

DOI: 10.13718/j.cnki.xdzk.2025.06.020

吴广, 王云香, 刘礼, 等. 融合图像分割与风格迁移的文物数字拓片技术探索 [J]. 西南大学学报(自然科学版), 2025, 47(6): 224-235.

# 融合图像分割与风格迁移的 文物数字拓片技术探索

吴广<sup>1,2</sup>, 王云香<sup>2</sup>, 刘礼<sup>2</sup>, 杨平恒<sup>3</sup>

1. 重庆市文物考古研究院, 重庆 400013; 2. 重庆大学 大数据与软件学院, 重庆 401331;
3. 西南大学 地理科学学院, 重庆 400715

**摘要:** 随着文化遗产数字化需求的不断增长, 传统手工拓片技术因存在潜在风险、操作复杂以及保存不便等问题, 已难以满足现代文物保护与展示的高标准要求。全新数字拓片技术凭借其高效、安全的特性, 在文化遗产保护领域展现出显著的应用价值, 为文物信息的数字化呈现提供了全新路径。为提升数字拓片的生成质量与效率, 提出了一种基于图像分割与风格迁移的创新方法。该方法通过结合 Segment Anything Model(SAM)的精准图像分割技术和基于深度学习的文物数字拓片风格迁移网络, 实现了文物内容与拓片风格的高度融合。具体而言, SAM 模型凭借其强大的泛化能力和 Prompt-based 交互式分割能力, 能够快速准确地分割出文物图像中的目标对象。在文物数字拓片风格迁移方面, 采用基于深度学习的卷积神经网络(CNN)架构, 通过训练神经网络学习特定拓片风格的特征, 将这些特征应用到目标图像上, 从而生成具有类似风格的数字拓片图像。技术架构采用编码器-解码器设计, 并引入拓片风格注意力模块, 同时利用细节保留损失函数优化模型训练, 确保生成的数字拓片既能保留文物的细节特征, 又能精准再现传统拓片的风格特质。实验结果表明, 该技术显著提高了自动化水平, 能够快速生成高质量的数字拓片。与传统图像处理和三维重构技术相比, 该方法无接触式操作避免了对文物的物理损害, 同时在细节保留和风格一致性方面取得良好平衡, 具有大规模应用的潜力, 并降低了总体成本。

**关键词:** 数字拓片技术; 文化遗产保护; 图像分割; 风格迁移

中图分类号: K877.4; TP391.41

文献标识码: A

文章编号: 1673-9868(2025)06-0224-12

开放科学(资源服务)标识码(OSID):



## Digital Rubbing Technology for Cultural Relics: Integrating Image Segmentation and Style Transfer

WU Guang<sup>1,2</sup>, WANG Yunxiang<sup>2</sup>, LIU Li<sup>2</sup>, YANG Pingheng<sup>3</sup>

收稿日期: 2024-11-05

基金项目: 重庆市科技局绩效激励引导专项课题(cstc2022jxj00022); 教育部人文社会科学研究规划基金项目(23YJA780003)。

作者简介: 吴广, 博士研究生, 副研究馆员, 主要从事文物考古数字化等方面的研究。

1. *Chongqing Cultural Relics and Archaeology Research Institute, Chongqing 400013, China;*
2. *School of Big Data & Software Engineering, Chongqing University, Chongqing 401331, China;*
3. *School of Geographic Sciences, Southwest University, Chongqing 400715, China*

**Abstract:** In the face of the burgeoning demand for the digitalization of cultural heritage, traditional manual rubbing technique has become increasingly inadequate in meeting the high standard requirement of contemporary cultural heritage protection and exhibition. This is attributable to its inherent risks, intricate operational procedures, and challenges in preservation. The novel digital rubbing technology, characterized by its efficiency and safety, has manifested remarkable application value within the realm of cultural heritage protection, thereby charting a new course for the digital representation of cultural heritage information. To enhance the quality and efficiency of digital rubbing generation, this paper proposes an innovative approach grounded on image segmentation and style transfer. This method achieved a high degree of integration between cultural heritage content and rubbing style by integrating the precise image segmentation capabilities of the Segment Anything Model (SAM) with a deep learning based style transfer network for digital cultural heritage rubbings. Specifically, the SAM model, leveraging its robust generalization ability and Prompt-based interactive segmentation prowess, is capable of swiftly and accurately segmenting the target objects within cultural heritage images. In the domain of style transfer for digital cultural heritage rubbings, a convolutional neural network (CNN) architecture founded on deep learning is employed. Through training the neural network to learn the distinctive features of specific rubbing styles and applying these features to target images, digital rubbing images with analogous styles are generated. The technical architecture incorporates an encoder-decoder design and a rubbing style attention module, and utilizes a detail preservation loss function to optimize model training. This ensures that the generated digital rubbings not only preserve the intricate details of cultural relics but also faithfully reproduce the stylistic characteristics of traditional rubbings. Experimental outcomes demonstrate that this technology significantly elevated the level of automation and is capable of rapidly producing high quality digital rubbings. When compared with traditional image processing and 3D reconstruction techniques, this non-contact method obviates the risk of physical damage to cultural relics. Moreover, it strikes a favorable balance between detail retention and style consistency, exhibits potential for large scale implementation, and reduces overall costs. This technology thus offers an efficient and innovative solution for the digital protection and dissemination of cultural heritage.

**Key words:** digital rubbing technology; cultural heritage preservation; image segmentation; style transfer

拓片堪称中国传统捶拓技艺的精髓,它能够把碑刻、墓碣、摩崖石刻以及器物上的文字与图案非常精准地复制到宣纸之上,进而变成可以反映不同时代社会文化、政治、经济、宗教以及艺术等诸多方面信息的一种重要的文献载体<sup>[1]</sup>。拓片一方面蕴含着卓越的工艺美术学价值,另一方面在历史学、考古学、书法研究等诸多领域占据着不可替代的地位。然而,传统手工拓片方法在实际应用中存在不足,比如对文物本体有可能会造成潜在的危害,对操作者自身的技艺水平有着很高的要求,而且对操作环境也有着颇为严格的要求(图 1)。除此之外,拓片的保存以及传播同样面临着重重困难,如采用折叠保存或者卷轴装裱的方式就会致使材料出现脆裂或者损坏的情况,所使用的黏合剂也容易使得纸张出现受潮、氧化以及霉变等问题,拓片遭受虫害以及人为损坏的风险也比较大,在使用以及展示的时候也有很多不便<sup>[2]</sup>。数字拓片技术

