引文格式:罗诗琦,罗斌,苏鑫,等.利用模拟数据进行遥感图像目标检测模型训练[J].武汉大学学报(信息科学版),2025,50(12):2569-2579.DOI:10.13203/j.whugis20230149

Citation:LUO Shiqi,LUO Bin,SU Xin,et al.Object Detection in Remote Sensing Images by Utilizing Synthetic Data[J].Geomatics and Information Science of Wuhan University,2025,50(12):2569-2579.DOI:10.13203/j.whugis20230149

利用模拟数据进行遥感图像目标检测模型训练

罗诗琦[1] 罗 斌[1] 苏 鑫[2] 张 婧[1] 刘 军[1]

1 武汉大学测绘遥感信息工程全国重点实验室,湖北 武汉,430079

2 武汉大学遥感信息工程学院,湖北 武汉,430079

摘 要:遥感图像目标检测技术应用广泛,但是训练数据获取受到天气、政策等限制且标注成本高昂,然而由计算机渲染得到的模拟数据生成速度快且成本低,因此将模拟数据应用于遥感图像目标检测模型训练。首先,基于《侠盗猎车手5》提出了自动获取图像及其标注的模拟数据采集系统,快速构建大规模遥感图像目标检测模拟样本集;然后,通过构建循环生成对抗网络将模拟遥感图像风格迁移至真实遥感图像;最后,在真实数据集UCAS-AOD和NWPU VHR-10上评估了所提方法的有效性。结果表明,风格迁移减小了模拟数据与真实数据之间的域差异,将迁移后的模拟样本集用于预训练可提升检测精度,在Faster R-CNN和YOLOv8上的平均精度最高分别提升了27.9%和18.5%。所提方法减小了遥感图像目标检测训练集的构建成本,可以为真实标注数据不足的场景提供有价值的解决方案。

关键词:模拟数据;迁移学习;目标检测;遥感图像;预训练

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-10-31

DOI:10.13203/j.whugis20230149 **文章编号**:1671-8860(2025)12-2569-11

Object Detection in Remote Sensing Images by Utilizing Synthetic Data

LUO Shiqi[1] *LUO Bin*[1] *SU Xin*[2] *ZHANG Jing*[1] *LIU Jun*[1]

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract: **Objectives:** Object detection in remote sensing images is widely used, but the cost of training set labeling is high and the image acquisition is limited by weather conditions and policies. However, synthetic data generated by computer rendering is fast and low-cost. Therefore, this paper proposes an object detection method in remote sensing images by utilizing synthetic data. **Methods:** First, a synthetic data collection system is developed based on Grand Theft Auto V (GTA5) to automatically obtain images and their annotations. And we construct a large-scale synthetic dataset named GTA5-Vehicle, which includes 29 657 instances of vehicles for object detection in remote sensing images. Then, the synthetic image style istransferred to real image style by constructing a cycle-consistent adversarial network while preserving their content. Finally, the effectiveness of the proposed method is evaluated on real datasets, including UCAS-AOD and NWPU VHR-10. To validate the generalization capability of the proposed method, both faster region-based convolutional neural network (Faster R-CNN) and YOLOv8 models are utilized in the object detection experiments. **Results:** The results demonstrate that style transfer can reduce the domain discrepancy between synthetic and real data, and yield an enhancement in detection accuracy by utilizing the transferred synthetic dataset for pre-training. Specifically, in the absence of real annotated data, the application of style transfer to the Faster R-CNN model yields an average precision improvement of 8.7% across both real datasets. By training the YOLOv8 model on transferred synthetic dataset, it also produces remarkable average precision scores, reaching 80.9% and 66.5% on the respective datasets with intersection

基金项目:国家科学自然基金(42230108)。

第一作者:罗诗琦,硕士,主要从事遥感图像目标检测等方面的研究。luoshiqi@whu.edu.cn

通信作者:刘军,博士,副研究员。liujunand@whu.edu.cn

over union set to 0.5. Moreover, when a limited amount of real annotated data is available, utilizing the transferred simulated sample set for pre-training can enhance the detection accuracy, resulting in the highest average precision improvements of 27.9% and 18.5% for Faster R-CNN and YOLOv8, respectively. **Conclusions**: The proposed method not only reduces the cost associated with constructing the training set for object detection in remote sensing images, but also offers a valuable solution for settings with limited real annotated data. The synthetic data and code are available at:https://lsq210.github.io/GTAVDataCollection/.

