

环境语义数据增强方法在古墓葬检测中的应用

陈思航^{1,2}, 于丽君^{1,2}, 朱建峰^{1,2}, 陈婕^{1,2}, 柳泽³, 王辉⁴, 聂跃平^{1,2}

1. 中国科学院空天信息创新研究院, 北京 100101

2. 中国科学院大学, 北京 101408

3. 自然资源部国土空间规划研究中心, 100034

4. 复旦大学, 200433

摘要: 在深度学习领域, 目标检测模型的性能通常依赖于样本数据充足且高质量的标注数据集。然而, 在实际的遥感考古探测中, 大规模、多样化的数据集获取困难且成本高昂, 导致样本数据不足, 模型在训练时容易过拟合。尤其在墓葬赋存环境差异较大的情况下, 不同环境之间的样本数量的差异, 会导致样本分布不均衡, 从而限制模型的环境适应能力。为了解决这些问题, 本文提出了一种基于扩散模型的遥感影像数据增强算法。通过设置不同的环境提示词, 在不破坏原始数据集标签分布的前提下, 丰富图像的视觉语义, 缓解数据样本稀缺和样本不均衡的问题。在阿勒泰地区的高分辨率墓葬影像数据集中, 本文评估了该算法在不同目标检测模型中的性能, 并横向比较了其与 Mosica、Mixup 等多种数据增强算法的效果。实验结果表明, 本文提出的算法在测试集上的平均精度比基线方法提高了 12.2%, 在异源数据集上的平均精度提升了 16.4%, 显著提高了模型的检测准确性、稳定性及跨数据集的适应性, 为小样本遥感考古目标识别提供了有效的技术支撑, 拓展了智能考古检测的研究思路。

关键词: 深度学习; 墓葬检测; 数据增强; 扩散模型; 遥感考古; 小样本学习; 环境语义增强; 模型泛化; 图像生成

中图分类号: **文献标识码:** **文章编号:**

引用格式:

收稿日期: 修回日期:

基金项目: 国家重点研发计划(2020YFC1521900), 国家自然科学基金项目(41801134)

第一作者简介: 陈思航, 1999 年生, 男, 硕士研究生, 研究方向为人文考古、遥感图像处理等。E-mail: chensihang22@mails.ucas.ac.cn

通信作者: 于丽君, 1982 年生, 女, 研究员, 研究方向为遥感考古。E-mail: yulj@aircas.ac.cn

1 引言

作为考古研究的重要对象,墓葬的历史渊源极为深厚,通过研究分析墓葬形制与空间格局的变迁,可以从侧面论证该地区的文化属性与变迁(苏雨露,2024),在考古研究中具有重要的意义。

随着遥感技术的蓬勃发展,考古学也迎来了新一轮的技术革新。相较于传统考古方式,遥感技术可以快速获取遗迹的光谱、纹理与几何特征(于丽君,2018),大大节省了人力成本,提高了田野考古的效率。深度学习在海量数据处理和特征自动提取方面具有巨大优势,可在大范围地区的田野考古调查中发挥重要作用。目前基于深度学习的墓葬的目标检测已经取得了实质性的进展,例如,Caspari等基于谷歌地球构建了欧亚草原上的小尺度墓葬数据集,利用CNN模型对遥感影像上的墓葬进行预测,精度远高于传统机器学习方法(Caspari等,2019);Verschoof-van等提出了基于Faster R-CNN的考古遗迹自动检测流程,并在LIDAR数据上成功实现了对圆形墓葬的提取分类(Verschoof-van等,2019);Chen等基于阿尔泰山脉的GF-2卫星影像利用Cascade R-CNN架构的物体检测模型足探测到了墓葬的空间分布(Chen等,2021);Berganzo-Besga等结合深度学习和随机森林算法,使用哨兵数据成功检测到加利比亚地区不同环境下的墓葬(Berganzo-Besga等,2021);Soroush等基于U-Net模型对伊拉克地区CORONA影像上的坎儿井进行提取,也取得了良好的成果(Soroush等,2022)。

虽然古代墓葬遗迹数量众多,但相比于深度学习

方法对样本量的需求,往往会面临数据样本量缺乏的情况,导致模型过于关注样本集内容,容易产生过拟合现象。针对少量样本数据,对其进行增强或增广是一套行之有效的解决方案,其本质是在辅助数据或转换规则的帮助下,对原始数据分布进行扩展,从而获取结构更复杂或数量更多的数据。图像混合增强是提升样本数据数量与质量的重要手段(Yuan Chih等,2023)。经典的增强算法如CutMix(Sangdoo Yun等,2019)、Mixup(Zhang等,2018)采用线性插值运算对图像进行混合增强;AugMix(Hendrycks等,2020)、IPMix(Huang等,2021)采用随机组合的方式模拟色彩结构语义变化;PixMix(Hendrycks等,2021)引入了复杂分形图像对原图像进行增强,大大提升了数据的安全系数;DiffuseMix(Islam等,2024)使用预先定义的条件提示库,将原始图像和扩散模型生成的图像进行混合。这些图像增强算法在许多数据集上都能带来质量的提升,具有较高的实用价值,但目前国内外将图像增强技术应用在遥感考古目标检测中的相关研究较少(Yin cui等,2021),考古目标检测的小样本问题仍十分突出。

不同地区、不同历史年代、不同聚落的墓葬往往在遥感影像上表现出较大的差异,且容易与非墓葬目标混杂,大部分需基于实地勘探才能确定遥感解译的墓葬是否准确。尤其在高寒区、荒漠地带、植被覆盖区等特殊区域和大范围调查区域,实地调查工作需花费大量的人力物力。这就导致了大型墓葬数据集构建成本较高。此外,复杂的赋存环境会为墓葬目标的识

别工作带来不可忽视的影响。不同环境之间的样本数量的差异，会导致模型的训练结果出现偏向性，对某些样本数据相对较多的背景环境具有良好的检测能力，样本数量较少的环境表现欠佳。现有的大多数图像增强技术无法模拟不同环境下的样本图像，使得样本数据分布不均衡，影响模型的精度。

为应对上述问题，本文构建了一套高分辨率墓葬影像数据集，分别测试了多种主流目标识别模型的效果。在此基础上，本文提出了一种环境语义增广算法，用于提高少样本墓葬目标的识别精度，提升模型的泛化能力，为少样本遥感考古目标检测提供了新思路。

2 研究区概况

阿勒泰地区位于新疆维吾尔自治区北部，地区内地形地貌复杂多样，北临阿勒泰山，南接准葛尔盆地，西南与沙吾尔山为邻，地势东北高西南低，主要地貌为山地、草原和沙漠（苏金娟 等，2023；卢梦等，2024）。

阿勒泰地区是早期东西方文化交流的重要通道，墓葬数量众多，例如切木尔切克乡克尔木齐古墓群、富蕴县塔勒德萨依墓地、阿勒泰市克孜加尔墓地（曾宝栋，2017）数量、规模和考古学特征都具有代表性。

阿勒泰地区墓葬目标在遥感影像上表现分为两大类：一类是斑点型，另一类是圆环型，两类的边界都比较模糊，自动提取这些墓葬目标比较困难（KIM 等，2013）。不同历史时期的墓葬对象可能分布在差异较大的环境中，例如沙尔依吾依希克古墓群分散在空旷的裸地，阿克托别墓地分布在农田、低矮灌木覆盖的

环境下，克亚乌特开勒古墓群处于居民区中，阿克布拉克墓群位于终年覆雪的雪山山脚下，如图2.1所示，这些墓葬所处区域展现出鲜明的环境特征，在遥感图像上的结构也表现出一定的差异性。