能够很好解决上述问题。

图像分割属于计算机视觉领域里的一项基础技术,其作用在于能够把图像分割成为语义清晰的各个区域,进而为后续目标对象的提取以及相关处理工作打下基础。近些年,借助深度学习方法,使得分割工作在精度以及效率这两方面都有了颇为明显的提升<sup>[3-4]</sup>。而 Segment Anything Model(SAM)作为当下最新的一项技术,凭借其自身通用性以及交互式优化能力,从众多技术当中脱颖而出,尤其适用于复杂图像特征分割处理,如此一来,能够在很大程度上减少人工方面的干预,并且还能有效提升工作的效率<sup>[5]</sup>。



图1 野外文物手工拓片

近年来,风格迁移技术在图像生成与艺术风格转换领域取得显著进展,其核心突破在于通过神经网络实现内容与风格的分离与融合。文献<sup>[6]</sup>提出的基于卷积神经网络的风格迁移方法,为该领域的发展奠定了理论基础,而 Adaptive Instance Normalization(AdaIN)则是在此基础上进一步达成了实时风格迁移的效果<sup>[7]</sup>。最新出现的自注意力生成对抗网络(SAGAN),将自注意力机制引入传统生成对抗网络,显著提升了生成图像的细节质量与风格表现力,使模型在生成过程中能够更好地捕捉全局上下文信息,生成更加逼真、细腻且风格多样的图像<sup>[8]</sup>。

文物数字拓片要求既要完好保留文物图像自身的纹理细节,又要再现传统拓片的风格,这无疑给风格迁移模型设定了更高层级的要求,也是本文围绕风格注意力模块展开深入探究的核心出发点。

本文对图像分割技术以及风格迁移技术加以分析,提出一种将图像分割和风格迁移相结合的创新方法,在提升文物数字拓片生成质量的同时提高生成效率,进而为文物保护、数字化留存及文化传播提供全新方案。

## 1 数字拓片的相关概念

### 1.1 数字拓片与拓片数字化

数字拓片侧重于通过数字化手段再现文物的物理特征,拓片数字化则是将传统拓片转化为数字资产,二者在技术路径与应用目标上存在本质区别。在文物保护数字化领域,“数字拓片”具体指借助计算机图像处理技术,对文物(如石刻等)的数字影像进行处理,生成具有传统拓片视觉特征的数字图像<sup>[9]</sup>。该方法通过数字技术重现传统拓片的外观与细节,其核心在于直接从文物本体获取数字影像,而非传统纸质拓片。

“拓片数字化”就是把已存在的传统纸质拓片,借助影像扫描或摄影等手段,将其转变为数字格式。在此基础上,还要对这些转化后的数据展开存储、管理以及处理等相关操作。如此一来,便能达成传统拓片数字化保存并利用的目的。此过程实际上就是对传统拓片这一物理载体进行数字记录。

虽说两者均牵涉到数字化技术的运用,可它们的核心差异体现在目标以及对象方面:“数字拓片”着重于凭借数字手段通过文物图像生成具有传统拓片视觉效果的数字图像;而“拓片数字化”则是把已有的纸质拓片转变成利于存储和管理的数字格式。

### 1.2 数字拓片技术概况

与传统拓片相比,数字拓片技术的数据采集、处理与输出过程具有高效且不受现场环境限制的特点,生成图像的质量与精度具有较高可控性,工艺适用性广泛,并支持二维矢量纹理转换<sup>[10]</sup>。此外,数字拓片生成的图像效果更加清晰逼真,纹饰与字体的线刻细节更为丰富,立体感也更加突出。

目前,数字拓片生成技术主要分为两大类。一类基于图像处理技术,通过高分辨率摄影或扫描获取浮雕文物的图像数据,再利用计算机视觉与图像处理算法生成具有传统拓片效果的数字图像。这类方法侧重于二维信息的提取与处理,适用于纹理复杂、表面反光较强的文物,其优势在于操作简便、成本较低,且能够快速生成高质量的数字拓片。另一类采用三维扫描与重构技术,这类方法能够捕捉文物的微小形态特征,保留丰富的空间信息,尤其适用于形态复杂的立体文物。通过获取文物表面的三维形态数据,进行高精度重建与纹理映射,从而实现数字拓片的高保真再现<sup>[11]</sup>。

图像技术在数字拓片处理中的应用主要涉及文物的图像采集与后期处理。首先,通过高分辨率摄影或摄像技术获取文物的图像数据。随后,对采集到的刻文图片进行一系列精细化处理,包括图片倾斜校正、背景纹理处理、字符的自动或手工校正与分割、风格化处理以及按实际尺寸进行比例缩放等。这些处理旨在确保生成的数字拓片在细节呈现上达到高保真标准,从而为文化遗产的数字化保护与研究提供高质量的图像资料。早期文物图像处理多依赖 Photoshop 等图像编辑软件,然而,这一过程需要大量人工操作,效率较低<sup>[12]</sup>。为提升自动化水平,文献[13]通过 MATLAB 编程实现局部阈值分割与均值滤波,初步实现了数字拓片的自动生成,但该方法仍需人工干预,自动化程度有限。此外,此类方法对图像质量要求较高,光照不均匀和噪声干扰易导致处理效果不稳定。文献[14]提出了一种基于三维模型的碑刻数字拓片生成技术。通过对碑刻三维模型的特征进行分析,研究了基于顶点深度和法向量的数字拓片生成方法,并提出了基于深度图的优化方案,有效提升了生成质量。然而,当碑刻表面风化严重时,平面拟合生成的投影面可能出现偏差,导致该方法失效。此外,仅依赖顶点深度生成的深度图,可能无法有效解决因顶点法向量导致的边缘锯齿问题。尽管已有多种类似方法,但这些方法通常对图像拍摄的环境、分辨率和光照条件要求较高,且自动化程度有限。

基于三维扫描及重构技术的数字拓片方法利用非接触式扫描仪对文物的空间外形、结构及色彩进行数字化采集,获取其表面点的三维坐标值,即点云信息。将扫描得到的点云数据输入相关数据处理软件,经过预处理和三维重建,在虚拟空间中创建文物的三维模型。这种方法通过三维激光扫描设备或基于碑刻图像进行三维重建的碑刻三角网格生成数字拓片。例如,文献[15]提出了一种基于激光点云和近景影像的数字拓片生成方法,该方法通过点云和影像配准生成正射影像,并采用图像增强、局部自适应阈值二值化和图像融合技术,生成具有传统拓片效果的数字拓片。然而,对于大规模碑刻,该方法可能难以保持精度,并且需要人工选择特征点进行配准,可能引入人为误差,效率也较低。文献[16]使用高精度三维扫描设备获取墓志的三维空间信息,转换成三角网格模型后,手工选点勾勒碑刻形状,并利用图像处理技术填充三角面片,最终生成具有传统拓片视觉特征的数字图像。此方法依赖于高采样率、高分辨率和高精度的三维扫描设备,但生成的数据量庞大,需专业处理才能恢复文物原貌。此外,将三维数据映射到二维平面以制作刻文数字拓片,仍需进一步探索相关技术。

