

基于扩散模型的甲骨摹本图像生成技术研究

李邦^{a,b,c}, 杨婧^a, 丁增茂^a, 方嘉琪^a, 韩胜伟^a, 蔡佳泽^a,
张展^{a,b,c}, 郭安^{a,b,c}, 高峰^{a,b,c}, 刘永革^{a,b,c}

(安阳师范学院 a.甲骨文信息处理教育部重点实验室; b.计算机与信息工程学院;
c.河南省感知数据智能处理国际联合实验室, 河南 安阳 455000)

摘要: 甲骨摹本是甲骨文研究中的重要资料, 主要用于摹写甲骨表面的轮廓、字符及裂痕等关键信息. 然而, 传统摹本制作过程耗时长、劳动强度大, 且高度依赖专家经验. 针对上述问题, 本文提出了一种基于字符信息密度归一化的扩散模型训练策略, 在有效提取字符纹理特征的同时, 兼顾甲骨轮廓等整体图像纹理信息. 此外, 为满足摹本生成中对甲骨字符突出表达的需求, 本文引入零终端信噪比机制以控制模型的分布差异性. 实验结果表明, 该方法在甲骨摹本生成任务中显著提升了图像生成质量与空间表达效率, 为甲骨文字形研究提供了一种可靠的技术手段.

关键词: 甲骨文; 图像生成; 扩散模型; 摹本

中图分类号: TP391

文献标志码: A

文章编号: 1000-2367(2026)03-0052-08

甲骨文是迄今发现的中国最早的成熟文字系统, 是汉字的源头和中华优秀传统文化的重要根脉. 它不仅承载着丰富的历史信息, 也是汉字演变过程中独一无二的“活化石”, 在研究中国上古社会结构、文化思想及汉字起源等方面具有不可替代的学术价值. 鉴于甲骨本身的文物属性, 甲骨文研究通常不直接使用原始甲骨片, 而是以其图像记录材料为研究对象, 并收录于各类甲骨著录中. 目前最常见的图像记录形式主要包括拓片和摹本. 拓片通过“墨拓”方式记录甲骨表面信息, 是现阶段最主要的甲骨文载体; 而摹本则是在拓片基础上对甲骨信息进行二次提取, 通过人工摹写进一步突出字符、裂痕及纹理等关键信息. 因此, 相较于拓片, 摹本在信息呈现上更为直观, 是更理想的研究资料. 然而, 摹本制作高度依赖人工经验, 对于拓片中模糊不清的内容, 往往需要深厚的甲骨文功底才能准确摹写, 同时人工绘制不可避免地会在笔画细节上与原始字形产生偏差. 这使得甲骨摹本的绘制成为一项兼具高知识门槛和高劳动成本的工作, 摹本数量不足也逐渐成为制约甲骨研究的重要瓶颈. 目前已知的有字甲骨约 15 万片, 而收录摹本数量最多的著录为黄天树教授团队出版的《甲骨文摹本大系》, 其仅涵盖约 7 万片甲骨, 且资料整理与出版历时近十年, 充分反映了摹本绘制工作的艰巨性.

近年来, 随着人工智能技术在图像生成领域的快速发展, 相关方法在考古材料绘图任务中取得了显著进展. ZENG 等^[1]针对文物图像语义分割问题, 提出了一种基于多分支特征融合的生成对抗网络方法, 实现了对敦煌壁画线条和边缘信息的有效提取. 该方法通过多尺度特征融合与边缘检测模块, 实现了对文物轮廓的

收稿日期: 2024-12-10; **修回日期:** 2025-01-16.

基金项目: 国家自然科学基金(62506007); 古文字与中华文明传承发展工程规划(G1806); 河南省自然科学基金(242300420680); 河南省高层次人才国际化培养项目(GCC2025028).

作者简介: 李邦(1991—), 男, 河南濮阳人, 安阳师范学院副教授, 博士, 研究方向为甲骨文信息处理, E-mail: libang@aynu.edu.cn.

通信作者: 刘永革, E-mail: liuyongge@aynu.edu.cn.

引用本文: 李邦, 杨婧, 丁增茂, 等. 基于扩散模型的甲骨摹本图像生成技术研究[J]. 河南师范大学学报(自然科学版), 2026, 54(3): 52-59. (Li Bang, Yang Jing, Ding Zengmao, et al. Research on generation technology of oracle bone facsimile images based on diffusion model[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(3): 52-59. DOI: 10.16366/j.cnki.1000-2367.2024.12.10.0004.)