Key words: synthetic data; transfer learning; object detection; remote sensing image; pre-training

遥感图像目标检测技术在各个领域都发挥着重要作用,可以利用其进行交通流量监控[1]、自然灾害救援[2]、工程进度监测等,也能够完成对建筑物[3]、车辆[4]、人员[5]等目标的识别,对场景信息的迅速获取有着重要意义。

目前已经有许多有效的算法来解决遥感图像目标检测中的各种问题。传统方法基于人工设计的特征来检测目标。文献[6]提出了一种使用二级联线性模型和二进制线性规划来检测船舶的方法;文献[7]用迭代训练方法提高了不变Hough变换的适应性,以解决高分辨率遥感图像中近岸船舶的检测问题;文献[8]提出了一个由线性支持向量机(support vector machine,SVM)组成的框架以完成遥感图像中多类对象检测的任务。但人工设计的特征依靠主观经验,导致此类方法泛化能力受限。近年来,深度学习方法极大地提高了特征提取和描述的效率,促进了遥感图像目标检测技术的发展。基于区域的卷积神经网络(region-based convolutional neural network,R-CNN)被用于检测遥感图像中的各种目标[9-10],在检测精度方面取得了显著进展。此后,在R-CNN的基础上,Faster R-CNN[11]、Mask R-CNN[12]被提出,改进后应用于遥感图像目标检测任务。为进一步提升目标检测的速度,学者转向一阶段的目标检测网络,如文献[13]基于YOLOv2体系结构直接预测定向包围盒,用来检测遥感图像中任意方向的船舶;文献[14]利用改进的YOLOv3网络在具有不同背景干扰的遥感图像中检测高压电路绝缘体;文献[15]利用改进的YOLOv5模型,引入图像跨尺度特征融合与数据增强,增强了提取复杂目标特征的能力,并减少图像残缺导致的漏检。

然而,深度学习方法不仅依赖于高质量和大规模的训练数据集,而且要求待检测的图像与训练集分布尽可能相同。实际上,由于遥感图像数据来源不一,拍摄时的角度、高度差异较大,待检测的新图像往往与现有训练集存在较大的域差异。而重新构建数据集难度较大,首先,遥感图像的获取受到天气、政策等限制,部分特殊地区或军事目标的遥感图像甚至无法获取;其次,即使已有遥感图像,其复杂的背景、密集的小目标等特性导致标注过程耗时耗力、成本高昂。由此可见,数据成为了遥感图像目标检测领域的发展的一个重要限制因素[16-18]。

为解决特定场景的深度学习训练数据少、获取难、标注成本高这一问题,不少学者将注意力转向了模拟数据的研究。模拟数据由计算机生成,可以消除对人工搜集和标注数据的依赖,成本较低且解决了真实数据可能带来的隐私问题。根据模拟数据的生成方式可以将研究分为两类,第一类是通过融合目标和背景的方式得到合成图像[19-22]。其中,以文献[22]提出的新型遥感图像合成管道(remote sensing-image synthesis pipeline,RS-ISP)方法为例,该方法先通过复制粘贴操作生成粗糙的合成图像,然后使用协调方法来约束前景和背景的照明条件,最后生成用来训练遥感图像舰船检测模型的训练样本。此类方法能够有效地扩充遥感图像目标检测训练样本,从而提高检测精度。然而,局限性在于无法适用于复杂场景,正如现有方法大多是相对简单的海面或空地背景,一旦涉及到城市、道路等复杂场景,目标和背景融合时存在的语义和结构问题往往难以解决。第二类模拟数据生成的方法是基于游戏或三维渲染引擎构建虚拟三维场景。如文献[23]基于《侠盗猎车手5》(Grand Theft Auto V,GTA5)开发了一种快速获取真值的方法,成功构建大型的像素级标注来训练语义分割网络;文献[24]提出了一种结合模拟引擎内部缓冲区来获取逼真的计算机图像的方法;文献[25]基于虚幻引擎构建了一个大规模个人身份识别模拟数据集;文献[26]利用Unity引擎构建用于自主驾驶的模拟街道场景数据,并获得像素分割标签和深度图。这类方法主要集中在人群计数、自动驾驶等领域,但也有部分学者尝试将其应用在遥