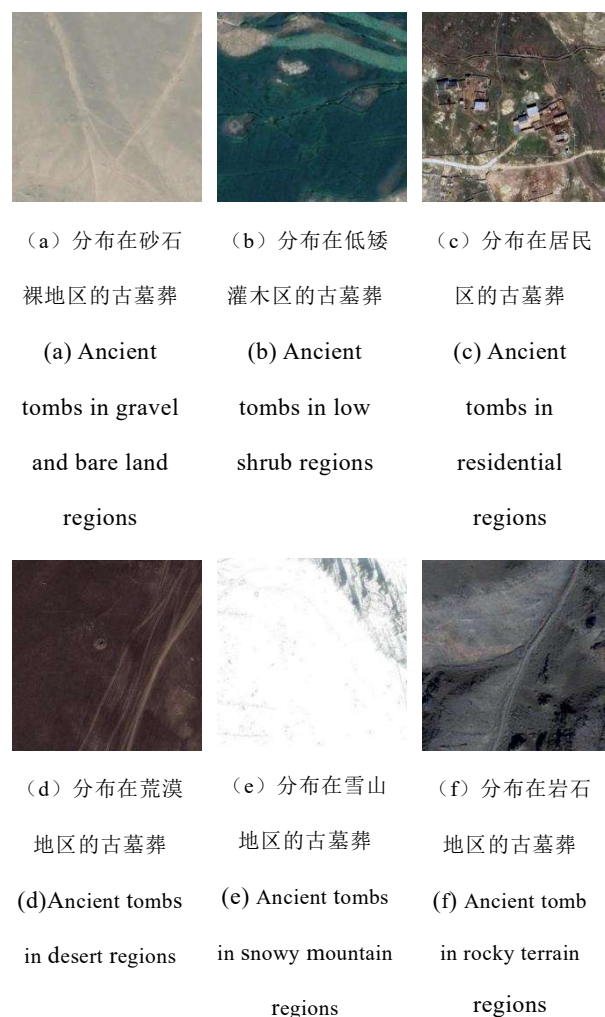


图 2.1 阿勒泰地区墓葬的赋存环境图

Figure 2.1 Surrounding Environment of Ancient Tombs in the Altai Region

目前在阿勒泰地区仍存在大量未发现的墓葬（邵会秋，2008），有待开展系统性的遥感考古探测工作。而由于样本数据的数量和质量存在很大的局限，为墓葬自动检测工作带来了一定的困难与挑战。因此本文通过构建一套数据增广算法，降低环境变化的影响，避免样本环境不均衡问题，增加数据集的鲁棒性和环

境复杂度,可有效增强深度学习模型在不同环境下的
识别能力。本文采用选择的样本点分布如图2.2所示。

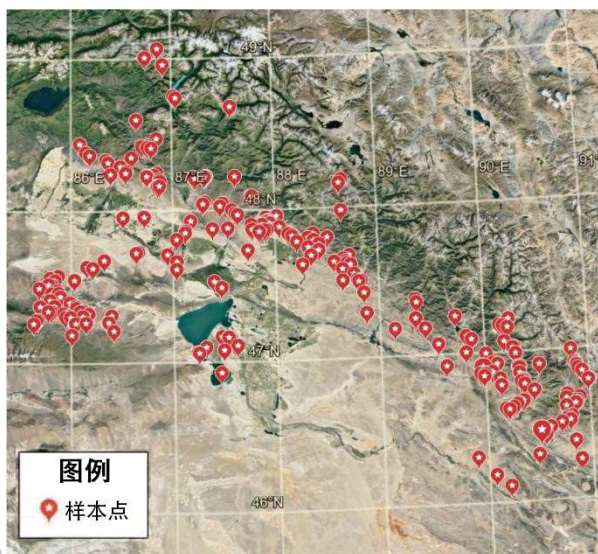


图 2.2 阿勒泰地区部分样本点的分布情况

Figure 2.2 Distribution of Selected Sample Points in the
Altai Region on Google Earth

3 研究方法

3.1 数据集构建

本文共收集了新疆阿勒泰地区523个墓葬的精确
坐标点,并从谷歌地球上获取了对应的卫星影像图。
这些影像的空间分辨率从0.3-1m不等,成像时间为
2019-2024年。通过目视解译剔除了难以辨识、目标边
缘极其模糊、质量较差的影像,一共得到了298张有效
的墓葬影像,影像为RGB三通道。每张影像上的墓
葬数量为1到17个不等,共计938个墓葬。将数据按照
6:2:2划分为训练集、验证集和测试集,每张影像的尺
寸均为640×640(像素),影像上的墓葬样本数量分
布如图3.1所示。

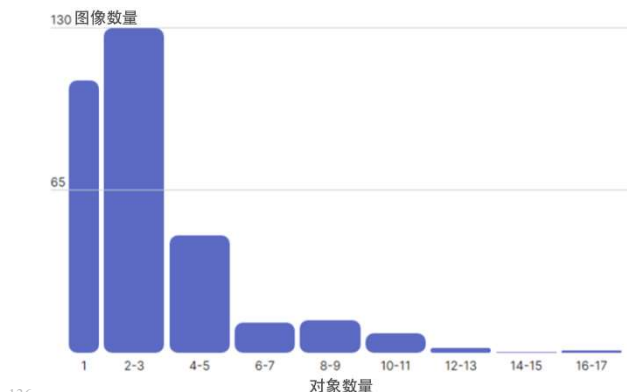


图 3.1 影像样本数量分布

Figure 3.1 Distribution of Image Sample Quantities

3.2 目标识别网络

目前,深度学习在目标识别领域的主要流派可分
为一阶段检测网络和二阶段检测网络。一阶段检测网
络如YOLO、SSD算法,其较高的精度和检测速度深得
工业应用的青睐,二阶段网络如Faster R-CNN,
Cascade R-CNN,尽管具有较高的精度,但其分块的架
构设降低了推理速度。

结合大范围田野考古调查高效率、高精度的应用
需求,本文选用推理速度更快的一阶段网络进行试验,
测试了YOLO系列的经典模型YOLO v5 (Zhu 等,
2021), YOLO v8 (Wang 等, 2023), YOLO 系列
的最新模型YOLO v9 (Wang 等, 2024), YOLO v10
(Wang 等, 2024)以及基于Transformer架构的目标检
测网络RT-DETR (Zhao 等, 2023)。

3.3 基于环境增强的遥感影像增广算法

3.3.1 算法背景

尽管DiffuseMix在各项分类任务中都取得了不错
的成绩,但不可忽视的是,在识别任务中,可能会因
为小尺度对象的空间分布而丢失视觉语义信息,从而

退化成原始图像与分形图像的结合。且DiffuseMix算法中采用的提示词并不适用于环境增强，无法模拟真实场景下的样本状况。

此外，由于谷歌地球提供的影像存在采集时相不一致、覆盖地域广泛等情况，导致数据集表现出较大的地理空间环境变异性。针对该问题，本文提出了一套适用于复杂环境背景的遥感影像数据增广算法。其架构如下图所示：