## 2 基于图像分割与风格迁移的数字拓片生成方法

本文基于图像分割与风格迁移相结合的方法,开发了一种深度学习模型,用于生成数字拓片图像。该方法的核心环节包括:精确分割文物主体图像,设计并构建一个集成编码器—解码器与风格注意力模块的文物数字拓片风格迁移网络,最终实现全数字化、智能化的一站式数字拓片生成方案。

在图像分割阶段,采用了 Facebook 的 SAM 模型,以实现图像中特定文物对象的精确分割。内容图像经过 SAM 模型的分析,精确识别并界定了分割区域。随后,根据用户定义的文物主体,通过将图像与生成的二值化分割掩码相结合,生成了一个专注于目标对象的增强型分割掩模。

在文物数字拓片风格迁移阶段,将图像分割阶段生成的增强型分割掩模、内容图像和风格图像一同输入到文物数字拓片风格迁移网络。本文设计的文物数字拓片风格迁移网络由编码器—解码器模块和风格注

注意力模块组成, 能够高效地产出既具有高质量又符合特定风格要求的图像。这些图像不仅反映了全局风格特征, 还捕捉到了局部细节。此外, 本文还引入了一种新颖的细节保留损失函数, 它在确保风格得到充分体现的同时, 也维护了内容图像的细节结构, 实现了内容与风格的完美融合。

## 2.1 图像分割

Segment Anything Model(SAM)是一种先进的图像分割模型, 它能够处理各种各样的图像分割任务, 其独特优势在于能够根据用户提供的输入点或边界框动态生成高精度分割掩码。该模型采用一个基于视觉 Transformer (ViT)<sup>[18]</sup> 的图像编码器来提取图像嵌入, 一个提示编码器来整合用户交互, 以及一个轻量级掩码解码器来预测分割掩码<sup>[19]</sup>, 各部件见图 2。

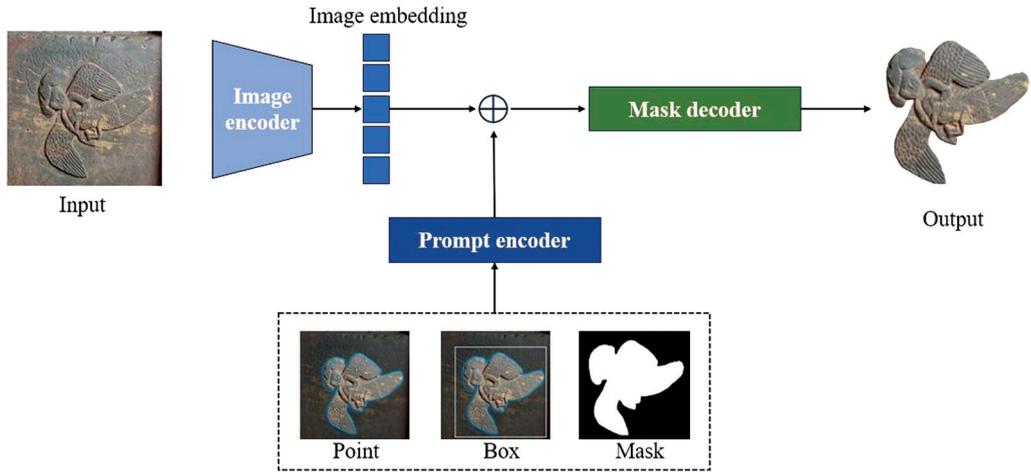


图 2 SAM 架构

SAM 考虑了两种类型的提示, 包括稀疏提示(如点、框)和密集提示(如掩码)<sup>[19]</sup>。具体而言, SAM 以原始图像 Input 作为输入, 采用位置编码与学习嵌入相结合来表示点和框。这些点由坐标  $(x_i, y_i)$  及其目标标签  $l_i$  给定, 或者通过由左上角  $(x_{lt}, y_{lt})$  和右下角  $(x_{rb}, y_{rb})$  确定的边界框来划定图像中的感兴趣区域。

在这一过程中, SAM 生成了一个分割掩码, 该掩码主要是用来标识出图像中前景对象所在区域及其和背景区域之间的边界。SAM 还允许用户跟模型展开交互操作, 用户可以持续添加或调整输入点和边界框, 如此一来, 便能逐步提升分割结果的精准程度。要是一开始生成的掩码无法准确区分前景和背景, 那么用户可以增添额外更多前景或背景点来改进分割效果。正是这样的交互式优化机制, 让 SAM 在各种各样的图像场景之下, 都能够给出高度精准的分割结果。

关于这一步骤的软件实现, 利用了官方 SAM GitHub 仓库中的可用代码。将该模型与 Python 集成非常简便, 用户只需将图像作为输入提供给模型, SAM 就会根据输入的提示条件自动生成所有必要的分割掩码。

最终, SAM 所生成的二值化分割掩码  $S$  将图像中的感兴趣对象区域标记为 1, 背景区域标记为 0。结合分割掩码  $S$  与原始图像  $Y$  可以得到分割后的图像  $Y^{\text{seg}}$ :

$$Y^{\text{seg}} = I \times Y + (1 - Y) \times 255 \quad (1)$$

分割后的图像  $I^{\text{seg}}$  中仅保留了原图像中感兴趣对象的像素值, 而背景部分则被设置为白色。

## 2.2 文物数字拓片风格迁移

文物数字拓片风格迁移网络用于将文物的正射图像  $Y^c$  (此处为经过图像分割处理的图片  $Y^{\text{seg}}$ ) 与拓片风格模板  $Y^s$  融合, 生成具有传统拓片风格的数字拓片  $Y^{\text{cs}}$ 。例如, 输入的图像是一块宋代文物浮雕, 上面有“羽人”浮雕图案, 而拓片风格模板是一张传统的手工拓片。预训练的 VGG-19 网络作为编码器提取浮雕图像的多层特征表示, 捕捉文物的形状和细节。然后, 使用对称解码器(遵循文献[7]中的设置)将嵌入拓片风格特征后的特征图逐步重建为高质量的数字拓片图像。拓片风格注意力网络通过注意力机制将拓片的墨色

深浅和纹理特征灵活嵌入到浮雕的鸟图案中,生成具有传统拓片风格的数字拓片图像。整个过程通过端到端的训练,确保生成的数字拓片在保留文物的细节的同时,精准再现拓片的墨色和纹理风格。