感图像解译领域,并取得了一定成果。如文献[27]基于City引擎构建了一个用于深度估计的大规模城市地区航空图像模拟数据集SyntCities;文献[28]基于Unity引擎构建了一个用于遥感图像舰船目标检测的模拟数据集UnityShip。总体而言,这类方法支持的场景更为丰富,生成的模拟数据具有较好的图像质量和保真度。但由于图像都来自计算机渲染而非真实世界,因此需要克服模拟数据与真实数据之间存在的域差异。

生成对抗网络(generative adversarial network,GAN)[29]是减小两个数据集之间域差异的常用方法,但标准GAN在具体应用时存在训练不稳定等缺陷,于是便衍生出了一系列变体。GAN的变体在遥感图像解译领域也得到了广泛应用,可用于可见光与红外图像的转换[30],有积雪和无积雪场景的卫星图像的转换[31]、光学图像到数字表面模型的转换[32]以及遥感图像到地图的转换[33]等。这些应用充分展示了基于GAN的方法在模拟数据与真实数据之间建立桥梁的巨大潜力。

在上述研究的基础上,本文创新性地将模拟数据应用于遥感图像目标检测的研究。首先,提出了模拟数据自动化采集系统并开源,快速、低成本地构建了一个大规模遥感图像目标检测模拟样本集;然后,为克服模拟数据和真实数据之间的域差异,构建循环生成对抗网络(cycle generative adversarial network,CycleGAN)实现跨域图像之间的特征迁移;最后,在真实数据上进行验证。本文方法极大地降低了遥感图像目标检测训练集的构建成本,并在真实数据标注不足的场景中有效提升了检测精度,具有重要的工程应用价值。本文构建的模拟数据自动化采集系统和模拟样本集均已开源,可从以下网址获取:https://lsq210.github.io/GTAVDataCollection/。

1　模拟样本集构建

1.1　模拟数据自动化采集系统

模拟数据可以通过渲染三维对象模型或用程序编写自定义环境获取,但这样渲染虚拟世界和模型耗费大量时间和资源,且构建出的虚拟场景与现实差距较大。更优的方案是从具有高保真内容的现代开放世界计算机游戏中提取逼真的图像和数据。因此,本文选择了GTA5作为数据采集的来源,游戏提供了丰富且高度真实的虚拟世界,包含城市、郊外、公路等非常广阔的场景,且内置大量的车辆、飞机、船只等模型,支持各种视角,能够满足遥感视角目标检测的需求。

本文设计的模拟数据自动化采集系统使用C#进行开发,通过插件实现对GTA5原生接口的调用,结合其他游戏模组和OpenIV等编辑器,完成遥感视角下模拟数据的采集工作。系统调用了相机位姿、车辆密度、天气等接口,GTA5在系统设定下自动加载模型和地图,渲染出遥感视角下的模拟场景。图1为该系统获取的不同场景下的模拟图像。



(a) 不同天气下的模拟遥感图像

(b) 同一地点不同时间的模拟遥感图像

(c) 不同相机高度的模拟遥感图像

图1　不同场景下的模拟遥感图像

Fig. 1　Synthetic Remote Sensing Images in Different Scenes

在得到模拟遥感图像后,还需要获取目标的标签信息。本文以车辆目标为例对系统原理进行介绍。首先,通过调用原生接口获取当前模拟世界中全部的车辆。由于渲染模拟世界时,视线内看不见的模型也一并加载,本文使用射线检测方法将视线内被完全遮挡的车辆目标去除。然后,对所有车辆进行遍历,得到每辆车的类别、车辆模型在三维世界中的坐标最大值和最小值。根据这两个三维的点可以构造出一个三维的包围盒,再通过计算将三维包围盒的8个点全部转为二维图像的像素坐标。对坐标进行排序可得到目标车辆的矩形包围盒左上角坐标和右下角

坐标。经过以上步骤,将目标的类别、位置信息和图像的相关信息等一同存入标签文本文件。系统按照设定的时间间隔自动采集训练数据,以采集时刻的时间戳命名图像和标签,二者一一对应。每个标签第一行是所对应图像的长和宽,与使用的计算机屏幕分辨率有关;第二行是相机的位姿,第三行至最后一行是每个实例的标注信息。系统同时将标注信息绘制在模拟图像中,以供用户检查。含有车辆标注信息的模拟遥感图像如图2所示,其中红色矩形框代表的是小型车辆的包围盒,绿色矩形框代表的是大型车辆的包围盒,矩形框下方是对应目标的编号和类别。