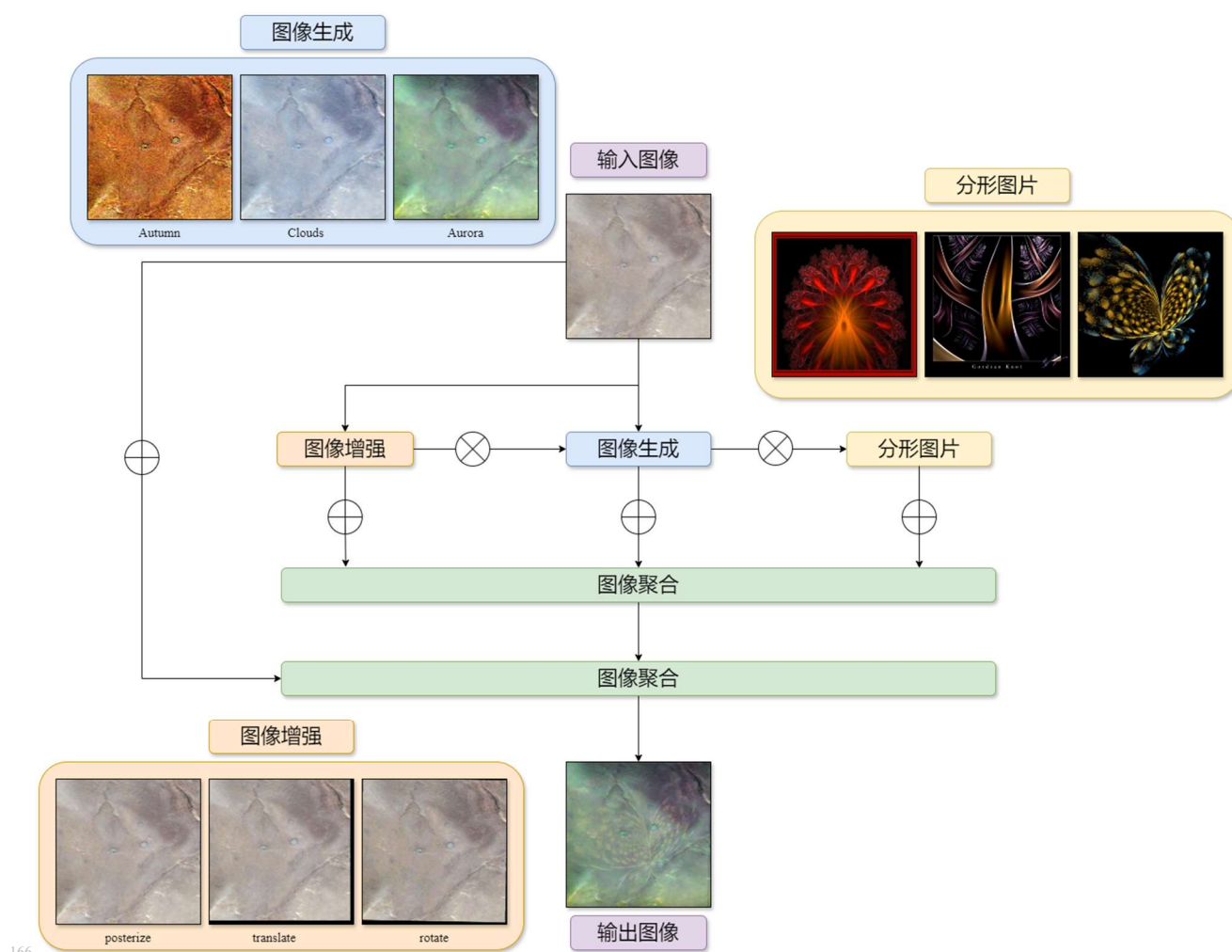


图 3.2 算法架构图

Figure 3.2 Algorithm Architecture Diagram

3.3.2 算法结构

基于增强图像视觉语义、保留主干信息的思想，我们设计了一套基于扩散模型的图像增强结构，通过引入短路连接的方式，保证了原始语义信息不会大幅丢失，并进一步扩展了图像语义空间。其主要结构包括合成图片生成、分形图形融合两部分。

扩散模型作为近年来最火爆的生成式模型，其各种变体在各种图像生成、视频生成、三维重建、图像编辑领域均取得了举世瞩目的成绩。在图像增强方面，采用微调扩散模型生成的合成数据可以提升分类性能，创建更多样、语义更加丰富的数据集，相较于对抗生成网络（GAN），扩散模型不需要依赖

大量真实影像 (Shekoofeh Azizi 等, 2023; Brandon Trabucco等, 2023; Zheng Li 等, 2023)。在本实验中, 考虑到考古目标的背景环境差异较大, 我们选择预训练的InstructPix2Pix (Brooks T 等, 2022) 扩散模型作为图像生成模型, 并参照ESRI的Sentinel-2卫星合成产品提供的土地利用类型、主要环境、气候来设置提示词 (Prompt), 该合成产品提供了九种主要的土地覆盖类型, 可以涵盖大部分基础地貌。基于扩散模型的图像生成工作流程如图 3.3 所示, InstructPix2Pix模型是通过图文对迁移构建的图生图指令微调模型, 输入的提示词经过预训练的文本编码器编码成一组嵌入词向量, 用于控制模型的输出。

输入的图片经过变分自编码器映射为一组高维特征, 通过UNet和交叉注意力层实现与词向量特征的多模态融合, 经过降噪过程逐步学习到词向量的风格, 并通过一个变分自编码器解码器生成最终的合成图像。图3.4展示了不同的环境提示词为遥感图像带来的不同视觉特征, 模拟目标在相应环境下可能的分布形态。经过实验发现, 基于我们设置的环境提示生成的图像更倾向于对背景、全局进行处理, 不会对识别目标结构进行大幅度的修改, 从而在维持结构语义的同时延展色彩语义。这些提示词可以根据具体任务进行调整, 以便适应不同环境不同要求下的数据增强工作。

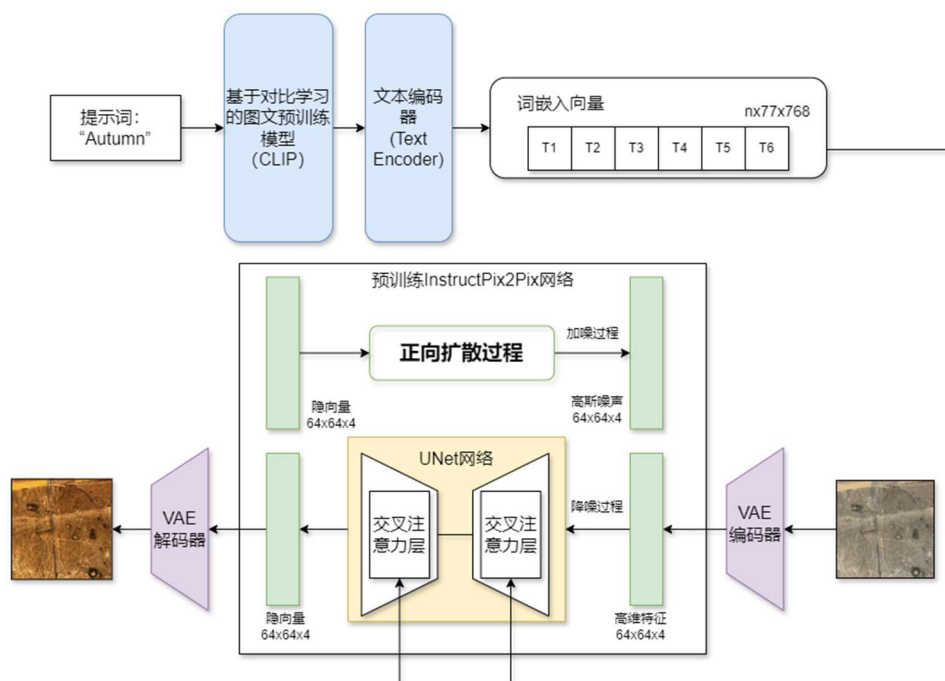


图 3.3 基于 Prompt 的图生图工作

Figure 3.3 Prompt-Based Image-to-Image Processing

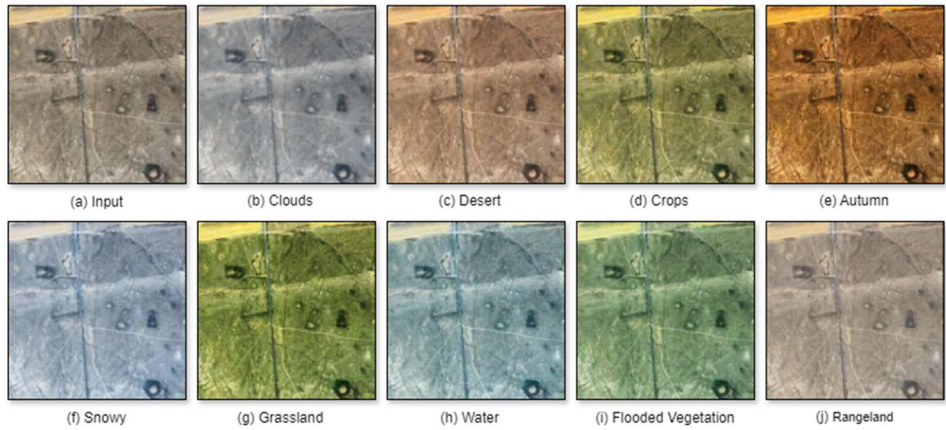


图 3.4 不同 Prompt 生成的合成图片

Figure 3.4 Composite Images Generated by Different Prompt

为了增强图像的安全性能，我们沿用PixMix和IPMix利用复杂分形图形提高图像数据安全性能的思想，将分形图形通过像素运算置入原始图像中，增强数据集的鲁棒性。这些分形图形筛选自DeviantArt艺术网站上由艺术家绘制的高度非偶然图形（Huang等，2021），其流程大致如图3.5所示。目前而言，尚未有相关文献将分形图像与遥感影像直接融合的案例，但这些分形数据集在多种经典图像识别数据集如COCO、ImageNet中被证实可以很好地提升数据集安全性，并且增加图像结构的复杂程度，避免模型过

拟合。与DiffuseMix的直接相加有所不同，我们更希望分形图形能够以背景纹理的方式存在，这是由于墓葬在影像上以稀疏小样本形式分布，原先的相加方式容易遮盖部分重要信息，且更像是在对无关背景做增强。在尝试了多种图像融合方式之后，我们发现基于随机贝塔分布的图像乘法效果较好，其对影像的增幅是基于原始影像像素的，视觉效果更近似背景纹理，减少对主要信息破坏的同时，增强了图像的结构语义。

优化得到的最佳权重，适用于原始图像与增强图像融合
 的通用任务。 I 表示图片向量 $[I_{il}^A, I_{ij}^G, I_{ik}^K]$ ，其中 I_{il}^A 表示
 原始图片经过简单增强后的结果， $I_{il}^A = A(I_i, o_l, k)$ 。 I_{ij}^G
 表示原始图片通过指令生成的合成图片， $I_{ij}^G =$
 $G(I_i, p_j) \otimes I_{ik}^A$ 。 I_{ik}^K 表示从集合中随机抽取的分形图片
 增强结果， $I_{ik}^K = f_k \otimes I_{ij}^G$ 。