精细分割与细节描绘.然而,与一般文物线图绘制相比,甲骨摹本的生成需要对具有丰富语义信息的甲骨字这一抽象符号进行准确分割,对图像生成任务提出了更高要求.一方面,甲骨风化与碎裂严重,拓片图像在尺寸和内容规模上差异极大,既包含仅有一两个字符的小型残片,也存在包含数百字符的大型甲骨.这种极端的尺度差异使传统图像分割方法在字符提取上面临挑战:全局归一化策略易导致字符尺度失真,而基于滑窗的多尺度方法虽能关注局部细节,却往往破坏拓片的整体结构,造成拼接轮廓不一致.另一方面,甲骨摹本要求对每个字符的形态与笔画细节进行精细刻画,对模型的分割精度和细节还原能力提出了更高要求.与以轮廓或整体结构为主要目标的传统文物语义分割不同,本研究更强调字符级精确分割与清晰化,对模型的字符语义建模能力提出新的挑战.

针对上述问题,本文选用具备更强高频信息重建能力的扩散模型作为基准框架,以提升细粒度生成效果.针对甲骨图像中内容信息密度不均衡,以及扩散模型固有分布差异性难以控制、训练稳定性不足等问题,提出了一种多尺度交替训练策略,以实现字符信息与图像纹理之间的平衡.同时,为进一步强化甲骨字形分布特征,本文引入零终端信噪比设置,以减小模型生成样本与真实摹本在风格上的差异.综合上述方法,实现了从甲骨拓片到高质量甲骨摹本的自动生成,为传统方法难以应对的复杂问题提供了一种新的解决思路.实验结果表明,与传统图像生成方法相比,该方法在甲骨字形等具有明确语义的信息提取任务中表现出更强的鲁棒性,其生成结果在一定程度上也具备优于人工摹写的客观一致性.

1 相关工作

1.1 图像语义分割模型

图像语义分割模型作为计算机视觉的核心技术之一,旨在为图像中的每个像素分配语义标签,是场景理解 and 目标检测的重要基础.近年来,深度学习的迅猛发展极大推动了语义分割领域的技术进步.从传统的基于手工特征的方法到基于卷积神经网络(CNN)的端到端模型,语义分割技术的精度和鲁棒性不断提升.近年来,编码器-解码器架构在语义分割任务中取得了较好的表现,ZHAO等^[2]提出了一种改进的生成对抗网络(Seg-GAN)用于图像语义分割任务.该方法将卷积条件随机场(ConvCRFs)引入到判别器网络中,以提高分割结果与真实分布之间的差异,并改善输出图像的细节.刘国奇等^[3]提出一种以Transformer为编码器的甲骨钻凿分割网络.尽管这些模型在开放场景的语义分割任务中表现出色,但面对高噪声、纹理复杂以及跨域应用的任务时,其泛化能力和鲁棒性仍存在不足.因此,针对不同应用场景设计任务特定的模型或结合生成式方法来增强分割性能,已成为当前的研究重点.

1.2 历史文物图像的语义分割

利用语义分割技术对历史文物图像进行自动特征提取长期备受考古研究的关注.近年来,研究者开始引入深度学习技术,用于提升文物图像分割的精度和鲁棒性.例如,ENAYATI等^[4]针对庞贝考古遗址的壁画碎片,提出了将图像分割为背景、碎片背景和母题区域的3类分割任务,以及将母题进一步细分为12类的语义分割任务,并结合优化的U-Net和YOLOv8模型,利用色彩空间和图像增强技术应对类别不平衡及碎片颜色、纹理变化等挑战.然而,现有的分割方法仍然面临多样化文物任务的挑战.历史文物图像往往因长期的自然老化和人类活动而呈现高噪声、不完整边界及严重的细节缺失,在壁画等特定文物上表现优异的分割模型未必适用于其他文物类型.

1.3 扩散模型

扩散模型(diffusion models)作为生成模型的最新发展方向,在图像生成领域展现了显著的性能优势.其主要机制是通过逐步在数据中加入噪声,然后在生成过程中反向去噪,最终生成高质量的图像.HO等^[5]首次提出了基于扩散过程的生成框架,显著提高了生成任务的稳定性与多样性.Stable Diffusion模型^[6]优化了这一框架,通过引入在潜在空间中运行的扩散模型(latent diffusion models, LDMs),大幅降低了计算复杂度,并保留了生成高分辨率图像的能力.

在语义分割任务中,扩散模型因其生成与判别能力的结合展现出广阔的应用前景.例如,WU等^[7]提出的DiffuMask利用扩散模型生成具有像素级标注的合成图像,显著缓解了高质量标注数据不足的问题.同

时,该方法在生成数据的多样性和质量上均表现出色,展示了扩散模型在辅助语义分割任务中的潜力.TIAN等^[8]基于扩散模型(stable diffusion)构建了一种无监督零样本语义分割框架,通过引入注意力模块对输入图像进行精确区域分割,而无需依赖人工标注信息.

