19.47	<b>21.13</b>	20.70	20.78
	40%~50%	17.80	17.10	18.04	17.92	<b>18.05</b>
	50%~60%	18.53	18.10	18.81	18.87	<b>18.99</b>
	60%~70%	17.22	16.83	17.49	17.52	<b>17.66</b>
SSIM	20%~30%	0.970	0.969	0.934	0.974	<b>0.974</b>
	30%~40%	0.954	0.949	0.906	0.956	<b>0.957</b>
	40%~50%	0.929	0.932	0.901	0.938	<b>0.939</b>
	50%~60%	0.940	0.943	0.902	0.949	<b>0.950</b>
	60%~70%	0.927	<b>0.943</b>	0.883	0.937	0.938



**Figure 7.** Qualitative comparison of character image inpainting results  
**图 7.** 字符图像修复效果的定性对比

实验结果表明, SCCGAN 算法在处理复杂损坏时, 难以准确重建文字结构; ATO-GAN 在小面积遮挡下能有效修复, 但随着掩码区域的扩大, 容易产生错误的纹理填充; MADF 算法在修复区域生成与掩码形状相似的伪影, 影响了整体的视觉自然度; T-FORMER 虽然在内容完整性上有所改善, 但无法有效恢复东巴文字精细的拓扑结构与边缘细节。相比之下, 本文提出的模型能够在全局语义与局部细节中, 生成结构连贯且纹理清晰的高质量修复结果, 与真实图像最为接近。

### 5.4. 消融实验

为验证所提方法中各模块的有效性与性能, 本文在实验数据集上开展消融实验, 相关实验结果如表 2 所示。

**Table 2.** Evaluation results of ablation study  
**表 2.** 消融实验的评估结果

算法	MAE	PSNR	SSIM
变体 1	0.0130	18.70	0.947
变体 2	0.0097	20.70	0.956
变体 3	0.0128	18.74	0.949
本文	<b>0.0095</b>	<b>20.78</b>	<b>0.957</b>

其中, 变体 1 表示基础的 CNN U-Net 模型。变体 2 表示在变体 1 的基础上将编解码器替换为 Transformer 块。变体 3 表示在变体 1 的基础上添加字形增强模块。变体 4 表示完整的本文模型。

实验数据表明, Transformer 模块的缺失会导致整体性能衰退, 验证了其在捕捉全局上下文与长距离依赖方面的关键作用。此外, 移除字形增强模块同样造成了修复质量的下降, 这不仅验证了该模块在精细化局部特征方面的有效性, 也表明了引入针对性的字形指导对于提升修复精度的必要性。

## 6. 结论

针对现实世界手写东巴文字中存在的损坏问题, 本文通过样本筛选、数据清洗和增强处理, 构建了高质量手写东巴字符数据集; 通过分析真实损坏东巴古籍, 采用 3 种掩码生成方法, 构建了模拟真实损坏的掩码数据集。然后, 提出融合字形增强和 Transformer 的东巴字符图像修复算法。利用字形增强模块在低频空间中提取并强化字形结构特征, 进而引导 Transformer 在缺失区域, 重建合理、清晰的东巴结构。实验结果表明, 本文算法在修复东巴字符内容、字形和风格上, 具有较好的性能。

## 致 谢

本研究基于中央民族大学“民族语言智能分析与安全治理实验室”提供的手写东巴文数据集 DB1404, 在此对该实验室表示衷心感谢。

## 基金项目

由国家级大学生创新训练计划项目(NO. 202510015013)资助。

## 参考文献

- [1] Wu, G., Ding, C., Xu, X., *et al.* (2016) Intelligent Recognition on Dongba Manuscripts Hieroglyphs. *Journal of Electronic Measurement and Instrumentation*, **30**, 1774-1779.
- [2] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R. and Van Gool, L. (2022) RePaint: Inpainting Using