3.3.4 算法时间复杂度分析

本文算法可分为图像融合、图像生成和图像增强
 三个模块，本小节将剖析各个模块的时间复杂度，为
 算法的优化和实际应用提供理论基础。

设图像的总像素量为 N 。在图像增强方面，我们对
 图像进行简单的增强操作，每一次操作均会对所有像
 素造成影响，涉及到不同的操作时间复杂度也不同。
 在本实验使用的增强操作池中，所有的操作复杂度均
 为线性，则记图像增强的复杂度为 $O_{aug} \sim O(N)$ 。同时，
 对图像进行融合的时间复杂度也是线性的，记作
 $O_{mix} \sim O(N)$ 。

图像生成的计算瓶颈取决于扩散模型的复杂度，
 一般来说，假设每个时间步的计算复杂度为 $C(N)$ ，其
 取决于网络模块的结构与深度，考虑到模型使用了注
 意力机制，其计算复杂度近似于 $O_{gen} \sim O(T * N^2)$ 。其
 中 T 表示推理时使用的时间步的数量。

综上所述，本文算法的时间复杂度近似为 $O =$
 $\sum_i^{\{aug, gen, mix\}} O_i \sim O(T * N^2)$ ，其中 T 表示推理时使用的

时间步的数量。其中 N 表示图像的像素总数。

为了更加直观揭示本文算法与传统算法在计算效
 率上的差异，我们以生成2000张合成图片为例，进行
 相关的时间开销计算。图片的宽高均为640像素，通道
 数为3。实验使用的图形处理器型号为RTX 4060，核心
 处理器型号为Intel(R) Core(TM) i7-14700F。

表3.1 不同算法的时间开销

Table 3.1 Time consumption of different algorithm

算法	总时间/s	平均时间/ms
本文算法	3,512.4	1756.2
Diffuse Mix	3,467.2	1733.6
Mosaic	29.4	14.7
MixUp	11.2	5.6

表3.1为计算结果，可以发现，扩散模型的推理开
 销相较于传统数据增强算法较大，故一般先进行图像
 推理工作，重构数据集后再进行目标识别模型的训练。
 训练过程的时间开销与图片数量和网络结构有关，并
 不需要额外的计算。

4 实验结果与分析

4.1 数据增广结果

我们使用本文提出的算法将训练集全部的179张
 影像进行增广，环境提示词设置如表4.1所示。

表4.1 预设环境提示词

Table 4.1 Preset Environmental Prompts

提示词 (Prompt)	解释	类型
River	模拟河流环境特征点	环境提示词
Built Area	模拟居民区、建成区域环境特征	环境提示词
Clouds	模拟带有云量的遥感影像特征	环境提示词
Crops	模拟作物区域影像特征	环境提示词
Desert	模拟沙漠区域影像特征	环境提示词
Flooded Vegetation	模拟湿地区域影像特征	环境提示词
Forest	模拟森林区域影像特征	环境提示词
Grassland	模拟草地区域影像特征	环境提示词
Rangeland	模拟牧场区域影像特征	环境提示词
Snowy/Ice	模拟冰雪区域影像特征	环境提示词
Water	模拟水体区域影像特征	环境提示词
Autumn	模拟秋季影像特征	季候提示词

在本实验中，使用扩散模型生成每一种环境提示词下的影像179张，并随机抽取分形图形进行增强，共计生成增强图片2148张，切片如图4.1所示。



图 4.1 增广图像集切片

Figure 4.1 Slices of Augmented Image Dataset

4.2 模型训练结果

为了分析不同数据增强算法在墓碑图像识别上的效果，以及比较主流的目标识别模型在遥感墓碑识别上的性能，本实验分别选取了YOLO v5、YOLO v8、YOLO v9、YOLO v10架构的小参数、中参数模型和RT-DETR的大参数、超大参数模型进行训练，其中各个模型参数量、FLOPS及在COCO数据集上的精度如下表所示：

表4.2模型基本数据

Table 4.2 Basic Model Parameters

模型	参数量 (M)	FLOPS (G)	mAP ₅₀₋₉₅
YOLO v5n	2.6	7.7	34.3
YOLO v5m	25.1	64.2	49.0
YOLO v8n	3.2	8.7	37.3
YOLO v8m	25.9	78.9	50.2
YOLO v9t	2.0	7.7	38.3
YOLO v9m	20.1	76.8	51.4
YOLO v10n	2.3	6.7	38.5
YOLO v10m	15.4	59.1	51.1
RT-DETR L	42	136	53.1
RT-DETR X	76	259	54.3

为了方便比对，我们在相同的实验环境、相同的墓碑数据集下，对这些模型性能进行测试与分析。在本实验中，所有的模型均不使用预训练权重。其中一

些重要参数如下：优化器选择SGD，初始学习率设置为0.01，动量设置为0.9，Batch大小设置为8，迭代次数设置为300，早停次数设置为75，设置验证置信度阈值为0.5，IoU阈值为0.6，在本阶段的实验中，我们对所有的模型采取Mosaic (Bochkovskiy A 等, 2004) 数据增强策略，该策略自YOLO v4起便成为了YOLO系列的默认增强方案，其核心在于随机对样本图片选取、放缩，并拼接成一张图像。本实验以其为基线方法，探究初始情况下各个模型在数据集上的适应性。

4.3 实验结果

为了验证本文提出算法的有效性，实验选择了基线方法Mosaic、CVPR2024图像增广算法DiffuseMix、经典图像增广算法MixUp和本文图像增广算法进行实验比较，最终结果如表4.3所示。

实验结果表明，YOLO v8系列在测试集上的表现最优；RT-DETR L与RT-DETR X是基于Transformer结构的目标识别网络，在验证集和测试集中均取得了不错的表现。而YOLO v9与YOLO v10的性能表现并没有与其他模型拉开差距，进一步分析发现，大参数网络几乎都取得了比相同结构下小参数网络更好的性能，但其训练开销较大，且当样本数量较少时，大参数结构很可能发生过拟合。

表4.3 模型性能表现

Table 4.3 Model Performance

模型	增强方法	(T/V*) AP ₅₀	(T/V) AP ₅₀₋₉₅	(T/V) Precision ₅₀	(T/V) Recall ₅₀
YOLO v5n	Mosaic	0.676 / 0.509	0.289 / 0.218	<u>0.76</u> / <u>0.76</u>	0.60 / <u>0.60</u>
	DiffuseMix	0.700 / 0.621	<u>0.332</u> / 0.274	0.80 / 0.71	0.62 / 0.58

