

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, 北京 100034;

4. 复旦大学 科技考古研究院, 上海 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。为此, 本文提出了一种基于扩散模型的遥感影像数据增强算法。该算法使用微调的扩散模型, 结合分形图形和随机图像增强链路, 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。本文在自建的阿勒泰地区高分辨率墓葬影像数据集中评估了该算法与经典数据增强算法的性能表现, 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

中图分类号: TP751/P2

引用格式: 陈思航, 于丽君, 朱建峰, 陈婕, 柳泽, 王辉, 聂跃平. 2026. 环境语义数据增强方法在古墓葬检测中的应用. 遥感学报, 30(2): 403-415

Chen S H, Yu L J, Zhu J F, Chen J, Liu Z, Wang H and Nie Y P. 2026. Application of environmental semantic data augmentation methods in ancient tomb detection. National Remote Sensing Bulletin, 30(2): 403-415 [DOI: 10.11834/jrs.20254414]

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目标检测已经取得了实质性的进展, 例如, Caspari 等 (2019) 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; Verschoof-van 等 (2019) 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; Chen 等 (2021) 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探测到了墓葬的空间分布; Berganzo-Besga 等 (2021) 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福尼亚地区不同环境下的墓葬;

收稿日期: 2024-09-27; 预印本: 2025-03-14

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

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

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

Soroush 等 (2022) 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。

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 (Chen 和 Lu, 2023)。经典的增强算法如 CutMix (Yun 等, 2019)、Mixup (Zhang 等, 2017) 采用线性插值运算对图像进行混合增强; AugMix (Hendrycks 等, 2019)、IPMix (Huang 等, 2023) 采用随机组合的方式模拟色彩结构语义变化; PixMix (Hendrycks 等, 2021) 引入了复杂分形图形对原图像进行增强, 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 (Ghiasi 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, 在样本数量较少的环境中表现欠佳。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。鉴于此, 本研究基于扩散模型提出了一种新的数据增强算法, 通过不同的环境提示词, 在不破坏原始数据集标签分布的情况下, 缓解目标的背景环境不均衡情况。最后, 使用自建的新疆维吾尔自治区阿勒泰地区遥感考古墓葬数据集开展实验对比分析。

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, 南接准噶尔盆地, 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。这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 1 所示。可见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。

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。本文选择的样本点分布如图 2 所示。

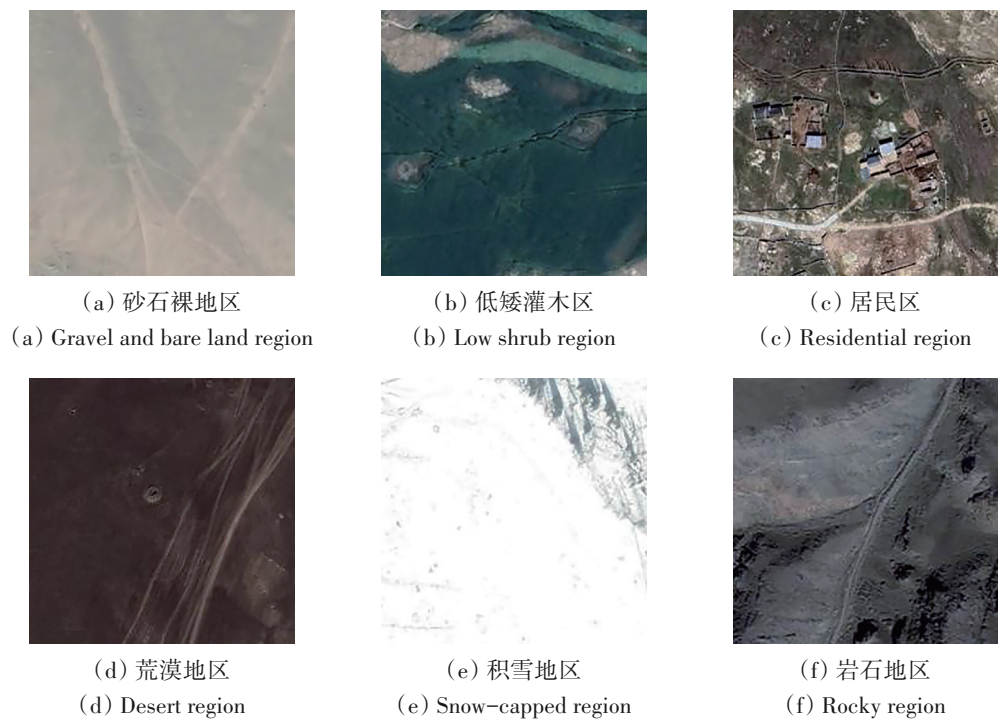


图1 阿勒泰地区古墓葬的典型赋存环境
Fig. 1 Surrounding environments of the ancient tombs distributed in the Altai region