图2 含有车辆标注信息的模拟遥感图像

Fig. 2 Synthetic Remote Sensing Image with Vehicle Annotations

本文提出的模拟数据自动化采集系统可支持自定义目标类别,可扩展至飞机、船只等目标检测。与其他生成模拟数据的方法相比,本文无需花费成本构建虚拟场景,即可得到丰富且逼真的虚拟世界,此外,调用原生接口得到目标标注的方式不额外占用计算机渲染资源,使得系统能持续稳定运行。理论上讲,一旦拥有足够的存储空间,使用该系统就可以获得任意数量、自定义场景的模拟图像,无需任何人工参与就同时得到了较为准确的标注。

1.2 遥感图像车辆目标检测模拟样本集

通过模拟数据自动化采集系统获取模拟遥感图像和车辆目标信息。在数据采集过程中,将相机设为俯视视角,时间设置为8时至18时,尽可能接近真实遥感图像生成条件。在具有车辆的场景中随机选择多个点,生成相机飞行轨迹。为保证车辆目标的广泛性和背景的多样性,本文选择了丰富的场景,包括城市道路、停车场、郊区、桥梁等,构建了面向遥感图像目标检测的大规模模拟样本集GTA5-Vehicle,共有1 520张图像,尺寸为2 048×1 125像素,其中准确标注了29 657个车辆实例,包括轿车、跑车、工业用车、摩托车等17个类别。更多关于数据的详细信息可通过本文所附的开源网址获得。

为了直观呈现模拟样本集的目标信息,本文统计了目标包围盒面积和目标密度,并与公开的真实遥感图像目标检测数据集UCAS-AOD进行对比分析,结果如图3、图4所示,其中目标密度是每张图上的目标个数。由图3可以看出,模拟样本集与真实样本集的目标包围盒面积分布很接近;由图4可以看出,两个数据集的目标密度大多集中在0~20之间。综合来看,GTA5-Vehicle在目标包围盒面积和目标密度的分布上与UCAS-AOD具有较高的相似度,但在数量上远超UCAS-AOD,因此可以作为真实数据有价值的补充。

表1列举了现有遥感图像车辆目标检测数据集与其他真实数据集的比较结果。由表1可见,本文所构建的GTA5-Vehicle数据集在图像、车辆样本和车辆类别等方面的数量都是最多的。

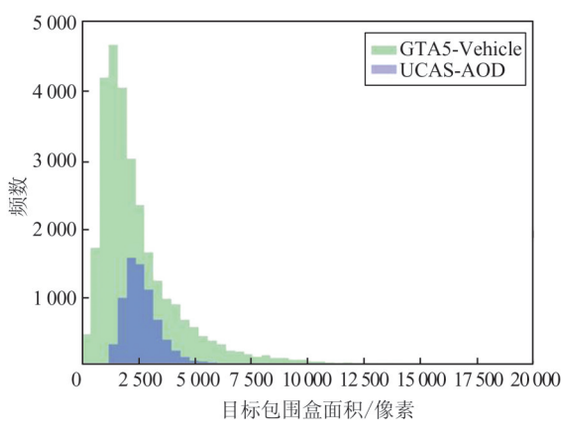


图3 目标包围盒面积直方图

Fig. 3 Histogram of Bounding Box Areas of Objects

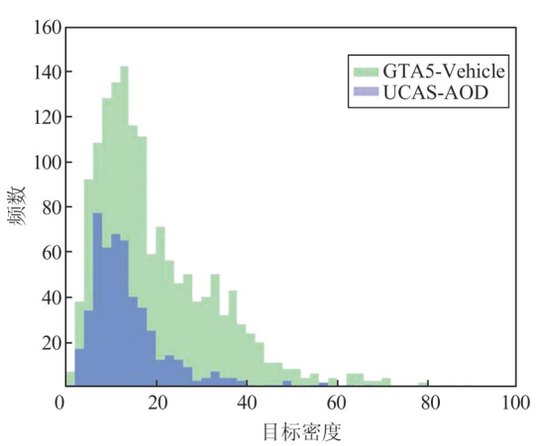


图4 目标密度直方图

Fig. 4 Histogram of Object Density

表 1　遥感图像车辆检测数据集对比

Table 1　Comparison of Vehicle Detection Datasets in Remote Sensing Images

数据集	图像数量/张	车辆样本/个	车辆类别	图片宽度/像素
UCAS-AOD[34]	510	7 114	1	~1 000
TAS[35]	30	1 319	1	792
NWPU VHR-10[36]	86	598	1	~1 000
VEDAI[37]	1 210	3 640	9	1 024
DLR 3K Vehicle[38]	20	14 235	2	5 616
GTA5-Vehicle	1 520	29 657	17	2 048