模型	增强方法	(T/V*) AP ₅₀	(T/V) AP ₅₀₋₉₅	(T/V) Precision ₅₀	(T/V) Recall ₅₀
	MixUp	<u>0.712</u> / 0.650	0.325 / <u>0.307</u>	0.72 / 0.72	<u>0.63</u> / 0.60
	Ours	0.721 / <u>0.646</u>	0.344 / 0.307	0.72 / 0.81	0.67 / 0.54
YOLO v5m	Mosaic	0.646 / 0.627	0.302 / 0.296	0.65 / 0.65	0.60 / 0.59
	DiffuseMix	0.698 / 0.710	0.331 / <u>0.326</u>	0.72 / 0.75	<u>0.66</u> / 0.66
	MixUp	<u>0.734</u> / 0.678	<u>0.359</u> / 0.310	0.77 / <u>0.74</u>	0.66 / <u>0.60</u>
	Ours	0.744 / <u>0.699</u>	0.368 / 0.357	<u>0.75</u> / 0.70	0.70 / 0.63
YOLO v8n	Mosaic	0.701 / 0.650	0.340 / 0.302	0.80 / <u>0.76</u>	0.59 / 0.56
	DiffuseMix	<u>0.721</u> / <u>0.686</u>	<u>0.366</u> / <u>0.354</u>	<u>0.79</u> / 0.73	0.64 / <u>0.63</u>
	MixUp	0.716 / 0.675	0.339 / 0.332	0.64 / 0.66	<u>0.71</u> / 0.65
	Ours	0.737 / 0.693	0.377 / 0.355	0.75 / 0.78	0.78 / 0.58
YOLO v8m	Mosaic	<u>0.731</u> / 0.636	0.343 / 0.318	<u>0.79</u> / 0.66	<u>0.63</u> / 0.64
	DiffuseMix	0.701 / 0.634	<u>0.355</u> / <u>0.330</u>	0.76 / 0.67	0.63 / 0.59
	MixUp	0.742 / 0.698	0.353 / 0.321	0.79 / <u>0.68</u>	0.66 / <u>0.61</u>
	Ours	0.702 / <u>0.676</u>	0.379 / 0.334	0.74 / 0.76	0.62 / 0.55
YOLO v9t	Mosaic	0.638 / 0.557	0.302 / 0.272	0.66 / <u>0.68</u>	0.56 / 0.47
	DiffuseMix	<u>0.717</u> / 0.692	<u>0.363</u> / 0.340	0.74 / 0.64	<u>0.64</u> / 0.66
	MixUp	0.717 / 0.685	0.369 / <u>0.335</u>	<u>0.74</u> / 0.66	0.61 / <u>0.63</u>
	Ours	0.706 / <u>0.686</u>	0.361 / 0.333	0.78 / 0.72	0.64 / 0.60
YOLO v9m	Mosaic	0.662 / 0.647	0.294 / 0.276	0.74 / <u>0.73</u>	0.58 / 0.53
	DiffuseMix	0.740 / <u>0.682</u>	0.357 / 0.352	0.80 / 0.69	<u>0.68</u> / <u>0.62</u>
	MixUp	<u>0.735</u> / 0.679	<u>0.362</u> / 0.333	<u>0.76</u> / 0.76	0.68 / 0.58
	Ours	0.731 / 0.701	0.376 / <u>0.350</u>	0.70 / 0.70	0.69 / 0.69
YOLO v10n	Mosaic	0.655 / 0.557	0.323 / 0.272	0.75 / 0.70	0.56 / 0.46
	DiffuseMix	<u>0.719</u> / <u>0.657</u>	<u>0.367</u> / <u>0.321</u>	0.81 / <u>0.73</u>	0.63 / <u>0.55</u>
	MixUp	0.693 / 0.611	0.323 / 0.284	0.61 / 0.64	0.69 / 0.51
	Ours	0.723 / 0.671	0.372 / 0.333	<u>0.76</u> / 0.76	0.67 / 0.64
YOLO v10m	Mosaic	0.708 / <u>0.638</u>	0.335 / 0.302	0.82 / 0.68	0.57 / <u>0.54</u>
	DiffuseMix	0.682 / 0.649	0.358 / 0.354	0.78 / 0.82	0.57 / 0.46
	MixUp	0.639 / 0.617	0.329 / 0.311	0.59 / 0.65	0.63 / 0.59
	Ours	<u>0.701</u> / 0.637	<u>0.355</u> / <u>0.340</u>	<u>0.77</u> / <u>0.73</u>	<u>0.61</u> / 0.52
RT-DETR L	Mosaic	0.695 / 0.671	0.304 / 0.296	0.87 / <u>0.78</u>	0.58 / 0.58

模型	增强方法	(T/V*) AP ₅₀	(T/V) AP ₅₀₋₉₅	(T/V) Precision ₅₀	(T/V) Recall ₅₀
RT-DETR X	DiffuseMix	0.717 / 0.735	0.307 / 0.303	0.80 / 0.80	<u>0.66</u> / 0.67
	MixUp	0.738 / 0.697	0.328 / 0.316	<u>0.84</u> / 0.72	0.72 / <u>0.63</u>
	Ours	<u>0.732</u> / <u>0.701</u>	<u>0.318</u> / <u>0.308</u>	0.79 / 0.77	0.64 / 0.60
	Mosaic	0.573 / 0.655	0.26 / 0.282	0.60 / <u>0.74</u>	0.60 / 0.59
	DiffuseMix	0.747 / <u>0.754</u>	0.315 / 0.306	0.69 / 0.70	0.73 / <u>0.76</u>
	MixUp	<u>0.773</u> / 0.754	0.345 / <u>0.340</u>	<u>0.74</u> / 0.74	0.75 / 0.72
	Ours	0.780 / 0.792	<u>0.342</u> / 0.342	0.76 / 0.72	<u>0.73</u> / 0.80
	Mosaic	0.678 / 0.615	0.312 / 0.283	<u>0.76</u> / 0.71	0.59 / 0.56
	DiffuseMix	0.714 / <u>0.682</u>	<u>0.345</u> / <u>0.326</u>	0.77 / <u>0.72</u>	0.65 / <u>0.62</u>
	MixUp	<u>0.720</u> / 0.674	0.343 / 0.319	0.72 / 0.70	<u>0.67</u> / 0.62
	Ours	0.728 / 0.690	0.359 / 0.337	0.75 / 0.75	0.68 / 0.62

* T 表示测试集，V 表示验证集，加粗数据表示同类比较最优结果，下划线数据表示同类比较次优结果

表4. 4 平均性能对比

Table 4.4 Average performance comparison

增强方法	(T/V*) AP ₅₀	(T/V) AP ₅₀₋₉₅	(T/V) Precision ₅₀	(T/V) Recall ₅₀
Mosaic	0.678 / 0.615	0.312 / 0.283	0.76 / 0.71	0.59 / 0.56
MixUp	0.720 / 0.674 (+6.2% / +9.6%)	0.343 / 0.319 (+11.3% / +12.7%)	0.72 / 0.70 (-5.3% / -1.4%)	0.67 / 0.62 (+13.6% / +10.7%)
DiffuseMix	0.714 / 0.682 (+5.3% / +10.9%)	0.345 / 0.326 (+10.6% / +15.2%)	0.77 / 0.72 (+1.3% / +1.4%)	0.65 / 0.62 (+10.2% / +10.7%)
Ours	0.728 / 0.690 (+7.4% / +12.2%)	0.359 / 0.337 (+15.1% / +19.1%)	0.75 / 0.75 (-1.3% / +5.6%)	0.68 / 0.62 (+15.3% / +10.7%)

* T 表示测试集，V 表示验证集，加粗数据表示同类比较最优结果，下划线数据表示同类比较次优结果

如表4.4所示，相较于基线处理方案，本文提出的相较于基线结构模型在验证集和测试集上差异较大，算法使得测试集中模型平均性能中的AP₅₀提升到了本文算法驱动模型在验证集和测试集上差异较小，0.728，提升率为7.4%，AP₅₀₋₉₅、精确率、召回率的提升性能更加稳定。DiffuseMix和MixUp两种增广方法均分别为15.1%、-1.4%、15.2%；而在验证集中AP₅₀、AP₅₀₋₉₅、精确率、召回率的提升分别为12.2%、19.1%、5.7%、略高于DiffuseMix方法，但在验证集上与后者存在一定差距。相较于这两种的图像增广方法，本文算法均

表现出较大的优势。在验证集与测试集上，本文算法-5.5%。在稳定性方面，由于后两种算法均使用了扩散模型图像生成技术，最终的结果相对稳定；而前两种算法过多的学习原始数据特征，泛化能力有所下降。

综合考虑四种增强方式下模型的平均性能，绘制出的模型性能图如图4.2所示。其中横轴表示 AP_{50-95} 的数值，纵轴表示 AP_{50} 的数值，上面的数字表示 AP_{50} 的具体值，下方的数字表示 AP_{50-95} 的具体值。其大小排序以更为严格的指标 AP_{50-95} 进行。从图中可以看出，本文提出的算法在验证集、测试集上的表现均优于DiffuseMix算法和MixUp算法，远高于基线（Mosaic）算法。验证了利用扩散模型模拟不同环境对遥感数据进行增强的可行性。

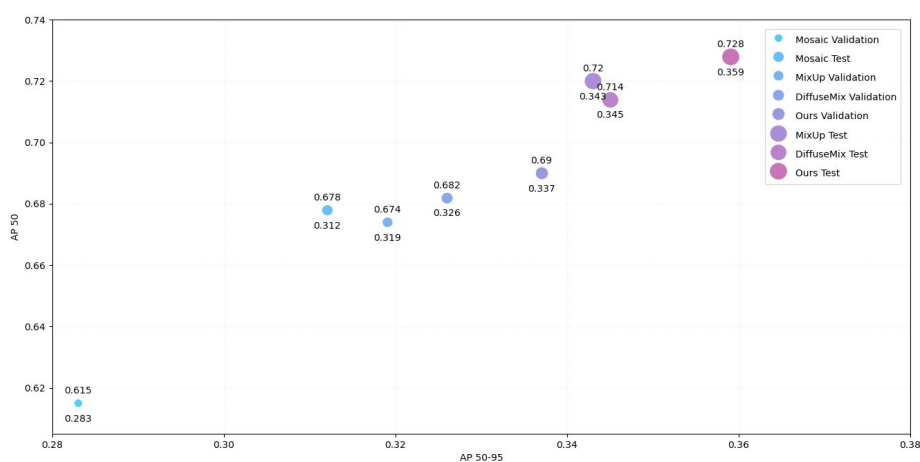


图 4.2 四种数据增强算法下的模型平均性能分布

Figure 4.2 Average Model Performance Distribution under Four Data Augmentation Algorithm