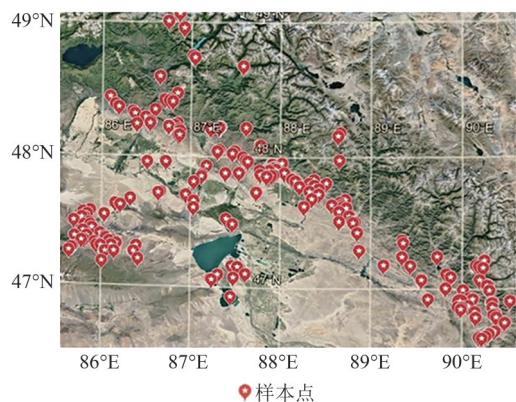


图2 阿勒泰地区部分样本点分布
Fig. 2 Distribution of selected sample points in the Altai region

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为0.3—1.0 m，成像时间为2019年—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，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1—17个，共计938个墓葬。将全部数据按照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像素×640像素。

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实验，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，2024)。

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

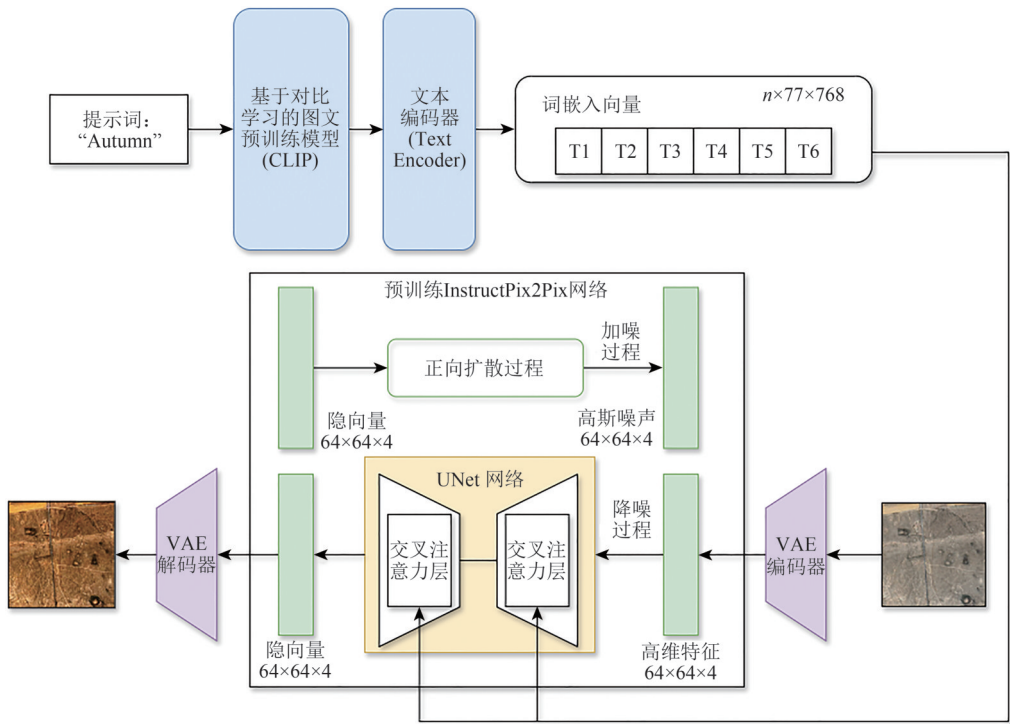


图4 基于 Prompt 的图生图工作

Fig. 4 Prompt-based image-to-image processing

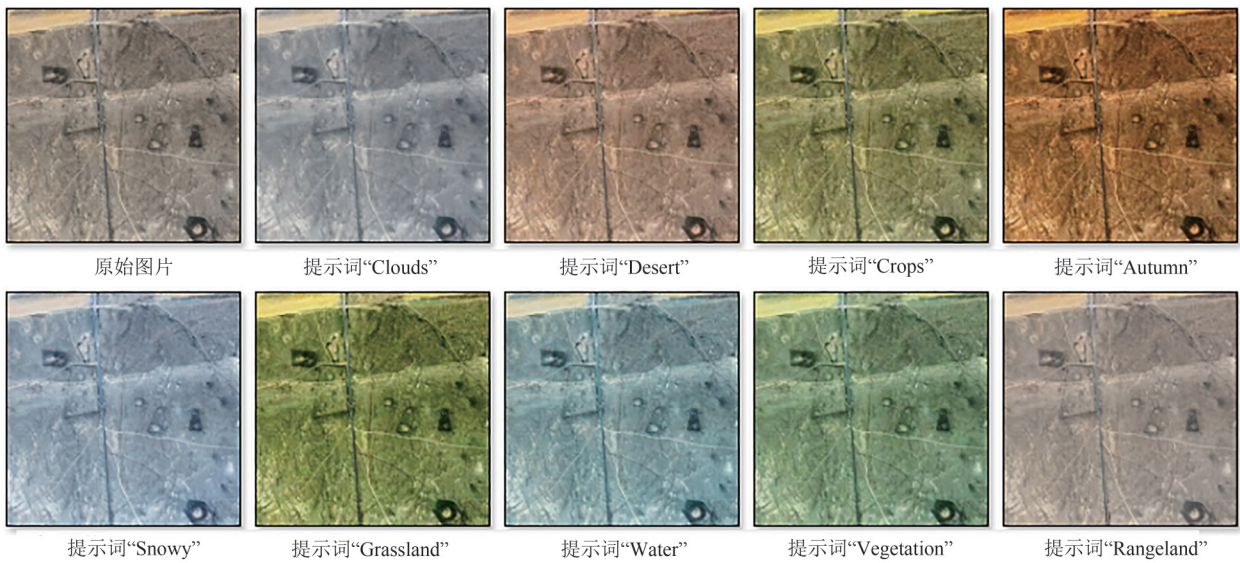


图5 不同 Prompt 生成的合成图片

Fig. 5 Composite images generated by different prompt

为了增强图像的安全性能，本研究沿用 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，2023），具体流程如图 6 所示。目前而言，尚未有相关文献将分形图形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，本研究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，本研究发