文物数字拓片风格迁移网络的结构如图 3 所示,输入为文物的正射图像  $Y^c$  和风格拓片模板  $Y^s$ 。网络利用预训练的 VGG-19 编码器分别提取这两张图像的特征,生成文物内容特征  $E^c = E(Y^c)$  和拓片风格特征  $E^s = E(Y^s)$ 。用  $E_{i,j}^c$  和  $E_{i,j}^s$  分别表示 Conv $i$ - $j$  层输出的文物内容特征图和拓片风格输出特征图。为了有效融合文物图像的语义信息与拓片的风格特征,选取 Conv4-1 和 Conv5-1 层的输出特征图( $E_{4,1}^c$ 、 $E_{4,1}^s$ 、 $E_{5,1}^c$  和  $E_{5,1}^s$ )作为后续拓片风格注意力网络的输入。其中,Conv4-1 层的特征图分辨率较高( $28 \times 28$ ),能够保留较多的细节信息,适用于文物正射图内容特征的提取;而 Conv5-1 层虽分辨率较低( $14 \times 14$ ),但其语义表达能力更强,有助于捕捉拓片风格图像的全局与局部风格特征。

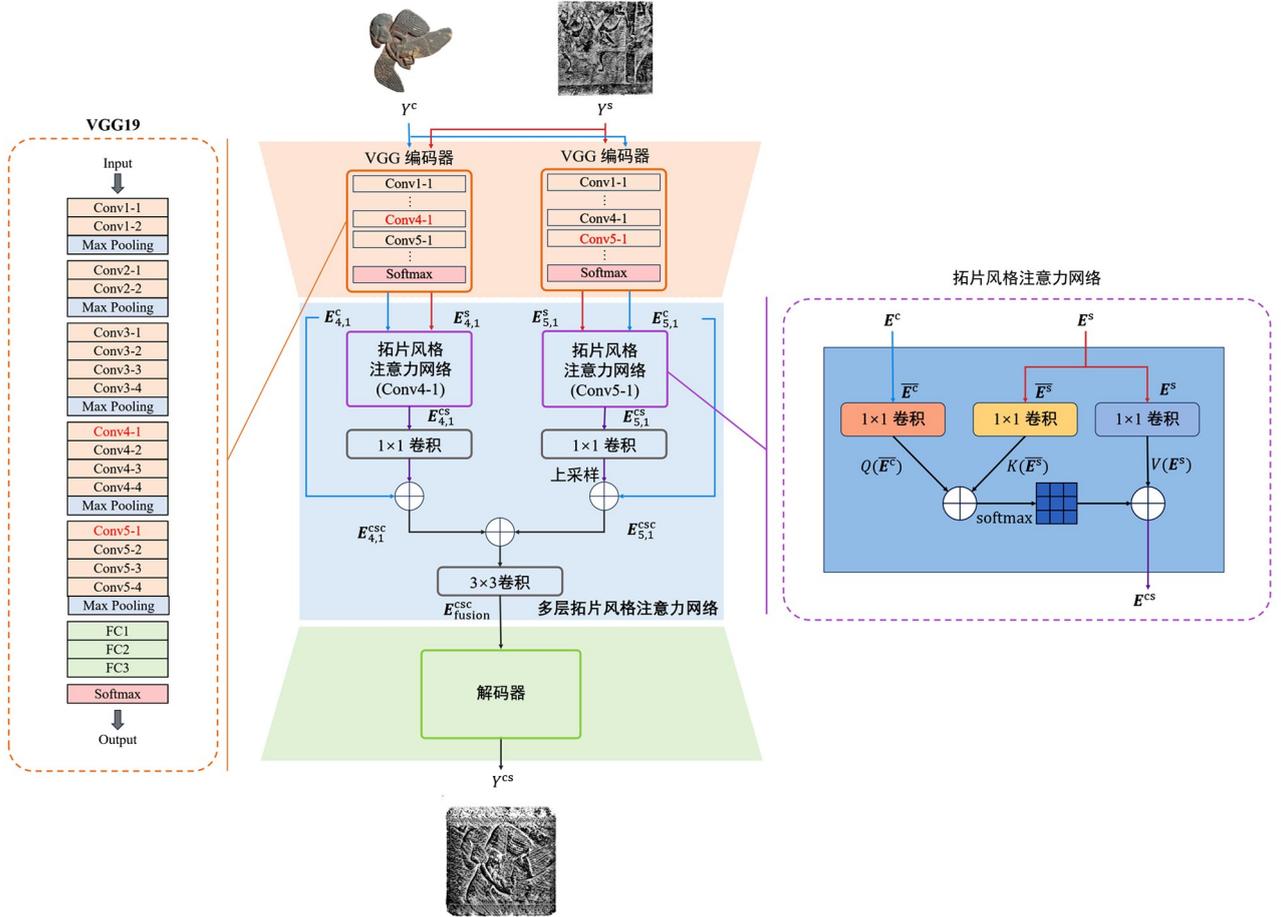


图 3 文物数字拓片风格迁移网络框架图

在特征提取后,文物内容特征  $E^c$  (如“羽人”的翅膀纹理、头部线条等)和拓片风格特征  $E^s$  (如墨色的深浅变化、线条的疏密安排等)被输入至拓片风格注意力网络模块,用于建立二者之间的显式映射关系,从而生成文物数字拓片特征图  $E^{cs}$ 。该网络结构在文献[20]中首次提出,计算公式如下:

$$E_m^{cs} = \sum_{n=1}^d \text{softmax}(Q(\overline{E}_m^c)^\top K(\overline{E}_n^s)) V(E_n^s) \quad (2)$$

式中:  $E^c$ 、 $E^s \in \mathbb{R}^{d \times d}$ ;  $E_m^c$ 、 $E_n^s$  分别表示矩阵  $E^c$  和矩阵  $E^s$  的第  $m$  列和第  $n$  列,其中  $m$ 、 $n \in [1, d]$ ,两者均为一个  $d \times 1$  的列向量;  $Q(\overline{E}^c)$ 、 $K(\overline{E}^s)$ 、 $V(E^s)$  为  $\overline{E}^c$ 、 $\overline{E}^s$ 、 $E^s$  分别与学习到的权重矩阵  $W_Q$ 、 $W_K$  和  $W_V$

的乘积;  $\text{softmax} = \frac{\exp(Q(\overline{E}_m^c)^\top K(\overline{E}_n^s))}{\sum_{j=1}^d \exp(Q(\overline{E}_m^c)^\top K(\overline{E}_j^s))}$ 。 $\overline{E}$  表示  $E$  的均值一方差通道归一化版本,其元素  $\overline{e}_{ij}$  的计算