4.4 模型泛化能力

为进一步验证模型的泛化能力，本文使用了一张一共包含75个墓葬目标的图像样本，并基于先前组获取自2022年10月的Worldview2卫星影像，该数训练好的模型进行推理，这些墓葬样本并没有在先据的空间分辨率为0.3m。本文在该影像上截取出27前的实验中使用。各个模型的性能如表4.4所示。

表4.5 Worldview2亚米级影像下各个模型性能

Table 4.5 Performance of Various Models on WorldView-2 Sub-Meter Resolution Images					
模型	增强方法	AP ₅₀	AP ₅₀₋₉₅	Precision ₅₀	Recall ₅₀
YOLO v5n	Mosaic	0.635	0.316	0.75	0.53
	DiffuseMix	<u>0.650</u>	<u>0.330</u>	0.60	<u>0.64</u>
	MixUp	0.527	0.246	0.52	0.51
	Ours	0.687	0.344	<u>0.67</u>	0.64
YOLO v5m	Mosaic	0.615	0.307	0.68	0.56
	DiffuseMix	<u>0.666</u>	<u>0.313</u>	0.64	<u>0.60</u>
	MixUp	0.571	0.282	0.60	0.55
	Ours	0.695	0.332	0.82	0.60
YOLO v8n	Mosaic	0.668	0.311	0.77	0.54
	DiffuseMix	<u>0.715</u>	<u>0.343</u>	<u>0.77</u>	<u>0.63</u>
	MixUp	0.595	0.285	0.70	0.54
	Ours	0.778	0.354	0.76	0.74
YOLO v8m	Mosaic	0.604	0.268	0.69	<u>0.53</u>
	DiffuseMix	<u>0.678</u>	<u>0.338</u>	0.97	0.4
	MixUp	0.572	0.265	0.69	0.48
	Ours	0.726	<u>0.334</u>	<u>0.80</u>	0.67
YOLO v9t	Mosaic	0.537	0.231	<u>0.73</u>	0.29
	DiffuseMix	0.713	0.328	0.75	<u>0.64</u>
	MixUp	0.574	0.293	0.67	0.53
	Ours	<u>0.702</u>	<u>0.321</u>	0.70	0.69
YOLO v9m	Mosaic	0.576	0.261	0.77	0.49
	DiffuseMix	<u>0.592</u>	0.275	0.61	0.60
	MixUp	0.582	<u>0.307</u>	0.61	<u>0.59</u>
	Ours	0.618	0.319	<u>0.70</u>	0.53

模型	增强方法	AP ₅₀	AP ₅₀₋₉₅	Precision ₅₀	Recall ₅₀
YOLO v10n	Mosaic	0.585	<u>0.304</u>	0.66	<u>0.54</u>
	DiffuseMix	<u>0.614</u>	0.295	<u>0.66</u>	0.54
	MixUp	0.544	0.262	0.57	0.48
	Ours	0.637	0.330	0.75	0.48
YOLO v10m	Mosaic	0.569	0.293	0.72	<u>0.45</u>
	DiffuseMix	<u>0.593</u>	0.319	0.84	0.36
	MixUp	0.492	0.267	0.53	0.45
	Ours	0.611	<u>0.312</u>	<u>0.72</u>	0.52
RT-DETR L	Mosaic	0.657	0.27	0.64	0.64
	DiffuseMix	<u>0.76</u>	<u>0.335</u>	<u>0.77</u>	<u>0.71</u>
	MixUp	0.611	0.294	0.70	0.60
	Ours	0.782	0.339	0.80	0.77
RT-DETR X	Mosaic	0.573	0.26	0.60	0.60
	DiffuseMix	0.784	0.353	0.85	<u>0.76</u>
	MixUp	0.706	0.292	<u>0.79</u>	0.64
	Ours	<u>0.776</u>	<u>0.338</u>	0.78	0.80
平均性能	Mosaic	0.602	0.282	0.70	0.52
	DiffuseMix	<u>0.677</u>	<u>0.323</u>	<u>0.75</u>	<u>0.59</u>
	MixUp	0.587	0.281	0.63	0.54
	Ours	0.701	0.332	0.75	0.64

409 * 加粗数据表示同类比较最优结果，下划线数据表示同类比较次优结果

410 与先前的结果类似，经过数据增强后的模型精
411 确率普遍较高。由于该组影像中存在不少的小型非
412 墓葬地物，与目标对象有着相似的形状，例如圆形灌
413 木、圆形建筑、土坑、圆形石块，这些相似地物使得
414 模型的误检率上升，通过设置合适的置信度阈值，可
415 以在一定程度上筛选排除小型非墓葬目标，使精确
416 率进一步提升。

417 就整体而言，在不同传感器拍摄的独立遥感影
418 像数据集下，基于本文算法增广下的模型平均 AP₅₀
419 可以达到0.701，精确率达到75%，召回率达到64%，
420 展现出良好的泛化能力性，可以初步迁移到其他墓
421 葬数据集中使用。与基线模型相比，本文算法在多种
422 指标上均有较大的提升，尤其体现在召回率的提升。
423 通过复杂图形融合和环境语义变换，模型可以从有
424 限的样本中学习到的更多的信息，进一步提升在不同

426 环境下的适应能力。与原始数据集上的亮眼表现不
427 同，MixUp 方法在异源数据集的表现最差，其 AP_{50}
428 与 AP_{50-95} 指标变化幅度分别为-18.5%和-18.1%，高于
429 基线模型的-11.2%和-9.6%。这说明该方法在原始数
430 据集上发生了过拟合，并不适用于模型迁移使用。而
431 本文算法的变化幅度为-3.7%和-7.5%，说明增广后

432 的数据集可以让模型学习到更多额外的特征，增强
433 模型的泛化能力和鲁棒性。如图 4.3 所示，通过模型
434 推理，该组影像样本上所有的墓葬对象均被准确识
435 别，这表明本文研究在引入实际田野考察墓葬探测工
436 作上具有一定的可行性。

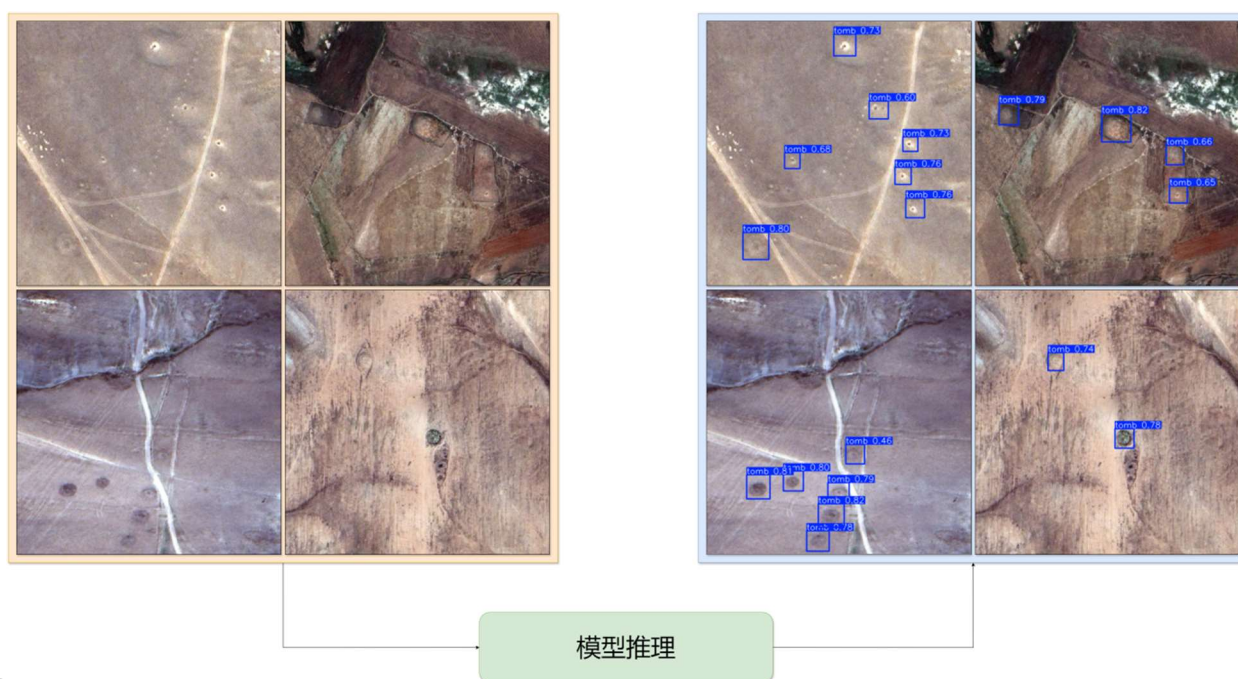


图 4.3 模型推理验证

Figure 4.3 Model Inference Validation

5 讨论

本文基于谷歌地球影像，构建了一套高分辨率的阿勒泰地区墓葬遥感影像数据集。针对阿勒泰地区样本分布较广、环境差异较大，导致模型识别性能下降、泛化能力下降等问题，本文提出了一种基于环境增强的数据增广算法，通过设置环境提示词，扩散模型可以在保留原始图像结构的情况下扩展环境语义，这些环境语义在一定程度上解决原始数据在不同环境下的样本不平衡问题，降低环境背景因子对模型判别的影响。基于该数据增广算法，显著提升了模型在数据集上的检测性能和泛化能力。与传统的图像增强方法相比，该方法可以在保留样本结构特征和标签位置的情况下，模拟不同环境下的样本状态，适用于环境差异较大的墓葬目标检测。本文将几种数据增广算法在不同源数据集上进行测试，结果表明本文所提出的方法具有良好的泛化能力。