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。

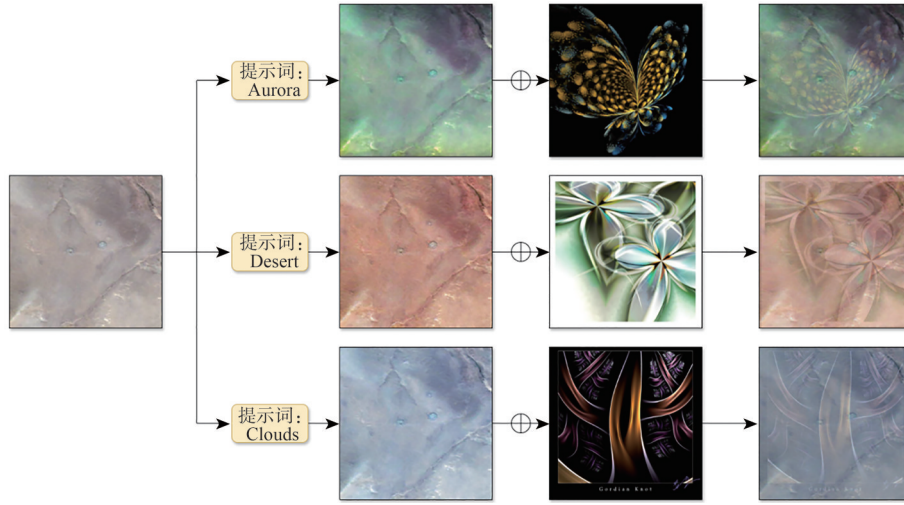


图6 引入分形图形的合成图片示例

Fig. 6 Composite images with introduced fractal patterns

3.3.3 具体方法

定义 I_i 为输入图片, D_{aug} 表示本文算法流程, Y 表示输出影像, G 表示合成图片生成操作, F 表示分形图形融合操作, A 表示图像增强操作, w 表示图像聚合过程中的权重系数, 则可以写作: $Y = D_{\text{aug}}(I, G, F, A, w)$ 。

将图像生成操作定义为 $G(I, p)$, 其中 I 表示输入图片, p 是从预定义好的指令集 P 中随机抽取的指令。在本实验中, 指令集 P 被定义为: $P = \{ \text{"River", "Autumn", "Built Area", "Clouds", "Crops", "Desert", "Flooded Vegetation", "Forest", "Grassland", "Rangeland", "Snowy/Ice", "Water"} \}$, 经过处理后得到合成影像 I^G 可表示为

$$I^G = G(I, p) \quad (1)$$

将简单的图像增强操作定义为 $A(I, o, k)$, 其中 I 表示输入图片, o 表示从预先定义好的图像增强操作集合 O 中随机抽取的结果, 在本实验中, 本研究设 $O = \{ \text{posterize, translate, equalize, shear, solarize, contrast, brightness, sharpness} \}$, k 表示执行的次数, 预设区间 $[1, 4]$ 的随机整数。经过图像增强后的影像可以表示为

$$I^A = A(I, o, k) \quad (2)$$

本研究定义图像乘法操作为 $\text{ImgA} \otimes \text{ImgB}$, 其中, ImgA 、 ImgB 表示两张大小相同的图像,

$\text{ImgA} \otimes \text{ImgB} = \left((\text{ImgA} \times 2)^\alpha \times (\text{ImgB} \times 2)^\beta \right) / 2$, 其中, α 与 β 表示从贝塔分布中取得的一组对应随机数, 整个算法可以表示为

$$Y = w \times I_i + (1 - w) \times (WI) \quad (3)$$

式中, W 表示权重向量 $[w_1, w_2, w_3]$, 为一组预先设置好的权重, 在本实验中设置为 $W = [0.12, 0.68, 0.2]$, 该组权重值为 AugMix (Hendrycks 等, 2019) 经过实验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 $[I^A, I^G, I^K]$, 其中, I^A 表示原始图片经过简单增强后的结果, I^G 表示原始图片通过指令生成的合成图片, $I^G = I^G \otimes I^A$, I^K 表示从集合中随机抽取的分形图形增强结果, $I^K = f_k \otimes I^G$ 。

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 3 个模块, 本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, 本研究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 $O_{\text{aug}} = O(N)$, 同时, 对图像进行融合时