随着扩散模型条件控制能力的提升,其在语义分割中的应用范围也得到扩展.例如,ControlNet^[9]通过结合输入条件(如语义分割图或边缘图),显式引导扩散过程,极大增强了模型的灵活性和可控性.这一能力为复杂场景下的分割任务(如高噪声文物图像或异质性强的分割目标)提供了新的解决方案.

综上所述,扩散模型作为生成式模型的一大突破,正日益成为语义分割任务中的重要工具.通过生成高质量的合成数据、支持无监督学习以及引入条件控制机制,扩散模型在复杂场景和高噪声条件下的分割任务中展现出独特优势,推动了该领域的发展.

2 本文方法

本研究的主要目标为提取甲骨拓片中的字符与轮廓并生成甲骨摹本,该任务要求生成摹本在符合拓片纹理的基础上,尽可能准确地生成甲骨字形.这就要求模型具有很强的低频信息捕获能力,以实现对图像细节的保真.考虑到生成对抗网络主要通过判别器对生成结果进行判断,对图像的细粒度特征的判别能力有限,本文使用扩散模型作为图像生成模型的基座,并借鉴 ControlNet 的方法,将拓片作为控制图像,提供甲骨内容信息引导模型扩散以生成相应摹本图像.同时,本研究还针对甲骨拓片固有的字符密度不一致和扩散模型的分布差异性控制等问题设计了训练策略,使模型可以学习不同尺度的拓片特征以及调整扩散链,从而保证模型输出结果的稳定性和高质量.

2.1 扩散模型

扩散模型包括一个正向过程,该过程通过添加高斯噪声 ϵ 将样本 z_0 转换为噪声样本 z_T ;还包括一个反向过程,该过程将噪声 z_T 通过迭代去噪得到 z_0 ,其中 $\epsilon \sim \mathcal{N}(0, 1)$, T 表示总的时间步数,前向过程可以表述为:

$$z_t = \sqrt{\alpha_t} z_0 + \sqrt{1 - \alpha_t} \epsilon, \quad (1)$$

z_t 是第 t 步的噪声特征,其中 $t \in [1, T]$, α 是一个超参数.

在反向过程中,给定从标准高斯随机采样得到的输入噪声 z_T ,模型 ϵ_θ 在条件 C 下估计每一步 t 的噪声.经过 T 次逐步细化迭代后, z_0 被导出为输出样本:

$$z_{t-1} = \frac{\sqrt{\alpha_{t-1}}}{\sqrt{\alpha_t}} z_t + \sqrt{\alpha_{t-1}} \left(\sqrt{\frac{1}{\alpha_{t-1}} - 1} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \epsilon_\theta(z_t, t, C). \quad (2)$$

扩散模型的训练通过优化去噪模型 ϵ_θ 在条件 C 估计对应的噪声 ϵ ,其目标是:

$$\min_{\theta} E_{z_0, \epsilon \sim \mathcal{N}(0, D), t \sim U(1, T)} | \epsilon - \epsilon_\theta(z_t, t, C) |. \quad (3)$$

2.2 拓片图像引导

为了将拓片图像特征有效地融入到预训练的扩散模型中,模型通过引入一个额外的分支,将拓片图像特征融合到隐藏层来控制扩散模型的去噪过程.具体而言,在生成过程中,噪声潜在变量的更新依赖于拓片图像的特征,从而为最终生成结果提供引导.

为实现这一目标,通过条件控制注入的方法^[9],将拓片图像的特征通过控制网络注入到扩散模型的去噪过程.具体而言,本方法通过引导模型在去噪过程中的状态变化,使得生成的摹本图像能够更准确地反映拓片图像中的结构信息.具体来说,在每一步的去噪过程中,控制网络的输出作为条件信息与扩散模型的内部状态共同作用,指导噪声的消除.公式表示为:

$$p_\theta(z_{t-1} | z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t), \Sigma_\theta(z_t, t, \mathcal{F}(c))), \quad (4)$$

其中, $\mathcal{F}(c)$ 表示条件提取模块,以实现将提取到的特征融合到生成模型以控制去噪的过程.