本文首次引入了使用环境变量的扩散模型，并结合复杂分形图片与随机增强算法对原始遥感影像进行数据增广，取得了良好的结果，为遥感考古少量样本识别、处理提供了新的参考思路。此外，我们评估了

多种先进目标检测模型在本文提出的墓葬数据集上的表现，从结果精度来看，基于Transformer的目标检测模型具有更大的优势，但由于架构本身的区别，训练和推理速度远低于CNN架构。

为了方便比较不同数据增强方式在本数据集上的综合表现，本文将模型平均AP₅₀和AP₅₀₋₉₅总结为表5.1和表5.2。相较于基线方法，本文提出的算法在测试集、验证集和异源数据集上都表现出更好的性能。在测试集上，本文算法分别令模型平均AP₅₀和AP₅₀₋₉₅提升了7.4%和15.1%，表明本文算法可以显著提升模型的识别能力。而在异源数据集上，本文算法使得模型的平均AP₅₀和AP₅₀₋₉₅分别提升了16.4%和17.1%，这表明本文算法在提升模型的泛化能力上具有较大的贡献。相较于其他数据增强方法，本文算法也取得了最好的结果。

在利用深度学习进行墓葬的自动检测中，使用数据增强技术有利于提高模型的稳定性，不需要引入额外的真实样本数据，就能在不破坏检测目标本身的情况下模拟不同环境下的样本状态，保留原始检测目标的重要信息的同时，降低环境背景带来的干扰信息，增强模型的泛化能力。

表5. 1 不同数据增强方式的平均AP₅₀性能比较

Table 5.1 Comparison of Average AP50 Performance for Different Data Augmentation Methods

数据集	Base (Mosaic)	DiffuseMix	MixUp	本文算法
测试集	0.678	0.714 (+5.3%)	<u>0.720</u> (+6.2%)	0.728 (+7.4%)
验证集	0.615	<u>0.682</u> (+10.9%)	0.674 (+9.6%)	0.690 (+12.2%)
异源数据集	0.602	<u>0.677</u> (+12.5%)	0.587 (-2.49%)	0.701 (+16.4%)

* 加粗数据表示同类比较最优结果，下划线数据表示同类比较次优结果

表5.2 不同数据增强方式的平均AP₅₀₋₉₅性能比较

Table 5.2 Comparison of Average AP₅₀₋₉₅ Performance for Different Data Augmentation Methods

数据集	Base (Mosaic)	DiffuseMix	MixUp	本文算法
测试集	0.312	<u>0.345</u> (+10.6%)	0.343 (+9.9%)	0.359 (+15.1%)
验证集	0.283	<u>0.326</u> (+15.2%)	0.319 (+12.7%)	0.337 (+19.1%)
异源数据集	0.282	<u>0.323</u> (+14.5%)	0.281 (-0.03%)	0.332 (+17.7%)

* 加粗数据表示同类比较最优结果，下划线数据表示同类比较次优结果

6 结论

由于历史环境变迁和人类活动影响，地表考古遗迹往往难以完好保存，导致遥感影像上晰、有效的样本较少，数据样本收集难度大。较少的样本数量可能会引起模型过于关注样本特征，从而导致过拟合现象。数据增广技术往往可以解决少量样本过拟合问题，提高深度学习模型的性能。本文将扩散模型引入到遥感影像的数据增广研究中，通过提升阿勒泰地区的墓葬数据集质量的方式，一定程度上改善了模型性能。针对复杂环境下的样本，引入环境提示词，根据不同环境特征对原始图像进行环境语义增广，可以帮助模型理解环境背景，避免因某种环境占比较多导致的样本不均衡问题，从而使模型具有更好的泛化能力，提升模型对样本特征的关注度。该方法对于在大范围地区和环境复杂条件下开展墓葬的考古调查具有重要的现实意义和应用价值。

在后续研究中，我们将进一步调整扩散模型参数与提示词，比较不同提示词模板生成的结果，以及其

对性能可能产生的影响。我们将考虑采用线性时间复杂度Mamba架构改进目标检测模型，并深入挖掘多尺度特征在遥感考古目标检测中的作用。

参考文献

- AZIZI S, KORNBLITH S, SAHARIA C, et al. 2023. Synthetic data from diffusion models improves imagenet classification. arXiv preprint arXiv:2304.08466 [J].
- BERGANZO-BESGA I, ORENGO H A, LUMBRERAS F, et al. 2021. Hybrid MSRM-Based Deep Learning and Multitemporal Sentinel 2-Based Machine Learning Algorithm Detects Near 10k Archaeological Tumuli in North-Western Iberia. Remote Sensing [J], 13: 4181.
- BOCHKOVSKIY A, WANG C-Y, LIAO H-Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]//
- BROOKS T, HOLYNSKI A, EFROS A A. InstructPix2Pix: Learning to Follow Image Editing Instructions[C]//
- Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective data augmentation with diffusion models. In ICLR 2023 Workshop on Mathematical and Empirical Understanding of Foundation Models, 2023.
- CASPARI G, CRESPO P 2019. Convolutional neural networks for archaeological site detection-Finding “princely” tombs.

- Journal of Archaeological Science [J], 110: 104998.
- CHEN F, ZHOU R, VAN DE VOORDE T, et al. 2021. Automatic detection of burial mounds (kurgans) in the Altai Mountains. ISPRS Journal of Photogrammetry and Remote Sensing [J], 177: 217-237.
- CHEN Y-C, LU C-S. Rankmix: Data augmentation for weakly supervised learning of classifying whole slide images with diverse sizes and imbalanced categories[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.2023:23936-23945.
- CHUN D, CHOI J, KIM H, et al. A study for selecting the best one-stage detector for autonomous driving[C]//2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC).IEEE,2019:1-3.
- DAVIS J, GOADRIC M. The relationship between Precision-Recall and ROC curves[C]//Proceedings of the 23rd international conference on Machine learning.2006:233-240.
- HENDRYCKS D, MU N, CUBUK E D, et al. 2019. Augmix: A simple data processing method to improve robustness and uncertainty. arXiv preprint arXiv:1912.02781 [J].
- Hendrycks D , Zou A , Mazeika M ,et al.PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures[J]. 2021.DOI:10.48550/arXiv.2112.05135.
- HUANG Z, BAO X, ZHANG N, et al. 2023. Ipmix: Label-preserving data augmentation method for training robust classifiers. Advances in Neural Information Processing Systems [J], 36: 63660-63673.
- ISLAM K, ZAHEER M Z, MAHMOOD A, et al. DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.2024:27621-27630.
- KIM W, NIE Y, ZHU J, et al. 2013. Local orientation based detection of circular soil marks of ancient graves by GA. Yaogan Xuebao-Journal of Remote Sensing [J], 17: 671-678.
- LI Z, LI Y, ZHAO P, et al. 2023. Is synthetic data from diffusion models ready for knowledge distillation? arXiv preprint arXiv:2305.12954 [J].
- Lu, M., Wang, H. S., & Xun, X. 2024. Spatiotemporal Evolution of Habitat Quality and Its Driving Factors in the Altai Region. Journal of Beijing Forestry University, 46, 27-39.(卢梦, 王鹤松, 荀雪 2024. 阿勒泰地区生境质量的时空演变及其驱动因素. 北京林业大学学报 [J], 46: 27-39.)
- Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision, pages 6023–6032, 2019.
- SOKOLOVA M, LAPALME G 2009. A systematic analysis of performance measures for classification tasks. Information processing & management [J], 45: 427-437.
- SOROUGH M, MEHRTASH A, KHAZRAEE E, et al. 2020. Deep learning in archaeological remote sensing: Automated qanat detection in the Kurdistan region of Iraq. Remote Sensing [J], 12: 500
- .Su, Y. (2024). A study on Han Dynasty tombs in the Shanxian region. Cultural Relic Identification and Appreciation, (6), 156–159.
- 苏雨露. 陕县地区汉代墓葬研究[J]. 文物鉴定与鉴赏, 2024, (06): 156-159.
- Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J Fleet. Synthetic data from diffusion models improves imagenet classification. arXiv preprint arXiv:2304.08466, 2023.
- Shao, H. Q. 2008. On Two Types of Bronze Cultures in the Altai Region of Xinjiang. Studies in Western Regions, 59-65.(邵会秋 2008. 试论新疆阿勒泰地区的两类青铜文化. 西域研究 [J]: 59-65.)
- Su, J. J., Liu, Y. P., Liu, L. Y., et al. 2023. Quantitative Classification and Ordination Analysis of Typical Plant Communities in the Altai Region, Xinjiang. Journal of Grassland Science, 32, 50.(苏金娟, 刘永萍, 刘丽燕, et al. 2023. 新疆阿勒泰地区典型植物群落数量分类与排序分析. 草业学报 [J], 32: 50.)
- TRABUCCO B, DOHERTY K, GURINAS M, et al. Effective Data Augmentation With Diffusion Models[C]//
- VERSCHOOF-VAN DER VAART W B, LAMBERS K 2019. Learning