间复杂度也是线性的，记作 $O_{\text{mix}} = O(N)$ 。

图像生成的计算瓶颈取决于扩散模型的复杂度，一般来说，每个时间步的计算复杂度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 $O_{\text{gen}} = O(T \times N^2)$ 。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 $O = \sum_i^{\{\text{aug}, \text{gen}, \text{mix}\}} O_i = O(T \times N^2)$ ，其中 T 表示推理时使用的时间步的数量， 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，本研究以生成 2000 张合成图片为例，进行相关的时间开销计算。图片的宽度和高度均为 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。

各类算法时间开销的结果见表 1，本文算法与 DiffuseMix 算法对单张图像的平均处理时间超过了 1.7 s，而 Mosaic 和 MixUp 算法的平均处理时间均小于 0.2 s。相对而言，基于扩散模型的图像处理方法推理时间开销较大，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。

表 1 不同算法的耗费时间

Table 1 Time consumption of different algorithms

算法	总时间	平均处理时间
本文算法	3512.4	1.756
Diffuse Mix	3467.2	1.733
Mosaic	29.4	0.147
MixUp	11.2	0.056

4 实验结果与分析

4.1 模型训练结果

为了分析不同数据增强算法在墓葬图像识别上的效果，以及比较主流的目标识别模型在遥感墓葬识别上的性能，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 数据集上的目标检测精度见表 2。

表 2 模型基本数据

Table 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	42.0	136.0	53.1
RT-DETR X	76.0	259.0	54.3

为了方便比对，本研究在相同的实验环境、相同的墓葬数据集下，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，本研究对所有的模型采取 Mosaic (Bochkovskiy 等, 2020) 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。

4.2 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。

如表 3 所示，验证结果表明 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；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3 模型性能表现
Table 3 Model performance

模型	增强方法	(<i>T:V</i>)AP ₅₀	(<i>T:V</i>)AP ₅₀₋₉₅	(<i>T:V</i>)Precision ₅₀	(<i>T:V</i>)Recall ₅₀
YOLO v5n	Mosaic	0.676:0.509	0.289:0.218	0.76:0.76	0.60:0.60
	DiffuseMix	0.700:0.621	0.332:0.274	0.80:0.71	0.62:0.58
	MixUp	0.712:0.650	0.325:0.307	0.72:0.72	0.63:0.60
	Ours	0.721:0.646	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	0.331:0.326	0.72:0.75	0.66:0.66
	MixUp	0.734:0.678	0.359:0.310	0.77:0.74	0.66:0.60
	Ours	0.744:0.699	0.368:0.357	0.75:0.70	0.70:0.63
YOLO v8n	Mosaic	0.701:0.650	0.340:0.302	0.80:0.76	0.59:0.56
	DiffuseMix	0.721:0.686	0.366:0.354	0.79:0.73	0.64:0.63
	MixUp	0.716:0.675	0.339:0.332	0.64:0.66	0.71:0.65
	Ours	0.737:0.693	0.377:0.355	0.75:0.78	0.78:0.58
YOLO v8m	Mosaic	0.731:0.636	0.343:0.318	0.79:0.66	0.63:0.64
	DiffuseMix	0.701:0.634	0.355:0.330	0.76:0.67	0.63:0.59
	MixUp	0.742:0.698	0.353:0.321	0.79:0.68	0.66:0.61
	Ours	0.702:0.676	0.379:0.334	0.74:0.76	0.62:0.55
YOLO v9t	Mosaic	0.638:0.557	0.302:0.272	0.66:0.68	0.56:0.47
	DiffuseMix	0.717:0.692	0.363:0.340	0.74:0.64	0.64:0.66
	MixUp	0.717:0.685	0.369:0.335	0.74:0.66	0.61:0.63
	Ours	0.706:0.686	0.361:0.333	0.78:0.72	0.64:0.60
YOLO v9m	Mosaic	0.662:0.647	0.294:0.276	0.74:0.73	0.58:0.53
	DiffuseMix	0.740:0.682	0.357:0.352	0.80:0.69	0.68:0.62
	MixUp	0.735:0.679	0.362:0.333	0.76:0.76	0.68:0.58
	Ours	0.731:0.701	0.376:0.350	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	0.719:0.657	0.367:0.321	0.81:0.73	0.63:0.55
	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	0.76:0.76	0.67:0.64
YOLO v10m	Mosaic	0.708:0.638	0.335:0.302	0.82:0.68	0.57:0.54
	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	0.701:0.637	0.355:0.340	0.77:0.73	0.61:0.52
RT-DETR L	Mosaic	0.695:0.671	0.304:0.296	0.87:0.78	0.58:0.58
	DiffuseMix	0.717: 0.735	0.307:0.303	0.80:0.80	<u>0.66:0.67</u>
	MixUp	0.738:0.697	0.328:0.316	<u>0.84:0.72</u>	0.72:0.63
	Ours	<u>0.732:0.701</u>	<u>0.318:0.308</u>	0.79:0.77	0.64:0.60
RT-DETR X	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>0.773:0.754</u>	0.345:0.340	<u>0.74:0.74</u>	0.75:0.72
	Ours	0.780:0.792	<u>0.342:0.342</u>	0.76:0.72	<u>0.73:0.80</u>
平均性能	Mosaic	0.678:0.615	0.312:0.283	<u>0.76:0.71</u>	0.59:0.56
	DiffuseMix	0.714: <u>0.682</u>	<u>0.345:0.326</u>	0.77:0.72	0.65: <u>0.62</u>
	MixUp	<u>0.720:0.674</u>	0.343:0.319	0.72:0.70	<u>0.67:0.62</u>
	Ours	0.728:0.690	0.359:0.337	0.75:0.75	0.68:0.62