2.3 面向字符提取的扩散策略

2.3.1 字符信息密度归一化

由于甲骨片中文本的信息密度存在不一致性,并且甲骨片的尺寸与文本数量呈正相关,直接将所有甲骨

片统一放缩至相同尺寸进行生成并不合适.为应对这一挑战,本研究提出了一种基于切片策略的生成与推理方法.在生成过程中,输入图像通过切片操作进行分块,以适应不同尺寸的输入.设输入图像为 $I \in R^{H \times W \times C}$,其中 H 和 W 分别为图像的高度和宽度, C 为通道数.切片操作将图像分割为若干个子图像 $I_s \in R^{h \times w \times c}$,其中 h 和 w 分别是切片后的子图像的高度和宽度,满足 $h \leq H$ 且 $w \leq W$.

然而,切片操作可能会导致模型丧失拓片中的全局信息.因此,在去噪阶段,没有对输入图像进行切片处理,仅在变分自编码器(VAE)过程中进行切片.设 VAE 的输入为 I_s ,其生成过程可表示为: $I_s = \epsilon(I_s)$ 和 $I'_s = \mathcal{D}(z_s)$,其中, ϵ 和 \mathcal{D} 分别表示编码器和解码器, z_s 为编码后的潜在变量.

为有效提升训练效率并支持并行化训练,本方法采用了交替训练策略.具体而言,在不同的训练步长 t 中,图像的放缩与切片操作交替进行.设训练步长 t 分为两类,分别为放缩阶段 t_1 和切片阶段 t_2 ,在这些阶段中分别执行如下操作.

放缩操作:将输入图像统一缩放至固定尺寸 I_{resize} (见式(5));切片操作:将图像分割为多个子图像 I_{slice} (见式(6)):

$$I_{\text{resize}} = R(I), \quad (5)$$

$$I_{\text{slice}} = S(I). \quad (6)$$

交替训练的策略设计使得模型在训练的过程中,通过不同的数据形式关注到不同层面的图像细节.在放缩阶段,由于甲骨字形产生了不同程度的放缩失真,模型难以很好地学习字符细节,但由于图像整体信息得到了保全,因此可以较好地学习全局信息;相反,在切片阶段,模型并不能获取并学习全局信息,因此字符的局部细节得到了更多的关注.

2.3.2 分布差异性控制

甲骨摹本可以被视为对原始拓片信息的高度提取.然而,尽管扩散模型在大规模图像数据集上经过预训练,具备强大的生成能力,其在提取甲骨字形的过程中仍存在一定的不稳定性.因此,直接应用 ControlNet 时,生成的图像往往与目标分布之间存在较大差异.具体而言,在时刻 $t = T$ 时,输入模型的信号并非纯粹的噪声,而是仍然包含了一定量的残余信号.这些泄露信号包含了图像的低频信息,例如每个通道的均值.模型在推理过程中会从这些泄露的信息中执行去噪声过程.为了提高生成的稳定性,引入了零终端信噪比策略.具体而言,通过在公式(1)中控制 $\sqrt{a_t}$ 的混合信号量,并保持 $\sqrt{a_1}$ 不变,将 $\sqrt{a_T}$ 设置为零,并对中间阶段的 $t \in [2, \dots, T-1]$ 的 $\sqrt{a_t}$ 进行线性调整.该策略有效地减少了额外信号的干扰,促使模型在生成过程中更加专注于目标图像的细节,从而实现更加稳定且精确的字符图像生成.

3 实验分析

3.1 数据集

为了满足训练需求,需要使用匹配的甲骨拓片图像与摹本图像对.本实验采用了 2024 年发布的甲骨文多模态数据集^[10](oracle bone inscriptions multi-modal dataset,简称 OBIMD).OBIMD 具有成对的拓片、摹本图像,并对其内容进行了详细标注.这些摹本图像为人工绘制,摹本中的字形虽然确实为正确的甲骨文字形,但在笔画细节上与拓片上的原始字形依然存在一定的误差,字形的宽窄、高低等差距也导致专家摹本无法与拓片图像达到逐像素对齐,甚至存在部分摹本字形几乎不能与拓片字形在像素上重叠.因此,虽然在生成任务中将数据集中的专家摹本图作为真实数据(ground truth)进行训练,但本质上,该实验属于一个弱监督任务语义分割任务,真实数据中的摹本仅有部分像素实现了正确拓片内容的语义分割.