- 600 to look at LiDAR: The use of R-CNN in the automated
601 detection of archaeological objects in LiDAR data from the
602 Netherlands. Journal of Computer Applications in
603 Archaeology [J], 2: 31-40.
- 604 WANG A, CHEN H, LIU L, et al. YOLOv10: Real-Time End-to-End
605 Object Detection[C]//
606 WANG C-Y, YEH I H, LIAO H-Y M. YOLOv9: Learning What You
607 Want to Learn Using Programmable Gradient In
608 formation[C]//
609 WANG G, CHEN Y, AN P, et al. 2023. UAV-YOLOv8: A small-object-
610 detection model based on improved YOLOv8 for UAV aerial
611 photography scenarios. Sensors [J], 23: 7190.
- 612 Yuan-Chih Chen and Chun-Shien Lu. Rankmix: Data augmentation for
613 weakly supervised learning of classifying whole slide images
614 with diverse sizes and imbalanced categories. In Proceedings
615 of the IEEE/CVF Conference on Computer Vision and
616 Pattern Recognition, pages 23936–23945, 2023.
- 617 Yin Cui, Aravind Srinivas ,et al. 2021. Simple Copy-Paste is a Strong
618 Data Augmentation Method for Instance Segmentation[J].
619 2021.DOI:10.48550/arXiv.2012.07177
- 620 Yu, L. J. 2018. Spatial Analysis and Remote Sensing Detection of
621 Neolithic Archaeological Sites in the Taihu Basin. Beijing:
622 University of Chinese Academy of Sciences.(于丽君 2018.
- 623 太湖流域新石器时期考古遗址空间分析与遥感探测研究
624 [M]. 北京: 中国科学院大学.)
- 625 ZHANG H. 2017. MixUp: Beyond empirical risk minimization. arXiv
626 preprint arXiv:1710.09412 [J].
- 627 ZHAO Y, LV W, XU S, et al. DETRs Beat YOLOs on Real-time Object
628 Detection[C]//2024 IEEE/CVF Conference on Computer
629 Vision and Pattern Recognition (CVPR).2024:16965-16974.
630 10.1109/CVPR52733.2024.01605.
- 631 Zheng Li, Yuxuan Li, Penghai Zhao, Renjie Song, Xiang Li, and Jian
632 Yang. Is synthetic data from diffusion models ready for
633 knowledge distillation? arXiv preprint arXiv:2305.12954,
634 2023.
- 635 ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5
636 based on transformer prediction head for object detection on
637 drone-captured scenarios[C]//Proceedings of the IEEE/CVF
638 international conference on computer vision.2021:2778-2788.
- 639 Zeng, B. D. 2017. Discovery and Study of Stone Coffin Tombs in Altai
640 Region, Xinjiang. Silk Road Perspectives, 53-56.(曾宝栋
641 2017. 新疆阿勒泰地区的石棺墓发现和研究. 丝路视野
642 [J]: 53-56.)

Application of Environmental Semantic Data Augmentation Methods in Ancient Tomb Detection

Sihang Chen^{1,2}, Lijun Yu^{1,2}, Jianfeng Zhu^{1,2}, Jie Chen^{1,2}, Ze Liu³, Hui Wang⁴, Yueping Nie^{1,2}

1. *Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China.*

2. *University of Chinese Academy of Sciences, Beijing 101408, China*

3. *Research Centre for Spatial Planning, Ministry of Natural Resources, 100034*

4. *Fudan University, 200433*

Abstract: In the field of deep learning, the performance of object detection models typically relies on abundant and high-quality annotated datasets. However, in remote sensing archaeology applications, obtaining large-scale and diverse datasets is challenging and costly, leading to data scarcity issues that cause models to overfit during training. Particularly in cases where the environmental conditions of burial sites vary significantly, the overrepresentation of specific environmental features can lead to sample imbalance, which restricts the model's adaptability to different environments. To address these issues, this paper proposes a remote sensing image data augmentation algorithm based on diffusion models. By setting different environmental prompts, the algorithm enriches the visual semantics of images without altering the original dataset's label distribution, thereby alleviating data scarcity and sample imbalance problems. This study constructed a high-resolution burial site image dataset for the Altai region using Google Earth and evaluated the performance of various object detection models with this algorithm. Experimental results demonstrate that the proposed algorithm improves the average precision on the test set by 12.2% compared to the baseline method and enhances average precision on heterogeneous datasets by 16.4%. The proposed algorithm significantly improves the model's detection accuracy, stability, and adaptability across different datasets, providing effective technical support for small-sample remote sensing archaeological object recognition and expanding the research perspectives in intelligent archaeological detection.

Objective

The performance of deep learning-based object detection models in remote sensing archaeology largely depends on the availability of sufficient, high-quality annotated datasets. However, acquiring large-scale, diverse datasets is challenging and costly. This leads to issues such as data scarcity, sample imbalance, and overfitting in model training, especially when burial sites exist in significantly different environmental conditions. To overcome these challenges, this study proposes a data augmentation technique based on environmental semantic enhancement using diffusion models. The method aims to enrich the visual semantics of remote sensing images

while preserving the original dataset's label distribution, thereby improving the adaptability and accuracy of object detection models in archaeological applications.

Method

This research introduces an environment-enhanced remote sensing image augmentation algorithm leveraging diffusion models. The algorithm generates augmented images by setting different environmental prompts, simulating realistic burial site conditions such as deserts, grasslands, forests, and snow-covered regions. Additionally, the method incorporates fractal patterns to enhance image complexity and robustness. A high-resolution burial site dataset from the Altai region was constructed using Google Earth images, including 298 effective images covering 938 burial sites. Various object detection models, including YOLO v5, YOLO v8, YOLO v9, YOLO v10, and RT-DETR, were trained using different augmentation techniques such as Mosaic, MixUp, and DiffuseMix. The proposed algorithm was evaluated through comparisons with these conventional augmentation methods.

Result

Experimental results indicate that the proposed algorithm significantly improves model performance. Compared to the baseline Mosaic augmentation, the proposed method increased average precision (AP50) on the test set by 12.2% and on heterogeneous datasets by 16.4%. The method also improved AP50-95, precision, and recall, demonstrating enhanced model robustness and cross-dataset generalization. The augmentation approach successfully mitigated the negative impact of sample imbalance and environmental diversity on detection accuracy. Furthermore, performance evaluation on an independent WorldView-2 satellite dataset showed that the proposed algorithm improved generalization capability, outperforming conventional augmentation techniques.

Conclusion

The study presents an effective data augmentation approach for remote sensing archaeological target detection by integrating environmental semantic enhancement through diffusion models. This method successfully addresses the challenges of data scarcity and sample imbalance, improving model accuracy, stability, and adaptability across various environments. The findings highlight the feasibility of using generative models for archaeological applications, providing a novel perspective for enhancing small-sample object detection in

remote sensing archaeology. Future research will focus on optimizing diffusion model parameters, refining environmental prompts, and exploring transformer-based architectures to further enhance detection performance in complex archaeological settings.

Key words: Deep Learning; Ancient Tomb Detection; Data Augmentation; Diffusion Model; Remote Sensing Archaeology; Few-Shot Learning; Environment-Semantic Enhancement; Model Generalization; Image Generation

Supported by: National Key Research and Development Projects (grant no. 2020YFC1521900)、National Natural Science Foundation of China (grant no. 41801134).