注：*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所示，相较于基线处理方案，本文提出的算法使得测试集中模型平均性能中的AP₅₀提升到了0.728，提升率为7.4%，AP₅₀₋₉₅、召回率的提升分别为15.1%、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%、10.7%。本文算法生成的复杂图片有效避免了过拟合，相较于基线结构模型在验证集和测试集上差异较大，本文算法驱动模型在验证集和

测试集上差异较小, 性能更加稳定。DiffuseMix 和 MixUp 方法均表现出优于基线模型的性能, MixUp 在测试集上表现略高于 DiffuseMix 方法, 但在验证集上与后者存在一定差距。相较于比较的数据增强算法, 本文算法表现出较大的优势。在验证集与测试集上, 本文算法驱动下的模型平均精度没有发生显著变化, 而其他增强算法均表现出一定

程度的下降, 基线方法的下降幅度最多, 达到了 6.6%。在召回率上, 除却基线方法, 其余方法在验证集上均下降至 0.62, 一定程度上表现为验证集数据所带来的瓶颈。在 AP_{50} 与更加严格的 AP_{50-95} 指标上, 基线方法与 MixUp 算法均降低了 7% 以上, 本文算法与 DiffuseMix 算法的相对稳定, 降低幅度低于 6%。

表 4 平均性能对比
Table 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%)

注: 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。

综合考虑 4 种增强方式下模型的平均性能, 绘制出的模型性能图如图 7 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 7 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 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。

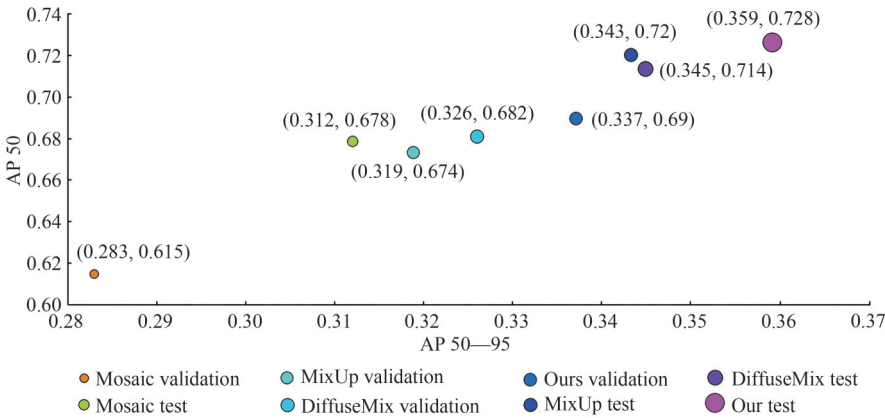


图 7 4 种数据增强算法下的模型平均性能分布对比

Fig. 7 Average model performance distribution under four data augmentation algorithm

4.3 模型泛化能力

为进一步验证模型的泛化能力, 本文使用了一组获取自 2022 年 10 月的 Worldview2 卫星影像, 该数据的空间分辨率为 0.3 m。本文在该影像上截取出 27 张一共包含 75 个墓葬目标的图像样本, 并基于先前训练好的模型进行推理, 这些墓葬样本并没有在先前的实验中使用。如表 5 所示, 与先前

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

表5 Worldview2亚米级影像下各个模型性能

Table 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	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	YOLO v10n	Mosaic	0.585	<u>0.304</u>	0.66	<u>0.54</u>
	DiffuseMix	<u>0.650</u>	<u>0.330</u>	0.60	<u>0.64</u>		DiffuseMix	<u>0.614</u>	0.295	<u>0.66</u>	0.54
	MixUp	0.527	0.246	0.52	0.51		MixUp	0.544	0.262	0.57	0.48
	Ours	0.687	0.344	<u>0.67</u>	0.64		Ours	0.637	0.330	0.75	0.48
YOLO v5m	Mosaic	0.615	0.307	0.68	0.56	YOLO v10m	Mosaic	0.569	0.293	0.72	<u>0.45</u>
	DiffuseMix	<u>0.666</u>	<u>0.313</u>	0.64	<u>0.60</u>		DiffuseMix	<u>0.593</u>	0.319	0.84	0.36
	MixUp	0.571	0.282	0.60	0.55		MixUp	0.492	0.267	0.53	0.45
	Ours	0.695	0.332	0.82	0.60		Ours	0.611	<u>0.312</u>	<u>0.72</u>	0.52
YOLO v8n	Mosaic	0.668	0.311	0.77	0.54	RT-DETR L	Mosaic	0.657	0.270	0.64	0.64
	DiffuseMix	<u>0.715</u>	<u>0.343</u>	<u>0.77</u>	<u>0.63</u>		DiffuseMix	<u>0.76</u>	<u>0.335</u>	<u>0.77</u>	<u>0.71</u>
	MixUp	0.595	0.285	0.70	0.54		MixUp	0.611	0.294	0.70	0.60
	Ours	0.778	0.354	0.76	0.74		Ours	0.782	0.339	0.80	0.77
YOLO v8m	Mosaic	0.604	0.268	0.69	<u>0.53</u>	RT-DETR X	Mosaic	0.573	0.260	0.60	0.60
	DiffuseMix	<u>0.678</u>	<u>0.338</u>	0.97	0.4		DiffuseMix	0.784	0.353	0.85	<u>0.76</u>
	MixUp	0.572	0.265	0.69	0.48		MixUp	0.706	0.292	<u>0.79</u>	0.64
	Ours	0.726	<u>0.334</u>	<u>0.80</u>	0.67		Ours	<u>0.776</u>	<u>0.338</u>	0.78	0.80
YOLO v9t	Mosaic	0.537	0.231	<u>0.73</u>	0.29	平均性能	Mosaic	0.602	0.282	0.70	0.52
	DiffuseMix	0.713	0.328	0.75	<u>0.64</u>		DiffuseMix	<u>0.677</u>	<u>0.323</u>	<u>0.75</u>	<u>0.59</u>
	MixUp	0.574	0.293	0.67	0.53		MixUp	0.587	0.281	0.63	0.54
	Ours	<u>0.702</u>	<u>0.321</u>	0.70	0.69		Ours	0.701	0.332	0.75	0.64
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						