公式如下:

$$\overline{e_{ij}} = \frac{e_{ij} - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (3)$$

式中:  $e_{ij}$  为  $\mathbf{E}$  的矩阵元素,  $\mathbf{E} \in \mathbb{R}^{H \times W}$  表示输入的特征,  $H$ 、 $W$  为输入特征图的高和宽;  $\mu = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \mathbf{E}_{ij}$

为输入特征  $\mathbf{E}$  的均值;  $\sigma^2 = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (\mathbf{E}_{ij} - \mu)^2$  为输入特征  $\mathbf{E}$  的方差;  $\epsilon$  是一个非常小的正则化项, 用于防止分母为零, 确保数值稳定性。

随后, 经过一次  $1 \times 1$  卷积操作后, 采用以下公式进行逐元素加权融合, 得到新的文物数字拓片特征图  $\mathbf{E}^{\text{csc}}$ :

$$\mathbf{E}^{\text{csc}} = \mathbf{E}^c \odot (\mathbf{I} + \mathbf{D} \cdot \mathbf{E}^{\text{cs}}) \quad (4)$$

式中:  $\odot$  表示逐元素乘法;  $\mathbf{I}$  为单位矩阵;  $\mathbf{D}$  为学习到的权重矩阵。

通过拓片风格注意力网络输出两个文物数字拓片特征图  $\mathbf{E}_{4,1}^{\text{csc}}$  和  $\mathbf{E}_{5,1}^{\text{csc}}$ 。首先对来自 Conv5-1 层的文物数字拓片特征图  $\mathbf{E}_{5,1}^{\text{csc}}$  进行上采样, 使其分辨率与来自 Conv4-1 层的文物数字拓片特征图  $\mathbf{E}_{4,1}^{\text{csc}}$  一致。然后将两者逐元素相加, 以融合不同层次的特征信息。最后, 将相加后的特征图通过一个  $3 \times 3$  的卷积层进行处理, 以整合信息并生成最终的融合特征图  $\mathbf{E}_{\text{fusion}}^{\text{csc}}$ , 其计算公式如下:

$$\mathbf{E}_{\text{fusion}}^{\text{csc}} = \text{Conv}_{3 \times 3}(\mathbf{E}_{4,1}^{\text{csc}} + \text{Upsampling}(\mathbf{E}_{5,1}^{\text{csc}})) \quad (5)$$

这一过程有效地结合了高分辨率的文物细节信息和低分辨率的拓片风格语义信息, 为生成高质量的数字拓片提供了基础。最后, 将融合后的文物数字拓片特征图  $\mathbf{E}_{\text{fusion}}^{\text{csc}}$  输入至解码器, 生成最终的数字拓片风格图像  $Y^{\text{cs}}$ 。

文物数字拓片风格迁移网络的总体损失函数如下:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{content}} + \beta \mathcal{L}_{\text{style}} + \mathcal{L}_{\text{detail}} \quad (6)$$

式中:  $\mathcal{L}_{\text{content}}$ 、 $\mathcal{L}_{\text{style}}$ 、 $\mathcal{L}_{\text{detail}}$  分别表示文物内容损失、拓片风格损失和细节保留损失;  $\alpha$  和  $\beta$  分别是不同损失的权重。

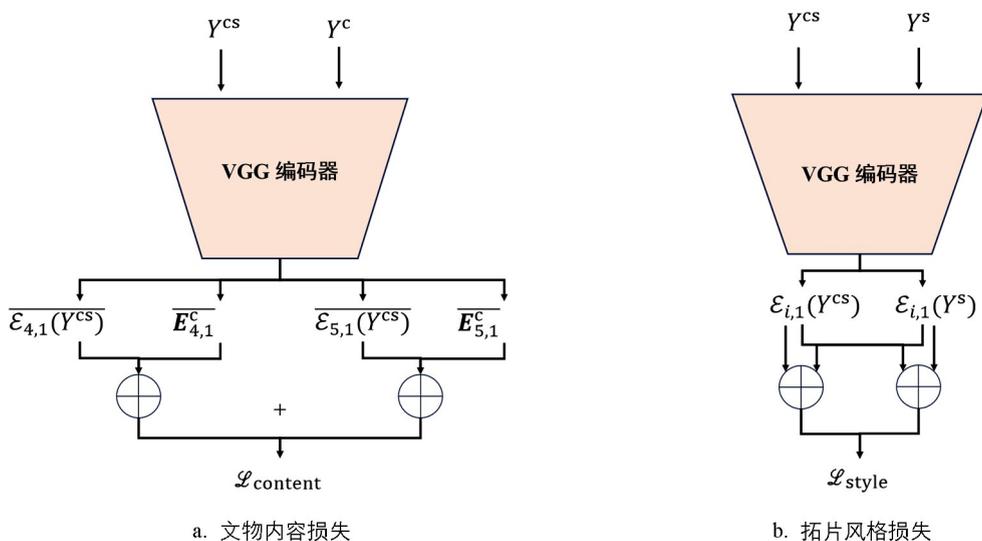


图 4 内容与风格损失

文物内容损失  $\mathcal{L}_{\text{content}}$  用计算归一化后的生成图像特征和原始文物图像特征之间的 Frobenius 范数, 可以确保生成的数字拓片图像与文物正射图像在高级语义特征(如形状、布局等)上保持一致(图 4a), 其计算公式如下:

$$\mathcal{L}_{\text{content}} = \|\overline{\mathbf{E}_{4,1}(Y^{\text{CS}})} - \overline{\mathbf{E}_{4,1}^{\text{c}}}\|_2 + \|\overline{\mathbf{E}_{5,1}(Y^{\text{CS}})} - \overline{\mathbf{E}_{5,1}^{\text{c}}}\|_2 \quad (7)$$

式中  $\mathbf{E}_{i,1}(Y)$  表示 Conv $_{i-1}$  层图像  $Y$  输入到编码器的第  $i$  层后得到的特征图。

拓片风格损失  $\mathcal{L}_{\text{style}}$  可提取拓片风格模板图像的纹理、色彩分布和视觉特征(如墨色的深浅、拓片的质感等),并将其融入到生成图像中(图 4b),其计算公式如下:

$$\mathcal{L}_{\text{style}} = \sum_{i=1}^L \|\text{mean}(\mathbf{E}_{i,1}(Y^{\text{CS}})) - \text{mean}(\mathbf{E}_{i,1}(Y^{\text{s}}))\|_2 + \|\text{std}(\mathbf{E}_{i,1}(Y^{\text{CS}})) - \text{std}(\mathbf{E}_{i,1}(Y^{\text{s}}))\|_2 \quad (8)$$