3.2 实验设置

所有实验均在 Pytorch v2.4 框架上进行,并使用官方的扩散模型框架 Diffusers v0.33.使用 4 张 NVIDIA 4090 显卡训练所有模型.底座模型使用 Stable Diffusion v1.5,在训练时冻结底座模型不用于反向传播,仅对 ControlNet 部分的参数进行反向传播以及参数更新,其中 ControlNet 使用的激活函数为 SiLU,并且将 ControlNet 的特征图作为残差注入到 Stable Diffusion v1.5 的下采样过程中.实验使用 OBIMD 数据集中的 10 077 组拓片摹本对.在训练时使用了 9 124 张拓片作为扩散模型的引导条件,并使用其对应的摹本

图片用于训练.测试阶段中使用 953 对拓片摹本进行测试.学习率使用 0.000 01 以及权重衰减设置为 0.000 1 的 Adam 优化器,精度设置为 fp16, BatchSize 设置为 8, 梯度累计设置为 4, 并且所有输入的图像的 size 为 512×512 . 所有模型都经过了 2 万次迭代训练并已经完全收敛.

3.3 实验结果对比分析

相较于《甲骨文摹本大系》中的摹本,生成摹本展现出更忠实于原始拓片细节的生成结果.比较结果如图 1 所示,生成摹本在甲骨字形(红色虚线框)、甲骨轮廓与纹理特征(蓝色虚线框)方面均能够比人工更好地凸显甲骨片的原有特征.这种高保真度的还原不仅体现在视觉上的相似性,更在于其对原始信息完整性和精确度的极致追求,为后续的分析与研究奠定了坚实的基础.

首先,在字形风格与笔画细节的还原方面,模型生成的摹本更加贴近原始甲骨拓片

图像,笔画细节丰富且变形较少,尤其在字形拐点的锐利与圆润处理方面表现出色,能够在书写风格方面辅助专家进行断代分组研究.其次,对于存在残缺或模糊的字形区域,模型生成的摹本能够对字形进行合理的补全和保留,避免了人工摹写过程中因部分字形笔画不完整或模糊所导致的省略现象,以及主观判断下争议较大的添补情况,为字形结构乃至辞例的完整性和准确性研究提供了重要参考.

最后,模型在甲骨片边缘轮廓与显著裂痕的处理上更加精准流畅,不仅高度还原了原始纹理特征,还能够有效区分噪声与关键纹理信息.例如,模型避免将边缘轮廓附近的噪声误判为甲骨片的残缺,从而在细节表现上超越了专家摹本绘制方式,增强了摹本图像的真实性和还原性,从而提高了甲骨摹本的学术和研究价值.

3.3.1 定性分析

实验采用定性分析对比了本文提出的模型与图像生成模型 CycleGAN^[11] 和 DiscoGAN^[12], 并进行了可视化分析,如表 1 所示.实验结果表明,无论是字形和边缘轮廓模糊的低质量甲骨拓片图像,还是字形和纹理清晰的高质量甲骨拓片图像,本文模型的生成效果均明显优于其他模型. CycleGAN^[11] 虽然在一定程度上能够生成较为完整的甲骨片边缘轮廓,但在字形语义方面存在缺陷.它生成的甲骨摹本图像,文字大多缺失,笔画表现杂乱且与非字形纹理相互融合,字形扭曲且边缘不平滑,难以清晰区分字形的语义信息.此外,其生成图像的背景画面充满噪点,进一步影响了整体效果.相比之下, DiscoGAN^[12] 的表现略优,生成图像的背景较为干净,没有明显噪点,且字形笔画较为流畅.然而,即便是在高质量的拓片图像输入情况下, DiscoGAN^[12] 依旧会丢失甲骨的边缘轮廓和部分字形细节.在低质量的拓片图像输入情况下,其性能则明显下降,生成结果画面杂乱,字形与非字形纹理交叠,稳定性较差.相较之下,本文模型克服了非字形纹理的干扰,表现出显著的优越性.无论是高质量还是低质量拓片图像输入,输出结果均具有良好的稳定性和一致性,能够完整、准确地生成清晰的甲骨摹本图像.在甲骨文的文字数量、位置、结构、笔画细节和甲骨片边缘轮廓等方面,本文模型均具有出色表现,为甲骨文的研究提供了有力支持.

3.3.2 定量分析

如上文所述,由于摹本生成任务属于弱监督语义分割任务.而结构相似性(SSIM)、峰值信噪比(PSNR)和局部熵等指标,作为传统的图像生成任务中逐像素对齐的评价标准,通常用于强监督学习任务,例如图像翻译、图像复原、图像去噪以及语义分割等.这些任务的共同特点是生成的图像需要在像素层面与真实样本(Ground Truth)高度一致,否则评分会因局部偏差显著下降.然而,专家摹本中的甲骨字形并非与拓片逐像素对应,局部熵、SSIM 和 PSNR 等传统指标在此任务中难以准确评价生成模型对甲骨字形特征的提取能力,无法有效反映其实际性能,并不适用于本任务.