注：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。

就整体而言，在不同传感器拍摄的独立遥感影像数据集下，基于本文算法增广下的模型平均AP₅₀可以达到0.701，精确率达到75%，召回率达到64%，展现出良好的泛化能力性，可以初步迁移到其他墓葬数据集中使用。与基线模型相比，本文算法在多种指标上均有较大的提升，尤其体现在召回率的提升。通过复杂图形融合和环境语义变换，模型可以从有限的样本中学习到的信息，进一步提升在不同环境下的适应能力。与原始数据集上的亮眼表现不同，MixUp方法在异源数据集的表现最差，其AP₅₀与AP₅₀₋₉₅指标分别降低了18.5%和18.1%，该方法极有可能在原始数据集上发生了过拟合，在跨域迁移上的表现不佳。本文算法的AP₅₀与AP₅₀₋₉₅指标分别降低了3.7%和7.5%，说明增广后的数据集可以让模型学习到更多额外的特征，增强模型的泛化能力和鲁棒性。

如图8所示，通过模型推理，该组影像样本上所有的墓葬对象均被准确识别，这表明本文研究在实际田野考墓葬探测工作中具有一定的可行性。

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。

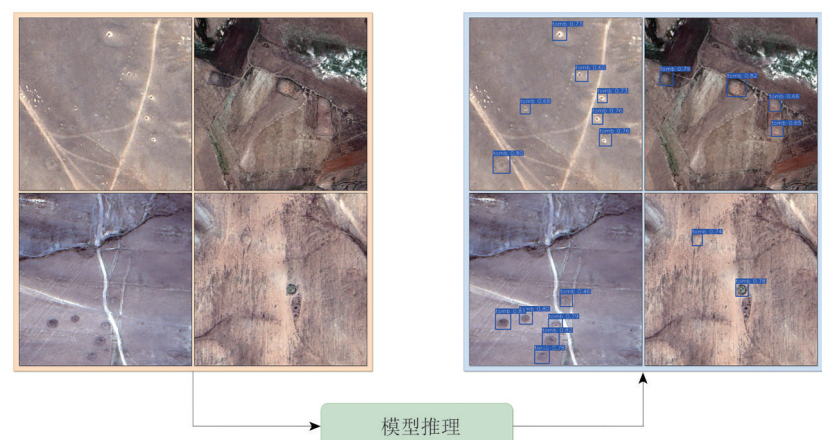


图8 模型推理验证
Fig. 8 Model inference validatio

本文首次引入了使用环境变量的扩散模型，并结合复杂分形图形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，本研究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 AP_{50} 和 AP_{50-95} 总结为表6和表7，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 AP_{50} 和 AP_{50-95} 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 AP_{50} 和 AP_{50-95} 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。

表6 不同数据增强方式的平均 AP_{50} 性能比较

Table 6 Comparison of average AP_{50} performance for different data augmentation methods

数据集	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%)

注：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7 不同数据增强方式的平均 AP_{50-95} 性能比较

Table 7 Comparison of average AP_{50-95} performance for different data augmentation methods

数据集	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%)

注：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。

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。

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，导致遥感影像上清晰、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,本研究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。本研究将考虑采用线性时间复杂度的 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(References)