式中:  $\mathbf{E}_{i,1}(Y)$  表示 Conv $_{i-1}$  层图像  $Y$  输入到编码器的第  $i$  层后得到的特征图,此处  $i=1, 2, \dots, 5$ ;  $L$  是用于计算风格损失的层数; mean 和 std 分别表示平均值和标准差函数。

为了综合考虑全局统计特征与文物内容特征和拓片风格特征之间的语义局部映射,本文引入了文献[20]中的身份损失函数。在本文的场景中,该函数能够有效比较生成的数字拓片图像与原始文物图像在像素级和特征级上的差异。此外,为了保留文物图像的高频细节(如纹理和边缘信息),本文进一步引入了梯度损失,并通过计算平方欧几里得距离来优化细节保留效果。利用如下细节损失  $\mathcal{L}_{\text{detail}}$  来确保生成拓片图像在细节上与原始文物图像保持一致。

$$\begin{aligned} \mathcal{L}_{\text{detail}} = & \lambda_{\text{detail1}} (\|Y^{\text{cc}} - Y^{\text{c}}\|^2 + \|Y^{\text{ss}} - Y^{\text{s}}\|^2) + \\ & \lambda_{\text{detail2}} \sum_{i=1}^L (\|E_i(Y^{\text{cc}}) - E_i(Y^{\text{c}})\|^2 + \|E_i(Y^{\text{ss}}) - E_i(Y^{\text{s}})\|^2) + \\ & \lambda_{\text{gradient}} (\|\nabla Y^{\text{cc}} - \nabla Y^{\text{c}}\|^2 + \|\nabla Y^{\text{ss}} - \nabla Y^{\text{s}}\|^2) \end{aligned} \quad (9)$$

式中:  $Y^{\text{cc}}$ (或  $Y^{\text{ss}}$ ) 表示由两个相同的内容(或风格)图像合成的输出图像;每个  $E_i$  表示编码器中的一个层;  $\lambda_{\text{detail1}}$  用于调整文物图像与拓片风格像素级的细节保留损失,  $\lambda_{\text{detail2}}$  用于调整文物图像与拓片风格特征级的细节保留损失;  $\lambda_{\text{gradient}}$  是梯度损失的权重;  $\|\nabla Y^{\text{cc}} - \nabla Y^{\text{c}}\|^2$  和  $\|\nabla Y^{\text{ss}} - \nabla Y^{\text{s}}\|^2$  分别表示生成数字拓片图像与原始文物图像在梯度上的差异,用以保留高频细节。这种细节可以确保文物轮廓图像清晰锐利,如保留“羽人”浮雕的翅膀、羽毛等精细纹理,使生成的拓片能够真实展现文物的原始质感。梯度损失在优化过程中能增强局部对比度,使文物主体与背景之间的界限更加分明,提升了生成图像的清晰度和视觉效果。

通过协同引入内容损失、风格损失与细节保留损失,模型在训练过程中得以同时优化图像的结构保真性、风格一致性与细节完整性。各类损失函数从不同层面对生成图像施加约束,形成多尺度、多目标的优化机制,从而确保迁移结果在保持文物图像基本语义结构的同时,充分还原拓片特有的墨色质感与视觉风格特征。

### 3 系统实现与流程

系统的实现以及流程涉及到数字拓片生成新方法的开发和落实,涵盖了数据处理、模型构建以及模型应用 3 个层面。它将现有的各类数据资源加以融合,运用 Facebook 的 SAM 模型针对文物主体来开展高效的图像分割工作,并且还集成了颇具创新性的风格迁移网络,从而达成全流程自动化以及高效运作的效果。

#### 3.1 数据处理层

在数据处理这一层面,首要需要收集高质量的文物图像数据并且做好标注。接着,要针对图像展开预处理工作,包括像图像增强、去除噪声、裁剪以及实现标准化等,其目的在于保证数据具备一致性以及清晰度。为了能够让数据的多样性得以增强,采用诸如旋转、翻转这类的数据扩增手段,进而构建出一个尺度上较为统一、类别又很丰富的高质量数据集。

本文选取 MS-COCO 作为内容图像的数据集,将 WikiArt 作为风格图像的数据集来开展网络训练。这两个数据集各自涵盖了大概 80 000 张训练图像。其中,MS-COCO 属于能够用于物体检测、分割以及图像

描述的大规模数据集, 而 WikiArt 则是一个把重点放在艺术风格识别以及风格迁移方面的艺术图像数据集。除此之外, 还引入了重庆大学大数据与软件学院实验室内部构建的小规模文物数据集, 这里面包含有 170 张正射图像以及 187 张拓片图像。借助数据增强技术, 比如旋转、翻转、缩放及随机裁剪操作把正射图像的数量扩充到了 12 240 张, 将拓片图像也扩充至 13 464 张, 其目的在于增强文物图像在迁移至拓片风格时候的实际效果及模型的泛化能力。

### 3.2 模型构建层

在文物数字拓片生成中, 模型构建属于核心步骤。采用图像分割与风格迁移技术来开展模型建模, 能够切实有效地实现拓片风格的转换, 并且可以处理不同种类的文物图像。图像分割技术首先会把文物图像里的主要特征和背景分离开来, 把细节部分精准地提取出来; 接着, 风格迁移技术会在保留文物原有特征的基础之上, 把拓片风格运用到图像之中, 最终实现将实体文物的正射图像转变成为数字拓片。

在训练的整个过程里, 采用的是 Adam 优化器, 把学习率设定成 0.000 1。所有图像当中相对较小的那个维度会被缩放至 512 像素, 在此过程中要保持图像的宽高比, 并且还要随机裁剪出  $256 \times 256$  像素的区域。由于网络是完全卷积的, 测试阶段能够处理任意尺寸的输入。为了进一步对模型进行优化, 采用了迁移学习的方式, 通过对预训练的模型权重做微调处理, 与此同时, 应用 L2 正则化以及 Dropout 技术来防止出现过拟合的现象, 而且还要定期运用验证集去评估模型的性能。

### 3.3 模型应用层

为达成自动化以及智能化的拓片生成目标, 专门开发出了数字拓片生成系统(图 5)。在使用该系统时, 用户仅仅需要上传文物图片, 并且从中挑选出要进行转换的部分, 随后系统便能在短短几秒之内完成其内部模型的相关处理操作, 进而自动生成与之相对应的拓片图像。此系统对于提升文物数字化管理水平以及保护效率方面有一定的促进作用, 同时在推动文物数字化展示上也能够发挥积极作用。