本质上,摹本生成任务要求模型具有从甲骨拓片上正确提取字形特征的能力.因此,在定量分析中,本文通过目标检测模型对生成摹本中字形的检测与识别的正确率作为评价生成图像质量的主要指标.具体而言,

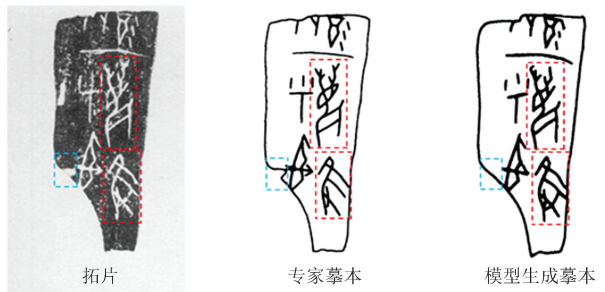


图1 专家摹本与本文模型生成摹本对比图

Fig.1 Comparison diagram of expert-created facsimiles and facsimiles generated by the model proposed in this paper

本文使用 OBIMD 数据集中的甲骨字位置和类别标注信息,训练了 YOLOv11^[13] 目标检测模型,用以检测与识别摹本图像中的甲骨字形.在模型训练中切分训练集与测试集的比例同摹本生成任务相同,为了保证性能的稳定,这里使用 YOLOv11 官方标准模型和优化器,并按照官方的实验对超参数进行设置,且未对模型结构做额外调整,最终该模型经过了约 30 万次迭代训练以保证评估结果.

表 1 不同模型生成的甲骨摹本图像比较

Tab. 1 Comparison of oracle bone facsimile images generated by different models

片号	输入图片	CycleGAN ^[11]	DiscoGAN ^[12]	Ours	专家摹本
合集 33					
合集 51					
合集 317					

为了全面分析目标检测模型在生成图像上的性能,从而衡量生成模型的字形提取效果,本文将测试集中的 953 张摹本用于对比算法输入,同时使用其对应的拓片图像用于摹本生成.本文选取了以下 4 个评价指标:mAP50、mAP50-95、Precision 和 Recall.这些指标分别从检测模型的整体精度、精确性和召回率等维度全面且间接反映生成图像中的字形信息质量.评价结果如表 2 所示.

表 2 目标检测模型在不同方法生成的摹本图像上的性能评估

Tab. 2 Evaluation of the performance of object detection models on facsimile images generated by different methods

测试样本	评价指标				测试样本	评价指标			
	mAP50	mAP50-95	Precision/%	Recall/%		mAP50	mAP50-95	Precision/%	Recall/%
专家摹写	0.557 6	0.372 5	69.93	47.70	CycleGAN ^[11]	0.041 9	0.018 7	45.56	3.89
Ours	0.350 8	0.196 3	58.61	30.35	DiscoGAN ^[12]	0.263 5	0.152 1	52.52	23.32

从实验结果可以看出,甲骨摹本的检测与识别任务对于 YOLOv11 依然存在很大的挑战性,来自 OBIMD 的专家摹本构成的测试集也仅能达到 69.93% 的准确率.这表明经过训练的 YOLOv11 已经具备了识别检测摹本中甲骨字的能力,生成结果中的甲骨文越接近甲骨文专家对字形的正确摹写,检测结果就越能展现出更好的性能.CycleGAN 的生成摹本仅得到了 4.19% 的 mAP50,原因在于该方法仅能较好地提取图像的全局特征,很难准确地生成字形,该现象在定性分析的表 1 中得以体现.DiscoGAN 的生成结果可以较清晰地对字符进行提取,但是与本文的方法相比,其生成字形的完整性还存在一定的不足.因此,该方法得到了较好的识别准确率 52.52%,但 mAP50 仅达到 26.35%.相较之下,本文的方法在所有指标上可以得到最接近专家摹本的分, mAP50 为 35.08%,准确率为 58.61%,这说明本文的方法在提取高频特征生成甲骨字形上具有更好的表现.

3.4 消融实验

在消融实验中,以 1 000 次迭代作为放缩阶段和切片阶段的完整循环,通过调整切片阶段在整个周期中的占比,分析切片操作与放缩操作对摹本生成结果的影响.实验结果如图 2 所示,当 t_2 占比为 0% 时,模型完全使用放缩策略进行训练,生成摹本可以完好地实现甲骨轮廓描绘,对于摹本的全局特征得到了较好地提取.但是,模型对字符细节的把控能力很弱,生成字符通常具有较大形变.随着 t_2 占比的增加,生成摹本中的

字符与拓片字符一致性在逐渐增强.但是,当 t_2 占比达到 50% 之后,尽管拓片上的字符得到了很好的体现,切片输入却严重影响了模型对生成摹本整体性的把握,摹本在甲骨片以外的部分开始出现大量错误纹理.特别是当 t_2 占比达到 100% 时,生成图像已经不能保证维持甲骨片的正常轮廓.通过对周期占比的消融实验分析,发现 t_2 占比不宜过高,维持在 40% 左右时,可以实现整体轮廓与字符细节提取的均衡.