- Kim J. M., Bader J., Alaniz S., Schmid C., Akata Z. (2025). DataDream: Few-Shot Guided Dataset Generation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds) *Computer Vision – ECCV 2024*. ECCV 2024. Lecture Notes in Computer Science, vol 15129. Springer, Cham. https://doi.org/10.1007/978-3-031-73209-6_15
- Berganzo-Besga I, Orenzo H A, Lumbreras F, Carrero-Pazos M, Fonte J and Vilas-Estévez B. 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*, 13(20): 4181 [DOI: 10.3390/rs13204181]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection. *arXiv preprint arXiv: 2004.10934* [DOI: 10.48550/arXiv.2004.10934]
- Brooks T, Holynski A and Efros A A. 2023. InstructPix2Pix: learning to follow image editing instructions//*Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE: 18392-18402 [DOI: 10.1109/CVPR52729.2023.01764]
- Caspari G. and P. Crespo, 2019. Convolutional neural networks for archaeological site detection – Finding “princely” tombs. *Journal of Archaeological Science*, 110: 104998 [DOI: 10.1016/j.jas.2019.104998]
- Chen F, Zhou R, Van De Voorde T, Chen X Z, Bourgeois J, Gheyle W, Goossens R, Yang J and Xu W B. 2021. Automatic detection of burial mounds (kurgans) in the Altai Mountains. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177: 217-237 [DOI: 10.1016/j.isprsjprs.2021.05.010]
- Chen Y C and Lu C S. 2023. RankMix: data augmentation for weakly supervised learning of classifying whole slide images with diverse sizes and imbalanced categories//*Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE: 23936-23945 [DOI: 10.1109/CVPR52729.2023.02292]
- Ghiasi G, Cui Y, Srinivas A, Qian R, Lin T Y, Cubuk E D, Le Q V and Zoph B. 2021. Simple copy-paste is a strong data augmentation method for instance segmentation//*Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE: 2917-2927 [DOI: 10.1109/CVPR46437.2021.00294]
- Hendrycks D, Mu N, Cubuk E D, Zoph B, Gilmer J and Lakshminarayanan B. 2019. AugMix: a simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv: 1912.02781* [DOI: 10.48550/arXiv.1912.02781]
- Hendrycks D, Zou A, Mazeika M, Tang L, Li B, Song D and Steinhardt J. 2021. PixMix: dreamlike pictures comprehensively improve safety measures. *arXiv preprint arXiv: 2112.05135* [DOI: 10.48550/arXiv.2112.05135]
- Huang Z L, Bao X N, Zhang N, Zhang Q Q, Tu X M, Wu B and Yang X. 2023. IPMix: label-preserving data augmentation method for training robust classifiers//*Proceedings of the 37th Conference on Neural Information Processing Systems*. New Orleans: Curran Associates Inc.: 63660-63673
- Islam K, Zaheer M Z, Mahmood A and Nandakumar K. 2024. Diffusemix: label-preserving data augmentation with diffusion models//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE: 27611-27620 [DOI: 10.1109/CVPR52733.2024.02608]
- Kim W, Nie Y P, Zhu J F, Deng B, Yu L J, Liu F and Gao H G. 2013. Local orientation based detection of circular soil marks of ancient graves by GA. *Journal of Remote Sensing*, 17(3): 671-678 [DOI: 10.11834/jrs.20132116]
- Li Z, Li Y X, Zhao P H, Song R J, Li X and Yang J. 2023. Is synthetic data from diffusion models ready for knowledge distillation?. *arXiv preprint arXiv: 2305.12954* [DOI: 10.48550/arXiv.2305.12954]
- Lu M, Wang H S and Xun X. 2024. Spatial-temporal evolution of habitat quality in Altay area and its driving factors. *Journal of Beijing Forestry University*, 46(3): 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*, 46(3): 27-39) [DOI: 10.12171/j.1000-1522.20230189]
- Shao H Q. 2008. Analysis of the two types bronze cultures in Altay, Xinjiang. *The Western Regions Studies*, (4): 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*, (4): 59-65) [DOI: 10.16363/j.cnki.xyj.2008.04.011]
- Soroush M, Mehrtash A, Khazraee E and Ur J A. 2020. Deep learning in archaeological remote sensing: automated qanat detection in the Kurdistan region of Iraq. *Remote Sensing*, 12(3): 500 [DOI: 10.3390/rs12030500]
- Su J J, Liu Y P, Liu L Y, Wu T Z and Wang M. 2023. Quantitative classification and ordination analysis of typical plant communities in Altay Prefecture of Xinjiang. *Acta Prataculturae Sinica*, 32(9): 50-67 (苏金娟, 刘永萍, 刘丽燕, 吴天忠, 王梅. 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*, 32(9): 50-67) [DOI: 10.11686/cyxb2022420]
- Su Y L. 2024. A study on Han Dynasty tombs in the Shanxian region. *Identification and Appreciation to Cultural Relics*, (6): 156-159 (苏雨露. 2024. 陕县地区汉代墓葬研究. *文物鉴定与鉴赏*, (6): 156-159) [DOI: 10.20005/j.cnki.issn.1674-8697.2024.06.038]
- Trabucco B, Doherty K, Gurinas M and Salakhutdinov R. 2025. Effective data augmentation with diffusion models. *arXiv preprint arXiv: 2302.07944* [DOI: 10.48550/arXiv.2302.07944]
- Verschoof-Van Der Vaart W B and Lambers K. 2019. Learning to look at LiDAR: the use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1): 31-40 [DOI: 10.5334/jcaa.32]
- Wang C Y, Yeh I H and Liao H Y M. 2024. YOLOv9: learning what you want to learn using programmable gradient information//*Proceedings of the 18th European Conference on Computer Vision*. Milan: Springer: 1-21 [DOI: 10.1007/978-3-031-72751-1_1]