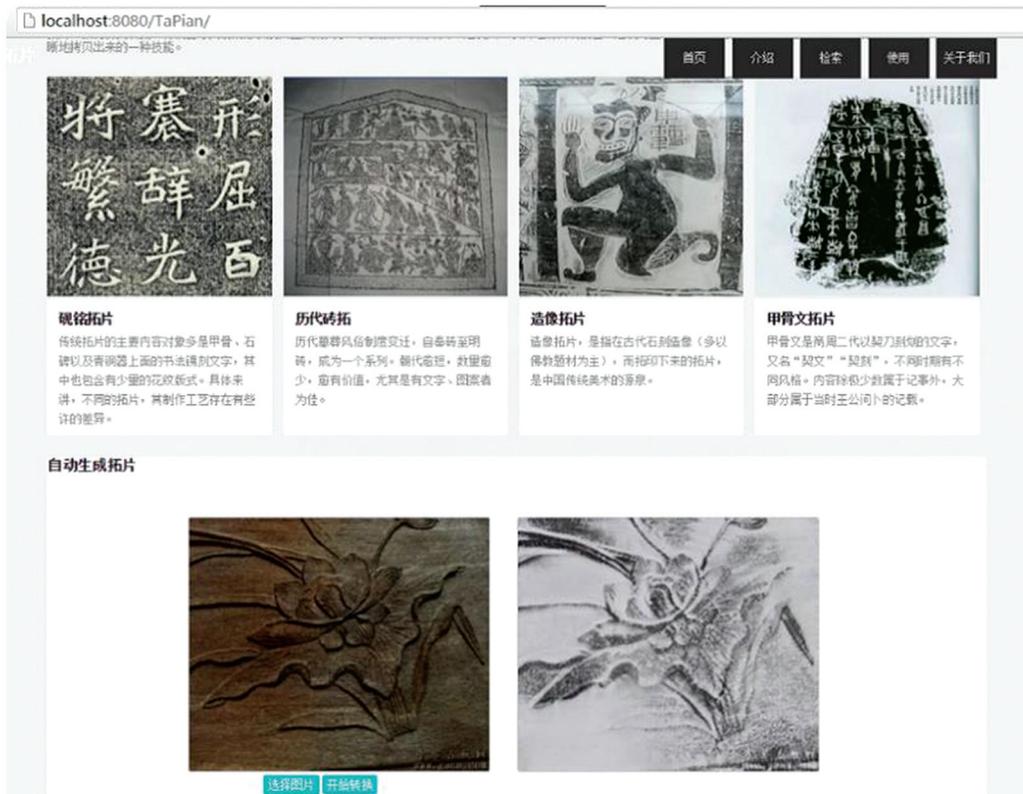


图 5 数字拓片生成系统截图

## 4 实验与分析

在此次研究当中,把自行构建的拓片图像以及文物正射图像相互结合起来,以此完成对模型的训练操作。当训练环节结束之后,只要输入文物图像,就能自动生成和该文物相对应的数字拓片图像,实验结果如图 6 所示。

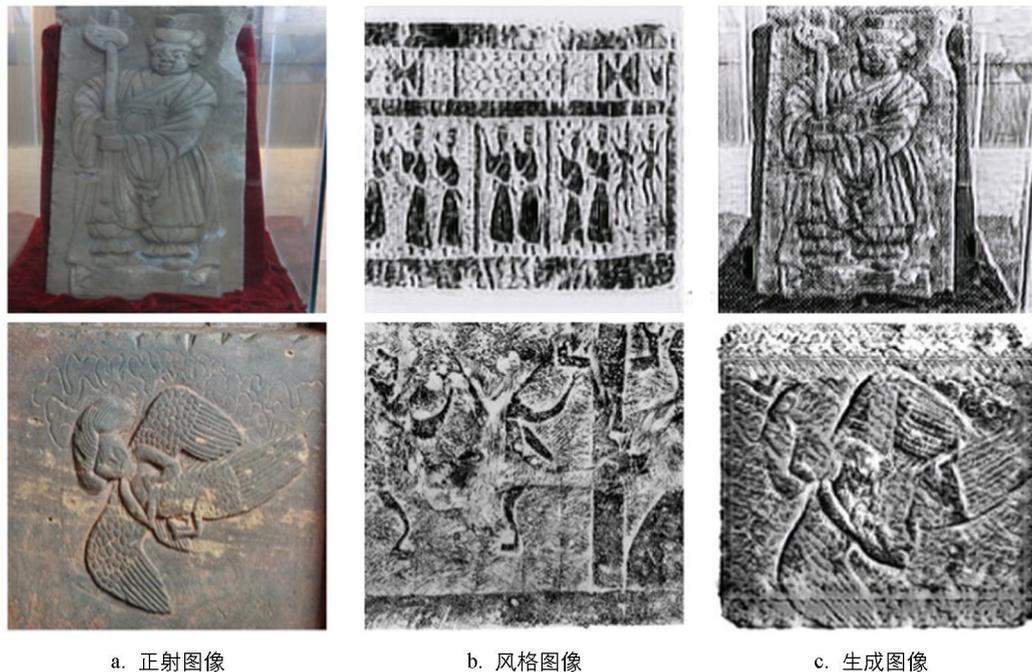


图 6 实验结果

将文物图像输入之后,模型可以依照用户所选定的区域,生成文物相对应的数字拓片图像(也就是图 5 所展示的那样)。模型一方面实现了拓片风格化,另一方面对文物的纹理细节以及独特特征尽量予以保留。研究选取涵盖不同年代、材质的文物图像样本进行测试,训练集与验证集按 8:2 划分,文物正射图像验证集共 2 448 张,拓片图像验证集共 2 692 张,从中随机抽取 1 000 张组成 10 组进行验证,每组验证集包含 100 张文物正射图像和 100 张拓片图像。选用图像结构相似性指数(SSIM)衡量模型对原始内容细节的保留程度,使用峰值信噪比(PSNR)衡量拓片风格迁移后图像的清晰度和整体质量。验证集评估结果如表 1 所示。实验结果表明:生成图像的 SSIM 指标平均能够达到 0.75,最高达到 0.82;PSNR 平均值为 28.1 dB,其中最高值为 29.3 dB,最低值为 27.8 dB。采用的 VGG-19 网络在特征提取上表现出色,能与对称解码器结合捕捉文物局部细节与整体结构。在处理不同年代且不同材质的文物图像时,本模型展现出了较好的适应性,生成的数字拓片可以十分细腻地呈现文物的微小纹理以及刻痕。