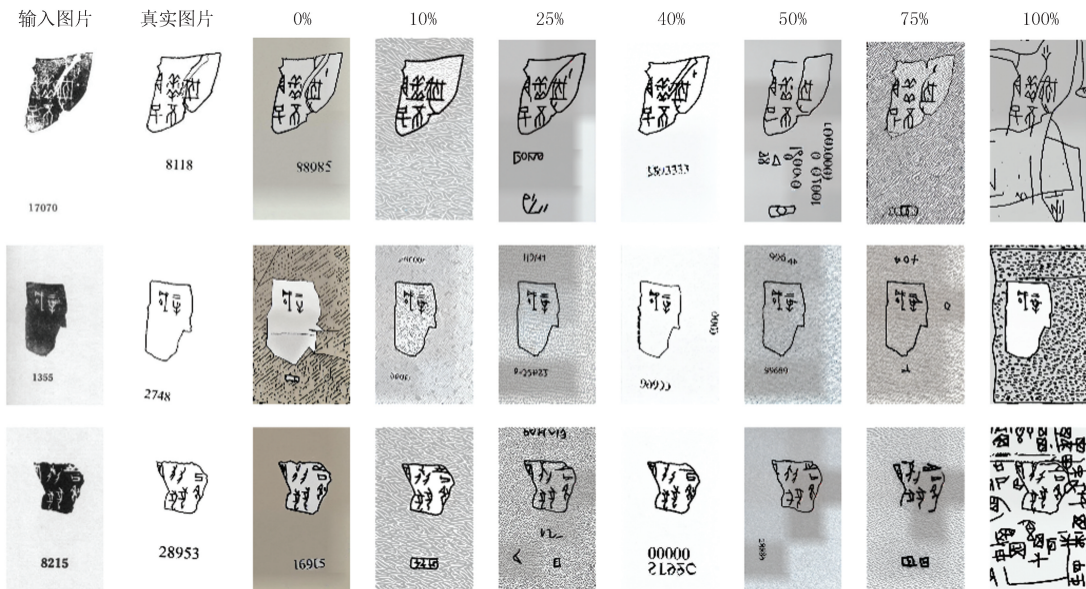


图2 不同的放缩阶段 t_2 占比生成摹本对比图

Fig.2 Comparison chart of replicas generated at different scaling stages t_2 ratio

为了展示引导系数在噪声扩散中起到的作用,我们对引导系数的不同取值对摹本生成结果的影响进行了对比,如图 3 所示.当引导系数比较小的时候,摹本生成结果的颜色偏暗,存在笔画粘连的情况,随着引导系数的增大,整体的对比度提高,并且粘连的笔画分开.当引导系数过大的时候,生成的摹本图片会出现失真,并且存在一定的笔画丢失、轮廓丢失等问题.实验结果表明,引导系数在 28 左右时能够达到最佳的平衡点,既能够有效增强图像对比度,又能够避免过度失真及细节丢失,生成的摹本图像在保持拓片特征的同时,清晰度和结构完整性得到了最优的保障.

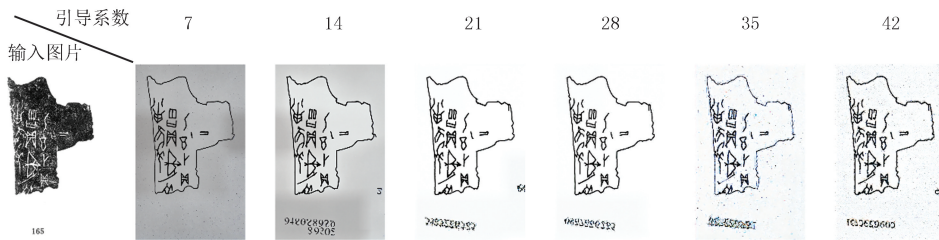


图3 使用不同的引导系数生成的拓片图像

Fig.3 The replicas images generated with different guidance scale

4 总结

本文提出了一种基于扩散模型的多尺度交替训练策略用于甲骨摹本生成,通过引导模型在不同尺度下协同关注字符细节与整体图像特征,有效缓解了字符密度不均导致的信息丢失问题.同时,引入零终端信噪比机制以控制分布差异性,增强了生成摹本中甲骨字符的表达能力.实验结果表明,所生成摹本在字符细节、笔画形态及整体轮廓还原方面均优于传统方法,且相较于人工摹写显著减少了笔画形变现象.该方法在文字数量、位置结构、笔画细节及甲骨片边缘轮廓等方面表现出良好的准确性与鲁棒性,可显著提升甲骨摹本的制作效率,为甲骨文字形研究及相关工具书编纂提供了有力的技术支持.