1.3　模拟遥感图像风格迁移

尽管在生成遥感图像目标检测模拟样本集时已经考虑到了成像时间、相机高度、背景等多个因素,但由于模拟遥感图像与真实遥感图像在色彩饱和度、纹理、清晰度等方面存在着不可避免的区别,模拟数据集和真实数据集仍存在一定的域差异。因此,为了进一步减小二者之间的域差异,本文构建了将模拟遥感图像风格迁移至真实遥感图像的CycleGAN。在图像内容不变的情况下,使用该网络将模拟遥感图像的风格转换成和真实遥感图像风格一致。

很多非遥感图像风格迁移方法已取得较好的成果[39],但无论是将照片转换成画作,还是生成艺术字,这些风格迁移方法注重的是艺术美感,同时也容易引入伪影和结构扭曲现象。本文研究目标是在风格迁移过程中尽可能保持图像内容的一致性,以确保迁移后的图像能够满足真实传感器所捕捉图像的色彩域要求,同时符合地物分布的合理性。鉴于此,在GAN基础上加入循环一致性损失[40],构建的CycleGAN如图5所示。

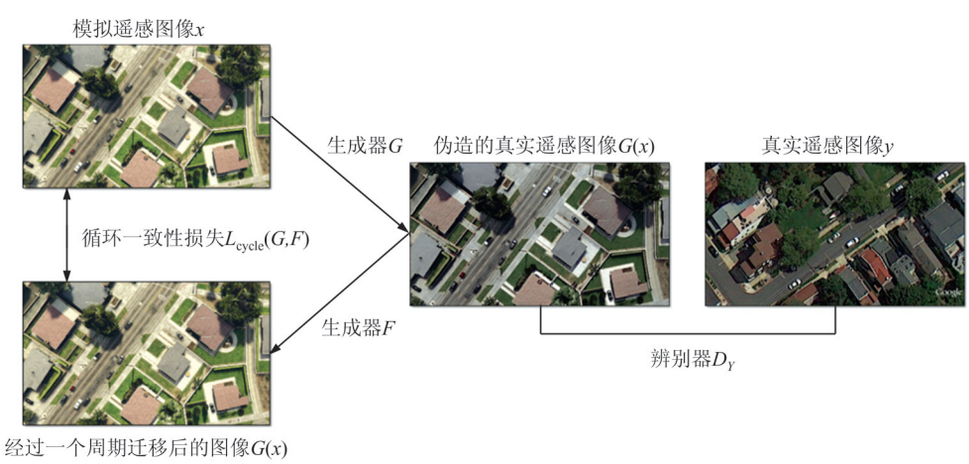


图 5　循环生成对抗网络框架

Fig. 5　Framework of Cycle Generative Adversarial Network

假设模拟遥感样本集和真实遥感样本集的样本空间分别是X和Y,给定了各自训练样本$\{x_i\}_{i=1}^{N}$和$\{y_j\}_{j=1}^{M}$,其中$x_i\in X, y_j\in Y$。目标是实现模拟遥感图像$\{x_i\}_{i=1}^{N}$到真实遥感图像$\{y_j\}_{j=1}^{M}$的风格迁移,即求得X到Y的映射。将映射关系称为生成器,可以构建两个生成器,G是一个将模拟遥感图像转化为真实遥感图像的生成器,F则是一个将真实遥感图像转化为模拟图像的生成器。将模拟遥感图像x输入生成器G后得到的图像称为伪造的真实遥感图像$G(x)$,将真实遥感图像y输入生成器F后得到的图像称为伪造的模拟遥感图像$F(y)$。同时,设置D_X和D_Y作为判别器,D_X用来辨别模拟遥感图像集$\{x_i\}$与生成网络输出的伪造的模拟遥感图像集$\{F(y)\}$,D_Y用来辨别真实遥感图像集$\{y_j\}$与伪造的真实遥感图像集$\{G(x)\}$。由此可以构造两个映射函数的对抗损失,以映射函数G和它的判别器D_Y为例,该部分的对抗损失L_{GAN}的计算式为:

$$L_{\mathrm{GAN}}(G,D_Y,X,Y)=E_{y\sim p_y(y)}\left[\lg D_Y(y)\right]+E_{x\sim p_x(x)}\left[\lg(1-D_Y(G(x)))\right] \tag{1}$$