表 1 验证集评估结果

指标	1 组	2 组	3 组	4 组	5 组	6 组	7 组	8 组	9 组	10 组
SSIM	0.74	0.76	0.73	0.77	0.75	0.78	0.72	0.82	0.74	0.75
PSNR	27.8	28.2	27.9	28.3	28.1	28.4	27.8	29.3	28.0	28.1

拓片风格注意力网络在此次研究里起到了极为关键的作用。拓片风格注意力网络会对不同拓片风格特征的权重进行动态调整,如此一来,其生成的数字拓片既能清晰展现文物的细致纹理,又能够契合传统拓片所具有的艺术风格,促使文物内容和拓片风格实现了高度融合,所生成的数字拓片从视觉效果上去看,和传统手工拓片极为相近,进而提升了数字拓片的表现力以及真实感。

相对传统的图像处理技术,本方法在自动化程度以及处理效率这两方面彰显出颇为显著的优势。传统的处理方法在很大程度上要依赖人工操作,并且对操作人员的技能要求也比较高,耗时较长,而且在面对大批量生产的情形时,想要实现处理结果的一致性是极为困难的。本文所提出的模型是依托深度学习技术构建起来的,达成了高度自动化的处理流程,只需要用户去选定文物主体区域,就能迅速生成质量颇高的拓片图像。这在很大程度上减少了人工干预的程度,进而使得拓片生成的效率以及一致性都得到了有效的提升。

与依靠三维重构的方式相比,本文所提方案在对计算资源的适配以及应用场景的契合方面,同样表现出了明显优势。虽说三维重构方法在精度层面取得了一定的突破,然而其往往需要借助极为昂贵的设备,并且要历经复杂繁琐的计算流程,这在很大程度上限制了大规模应用。本文采用的方法不但能够有效降低对设备的依赖程度,而且还可以高效地重现文物所具备的细节特征以及拓片风格特征,为开展文物数字化保护工作给出了一种更为经济且高效的解决办法。

## 5 结论

将图像分割技术和风格迁移技术融合起来,为文物数字拓片自动生成提供了一套全新的解决方案。运用 SAM 进行图像分割,再与深度学习风格迁移网络结合,成功地将文物内容与拓片风格有效地融合在一起,生成了接近手工拓片风格的文物数字拓片图像。这一技术一方面提升了文物数字拓片处理时的效率以及效果,另一方面也为多样化的文化表达以及传播形式创造了可能性。凭借实现文物信息的数字化以及可视化,推动了文化遗产的保护与利用,也为学术研究、公众教育以及文化交流等诸多方面的应用开拓了新的空间。

虽然本研究的方法在大多数情况下表现优异,但仍存在一些局限。

1) 关于细节丢失问题。处理纹理较为复杂或者表面损坏程度较为严重的文物图像,模型所生成的拓片图像没有办法把所有的细节都完整地还原出来,尤其是针对深浮雕文物图像,呈现出一定的局限性,出现细节丢失。未来的研究可以通过增强模型的特征提取能力和优化算法来进一步提升细节保留效果,以更好地满足文化遗产保护的高标准需求。

2) 拓片风格泛化问题。针对非典型或风格极为独特的拓片样式时,模型的 PSNR 评分会出现下降。具体而言,在处理复杂拓片风格时,生成图像的细节边缘和视觉效果与原始拓片相比存在较为明显的偏差,未能精准匹配原始拓片的独特风格特征。后续开展的研究工作,将引入更多拓片风格数据集开展训练,让模型学习更广泛的拓片风格特征以提升其泛化能力。

## 参考文献:

- [1] 张文艳. 墓志拓片文物保护路径 [J]. 文物鉴定与鉴赏, 2024(18): 32-35.
- [2] 朱晓丽, 张春新, 吴中福. “数字拓片”构想及实验研究 [J]. 同济大学学报(社会科学版), 2011, 22(3): 82-88.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [4] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFS [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [5] KIRILLOV A, MINTUN E, RAVI N, et al. Segment Anything [EB/OL]. (2023-04-05) [2024-11-26]. <https://arxiv.org/abs/2304.02643v1>.

- [6] GATYS L A, ECKER A S, BETHGE M. Image Style Transfer Using Convolutional Neural Networks [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 2414-2423.
- [7] HUANG X, BELONGIE S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization [C] //2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 1510-1519.
- [8] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-Attention Generative Adversarial Networks [C] //Proceedings of the 36th International Conference on Machine Learning (ICML). Long Beach: PMLR, 2019: 7354-7363.
- [9] 王书敏. 基于纹理特征方法的甲骨拓片文字定位研究 [J]. 信息系统工程, 2020(12): 141-142.
- [10] 信泰琦, 席艳峰, 李子豪. 基于多尺度积分不变量的碑刻数字拓片制作研究 [J]. 工程勘察, 2024, 52(8): 76-81.
- [11] 刘清珺, 刘晓晖, 刘海伦, 等. 探索自动化数字拓印技术对刻文的保护及传承 [M]. 北京联合大学, 北京数字科普协会. 互联网时代的数字博物馆. 北京: 电子工业出版社, 2018: 104-110.
- [12] 朱晓丽, 吴中福. 基于 photoshop 实现浮雕图像“数字拓片”的技术研究 [J]. 计算机科学, 2008, 35(12): 224-228.
- [13] 朱晓丽, 吴中福, 李颖. 基于 MATLAB 实现石刻浮雕图像“数字拓片”技术的研究 [J]. 计算机科学, 2009, 36(2): 268-270.
- [14] 张园林. 基于三维模型的碑刻数字拓片生成技术及应用 [D]. 杭州: 浙江大学, 2018.
- [15] 胡春梅, 薛惠敏, 夏国芳, 等. 基于激光点云和近景影像的数字拓片生成方法. 激光杂志, 2017, 38(9): 32-37.
- [16] 聂凡. 墓志数字拓片制作与虚拟展示—以周懿王及夫人墓志为例 [J]. 大众考古, 2023, (7): 48-53.
- [17] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J] [EB/OL]. (2015-08-10) [2024-08-11]. <https://arxiv.org/abs/1409.1556>.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale [EB/OL]. (2021-05-03) [2024-08-11]. <https://arxiv.org/pdf/2010.11929>.
- [19] ZHANG Y C, SHEN Z R, JIAO R S. Segment anything Model for Medical Image Segmentation: Current Applications and Future Directions [J]. Computers in Biology and Medicine, 2024, 171: 108238.
- [20] PARK D Y, LEE K H. Arbitrary Style Transfer with Style-Attentional Networks [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 5873-5881.

责任编辑 张枸

崔玉洁