- Wang G, Chen Y F, An P, Hong H Y, Hu J H and Huang T G. 2023. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 23(16): 7190 [DOI: 10.3390/s23167190]
- Yu L J. 2018. Spatial Analysis and Remote-Sensing-Based Detection of Neolithic Archaeological Sites in Lake Taihu Basin. Beijing: University of Chinese Academy of Sciences(Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences) (于丽君 . 2018. 太湖流域新石器时期考古遗址空间分析与遥感探测研究. 北京: 中国科学院大学(中国科学院遥感与数字地球研究所))
- Yun S, Han D, Chun S, Oh S J, Yoo Y and Choe J. 2019. CutMix: regularization strategy to train strong classifiers with localizable features//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 6022-6031 [DOI: 10.1109/ICCV.2019.00612]
- Zeng B D. 2017. Discovery and study of stone coffin tombs in Altai Region, Xinjiang. *Silk Road Vision*, (18): 53-56 (曾宝栋 . 2017. 新疆阿勒泰地区的石棺墓发现和研究. *丝路视野*, (18): 53-56)
- Zhang H Y, Cisse M, Dauphin Y N and Lopez-Paz D. 2017. mixup: beyond empirical risk minimization. *arXiv preprint arXiv: 1710.09412* [DOI: 10.48550/arXiv.1710.09412]
- Zhao Y A, Lv W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, Liu Y and Chen J. 2024. DETRs beat YOLOs on real-time object detection//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]
- Zhu X K, Lyu S C, Wang X and Zhao Q. 2021. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE: 2778-2788 [DOI: 10.1109/ICCVW54120.2021.00312]

Application of environmental semantic data augmentation methods in ancient tomb detection

CHEN Sihang^{1,2}, YU Lijun^{1,2}, ZHU Jianfeng^{1,2}, CHEN Jie^{1,2}, LIU Ze³, WANG HUI⁴, NIE Yueping^{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, Beijing 100034, China;

4. Institute of Scientific and Technological Archaeology, Fudan University, Shanghai 200433, China

Abstract: Deep learning-based object detection has become an important tool for large-scale burial site identification in remote sensing archaeology, with its performance highly dependent on sufficient and diverse annotated datasets. However, in practical archaeological applications, acquiring large-scale, high-quality samples is both expensive and time-consuming. Moreover, burial sites are often distributed across highly heterogeneous environments, leading to significant environmental imbalance within datasets. Under few-shot conditions, models tend to overfit to dominant background features and exhibit limited cross-scene generalization capabilities. This study aims to develop an environment-semantic data augmentation strategy that expands background diversity while preserving the original spatial structure and label distribution of burial targets. By simulating multiple environmental contexts through generative modeling, the proposed method can alleviate sample imbalance to a certain extent, enhance robustness to environmental variations, and improve cross-domain transferability, thereby providing a practical and scalable solution for few-shot remote sensing archaeological detection.

This study proposes a diffusion-based environment-semantic augmentation framework consisting of three components: environmental generation, fractal fusion, and random enhancement. A pre-trained InstructPix2Pix diffusion model is guided by predefined environmental prompts to simulate burial sites under diverse background conditions while preserving structural features. To further enhance robustness, fractal patterns are fused with images using a Beta-distribution-based weighted multiplication strategy, allowing texture-level enhancement without occluding targets. Additional random image operations are applied to increase variability. The method was evaluated on a self-constructed Altai burial dataset using multiple object detection models, with Mosaic, MixUp, and DiffuseMix serving as baselines. A WorldView-2 dataset was used for cross-domain testing.

The proposed method consistently improved detection performance across various models. Compared with Mosaic, the average AP_{50} across the model ensemble increased by 7.4% on the test set and 12.2% on the validation set, with AP_{50-95} achieving a maximum improvement of 19.1%. In transfer tasks on heterogeneous datasets, AP_{50} improved by 16.4%, outperforming both MixUp and DiffuseMix. Through environment-semantic augmentation, the method mitigated background variations across diverse natural environments and enhanced the model's capability to discern burial targets. Transfer experiments on WorldView-2 imagery further demonstrated that the approach improves model generalization and prevents excessive focus on background features.

This study presents an environment-semantic augmentation framework that integrates diffusion-based background simulation with fractal texture fusion for remote sensing archaeological detection. By enriching environmental diversity while preserving target structure, the proposed approach effectively addresses few-shot learning, environmental imbalance, and overfitting issues that commonly limit archaeological object detection models. Experimental results across multiple detection architectures and heterogeneous datasets demonstrate significant improvements in accuracy, recall, robustness, and cross-dataset generalization. The findings confirm that generative models, when properly guided by environmental semantics, can provide meaningful and controllable data diversity for small-sample scenarios.

Key words: deep learning, ancient tomb detection, data augmentation, diffusion model, remote sensing archaeology, few-shot learning, environmental semantic enhancement

Supported by National Key Research and Development Program of China (No. 2020YFC1521900); National Natural Science Foundation of China (No. 41801134)