- Denoising Diffusion Probabilistic Models. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11451-11461. <https://doi.org/10.1109/cvpr52688.2022.01117>
- [3] Krishnan, P., Kovvuri, R., Pang, G., Vassilev, B. and Hassner, T. (2023) TextStyleBrush: Transfer of Text Aesthetics from a Single Example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 9122-9134. <https://doi.org/10.1109/tpami.2023.3239736>
  - [4] Dong, Q., Cao, C. and Fu, Y. (2022) Incremental Transformer Structure Enhanced Image Inpainting with Masking Positional Encoding. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11348-11358. <https://doi.org/10.1109/cvpr52688.2022.01107>
  - [5] Cao, C., Dong, Q. and Fu, Y. (2023) ZITS++: Image Inpainting by Improving the Incremental Transformer on Structural Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 12667-12684. <https://doi.org/10.1109/tpami.2023.3280222>
  - [6] Liu, Y., Zhao, Q., Pan, F., Gao, D. and Danzeng, P. (2023) Structure Prior Guided Text Image Inpainting Model. *Journal of Image and Graphics*, **28**, 3699-3712. <https://doi.org/10.11834/jig.220960>
  - [7] Sun, J., Xue, F., Li, J., Zhu, L., Zhang, H. and Zhang, J. (2023) TSINIT: A Two-Stage Inpainting Network for Incomplete Text. *IEEE Transactions on Multimedia*, **25**, 5166-5177. <https://doi.org/10.1109/tmm.2022.3189245>
  - [8] Zhu, S., Fang, P., Zhu, C., Zhao, Z., Xu, Q. and Xue, H. (2024) Text Image Inpainting via Global Structure-Guided Diffusion Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 7775-7783. <https://doi.org/10.1609/aaai.v38i7.28612>
  - [9] Li, H., Du, C., Jiang, Z., Zhang, Y., Ma, J. and Ye, C. (2024) Towards Automated Chinese Ancient Character Restoration: A Diffusion-Based Method with a New Dataset. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 3073-3081. <https://doi.org/10.1609/aaai.v38i4.28090>
  - [10] Zhang, Y., Zhao, L., Wang, Z., *et al.* (2024) Incomplete Handwritten Dongba Character Image Recognition by Multiscale Feature Restoration. *Pattern Recognition*, **148**, Article ID: 110309.
  - [11] Luo, Y., Bi, X., Wu, L., *et al.* (2022) Dongba Pictographs Recognition Based on Improved Residual Learning. *CAAI Transactions on Intelligent Systems*, **17**, 79-87.
  - [12] Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., *et al.* (2023) TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 13094-13102. <https://doi.org/10.1609/aaai.v37i11.26538>
  - [13] Wang, T., Zhu, Y., Jin, L., Luo, C., Chen, X., Wu, Y., *et al.* (2020) Decoupled Attention Network for Text Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12216-12224. <https://doi.org/10.1609/aaai.v34i07.6903>
  - [14] Singh, S.S. and Karayev, S. (2021) Full Page Handwriting Recognition via Image to Sequence Extraction. In: Lladós, J., Lopresti, D. and Uchida, S., Eds., *Document Analysis and Recognition—ICDAR 2021*, Springer, 55-69. [https://doi.org/10.1007/978-3-030-86334-0\\_4](https://doi.org/10.1007/978-3-030-86334-0_4)
  - [15] Finder, S.E., Amoyal, R., Treister, E. and Freifeld, O. (2024) Wavelet Convolutions for Large Receptive Fields. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Computer Vision—ECCV 2024*, Springer, 363-380. [https://doi.org/10.1007/978-3-031-72949-2\\_21](https://doi.org/10.1007/978-3-031-72949-2_21)
  - [16] Deng, Y., Hui, S., Zhou, S., Meng, D. and Wang, J. (2022) T-Former: An Efficient Transformer for Image Inpainting. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 6559-6568. <https://doi.org/10.1145/3503161.3548446>
  - [17] Huang, W., Deng, Y., Hui, S., Wu, Y., Zhou, S. and Wang, J. (2024) Sparse Self-Attention Transformer for Image Inpainting. *Pattern Recognition*, **145**, Article ID: 109897. <https://doi.org/10.1016/j.patcog.2023.109897>
  - [18] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. arXiv: 1706.03762.
  - [19] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T. (2019) Free-Form Image Inpainting with Gated Convolution. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 4470-4479. <https://doi.org/10.1109/iccv.2019.00457>
  - [20] Hendrycks, D. and Gimpel, K. (2016) Gaussian Error Linear Units (Gelus). arXiv: 1606.08415.
  - [21] Zeng, Y., Fu, J., Chao, H. and Guo, B. (2023) Aggregated Contextual Transformations for High-Resolution Image Inpainting. *IEEE Transactions on Visualization and Computer Graphics*, **29**, 3266-3280. <https://doi.org/10.1109/tvcg.2022.3156949>
  - [22] Zhu, M., He, D., Li, X., Li, C., Li, F., Liu, X., *et al.* (2021) Image Inpainting by End-To-End Cascaded Refinement with Mask Awareness. *IEEE Transactions on Image Processing*, **30**, 4855-4866. <https://doi.org/10.1109/tip.2021.3076310>
  - [23] Liu, R., Wang, X., Lu, H., Wu, Z., Fan, Q., Li, S., *et al.* (2021) SCCGAN: Style and Characters Inpainting Based on CGAN. *Mobile Networks and Applications*, **26**, 3-12. <https://doi.org/10.1007/s11036-020-01717-x>