参 考 文 献

- [1] ZENG X, CHENG L, LI S, et al. Archaeology drawing generation algorithm based on multi-branch feature cross fusion[EB/OL]. [2025-01-13]. <http://dx.doi.org/10.21203/rs.3.rs-4409621/v1>.
- [2] ZHAO Z, WANG Y, LIU K, et al. Semantic segmentation by improved generative adversarial networks[EB/OL]. [2024-12-05]. <https://doi.org/10.48550/arXiv.2104.09917>.
- [3] 刘国奇, 李文格, 茹琳媛, 等. FEC-PVT: 基于 PVT 架构的甲骨钻凿图像分割网络[J]. 河南师范大学学报(自然科学版), 2026, 54(1): 8-17.
- LIU G Q, LI W G, RU L Y, et al. FEC-PVT: An oracle bone drilling image segmentation network based on PVT architecture[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(1): 8-17.
- [4] ENAYATI A, PALMIERI L, VASCON S, et al. Semantic motif segmentation of archaeological fresco fragments[C]//2023 IEEE/CVF International Conference on Computer Vision Workshops(ICCVW). Paris: IEEE, 2023: 1709-1717.
- [5] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[EB/OL]. [2024-11-13]. <https://arxiv.org/abs/2006.11239>.
- [6] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New Orleans: IEEE, 2022: 10674-10685.
- [7] WU W J, ZHAO Y Z, SHOU M Z, et al. DiffuMask: synthesizing images with pixel-level annotations for semantic segmentation using diffusion models[C]//2023 IEEE/CVF International Conference on Computer Vision(ICCV). Paris: IEEE, 2023: 1206-1217.
- [8] TIAN J J, AGGARWAL L, COLACO A, et al. Diffuse, attend, and segment: unsupervised zero-shot segmentation using stable diffusion [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2024: 3554-3563.
- [9] ZHANG L M, RAO A Y, AGRAWALA M. Adding conditional control to text-to-image diffusion models[C]//2023 IEEE/CVF International Conference on Computer Vision(ICCV). Paris: IEEE, 2023: 3813-3824.
- [10] LI B, LUO D, LIANG Y, et al. Oracle Bone Inscriptions Multi-modal Dataset[EB/OL]. [2024-11-13]. <https://arxiv.org/abs/2407.03900>.
- [11] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision(ICCV). Venice: IEEE, 2017: 2223-2232.
- [12] KIM T, CHA M, KIM H, et al. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks[EB/OL]. [2025-01-13]. <https://doi.org/10.48550/arXiv.1703.05192>.
- [13] BOESCH G. YOLO11: A New Iteration of 'You Only Look Once'[EB/OL]. [2024-11-13]. <https://viso.ai/computer-vision/yolov11/>.

Research on generation technology of oracle bone facsimile images based on diffusion model

Li Bang^{a,b,c}, Yang Jing^a, Ding Zengmao^a, Fang Jiaqi^a, Han Shengwei^a, Cai Jiase^a,
Zhang Zhan^{a,b,c}, Guo An^{a,b,c}, Gao Feng^{a,b,c}, Liu Yongge^{a,b,c}

(a. Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China; b. School of Computer & Information Engineering; c. Henan International Joint Laboratory for Intelligent Processing of Perceptual Data, Anyang Normal University, Anyang 455000, China)

Abstract: Oracle bone facsimiles are an essential resource in oracle bone inscription studies, serving to reproduce surface information such as contours, characters, and cracks of oracle bones. However, the traditional process of producing oracle bone facsimiles is time-consuming, labor-intensive, and highly dependent on expert knowledge. To address these challenges, this paper proposes a diffusion-model-based training strategy with character information density normalization, which enables the effective extraction of fine-grained character texture features with the preservation global image textures such as contours. Furthermore, to emphasize oracle bone characters in facsimile generation, a zero-terminal signal-to-noise ratio mechanism is introduced to control distributional discrepancies during diffusion. Experimental results demonstrate that the proposed method achieves higher image generation quality and improved spatial efficiency in oracle bone facsimile generation tasks, providing strong technical support for oracle bone character form analysis.

Keywords: oracle bone script; image generation; diffusion model; facsimile