式中,$E_{y\sim p_y(y)}\left[\lg D_Y(y)\right]$指的是从真实遥感图像样本集分布$p_y$中采样得到的样本$y$输入判别器$D_Y$后得到的结果,该值越接近1,意味着判别器$D_Y$越能够准确地识别真实遥感图像。对于

$E_{x\sim p_x(x)}\left[\lg(1-D_Y(G(x)))\right]$，$x$为从模拟遥感图像样本集分布$p_x$中采样得到的样本，$G(x)$为经过生成器$G$生成的伪造的真实遥感图像，那么$E_{x\sim p_x(x)}\left[\lg(1-D_Y(G(x)))\right]$的值越大，则$D_Y(G(x))$越接近0，意味着判别器$D_Y$越能准确地判断出这是个伪造样本，从而给出低分。

如上所述，判别器D_Y越能准确地区分真实遥感图像y和伪造的真实遥感图像$G(x)$，损失就越大，因此将其记为$\max_{D_Y}$；相反地，生成器G致力于使伪造的真实遥感图像$G(x)$和真实遥感图像y尽可能相似，导致判别器发生误判，即希望该损失最小，将其记为$\min_G$。该GAN训练的目标是$\min_G \max_{D_Y} L_{\mathrm{GAN}}(G, D_Y, X, Y)$。同样地，生成器$F$和它的判别器$D_X$的训练目标是$\min_F \max_{D_X} L_{\mathrm{GAN}}(F, D_X, Y, X)$。

然而，仅仅使用这一个损失是无法训练出有效的风格迁移模型，原因是有无数个生成器G都可以使$G(x)$达到与真实遥感图像y相近的分布。如G可以将全部的模拟遥感图像x都转化为样本空间Y中的同一张图片，也能达到骗过辨别器D_Y的效果。为此，本文引入循环一致性损失来对生成器进行约束，该损失L_{cycle}的计算式为：

$$L_{\mathrm{cycle}}(G,F)=E_{x\sim p_x(x)}\left[\|F(G(x))-x\|\right]+E_{y\sim p_y(y)}\left[\|G(F(y))-y\|\right] \tag{2}$$

式中，$F(G(x))$代表模拟遥感样本x经历了一个周期的迁移之后的图像，该图像与原始图像x越相同，那么$E_{x\sim p_x(x)}\left[\|F(G(x))-x\|\right]$越小。同理，$E_{y\sim p_y(y)}\left[\|G(F(y))-y\|\right]$代表真实遥感样本$y$经历了一个周期的迁移之后的图像$G(F(y))$与$y$之间的差异。

循环一致性损失的引入可以使得模拟遥感图像在迁移过程中保持内容不变，将对抗损失和循环一致性损失相加得到全部的损失，计算式为：

$$L(G,F,D_X,D_Y)=L_{\mathrm{GAN}}(G,D_Y,X,Y)+L_{\mathrm{GAN}}(F,D_X,Y,X)+\lambda L_{\mathrm{cycle}}(G,F) \tag{3}$$

式中，λ是控制两个损失权重的参数。

本文将模拟遥感样本集GTA5-Vehicle作为源域，真实遥感样本集UCAS-AOD作为目标域，输入CycleGAN，输出结果如图6所示。图像包含空地、树木、房屋、草地、水面等地物和车辆目标，从目视的角度看迁移后的遥感模拟图像在纹理、色彩、明暗等方面均更加接近真实遥感图像。

除了从主观的角度对迁移后的图像进行定性评价，一般非遥感图像风格迁移还会参考一些反映生成图像质量的客观指标，如峰值信噪比、结构相似性指数等。然而，本文风格迁移的目标是增强模拟遥感图像在实际遥感应用场景中的适用性。因此，本文提出在真实图像上进行目标检测实验结果来反映风格迁移效果，这种评估方法体现了遥感领域的独特性，提供了更有现实意义的结果。

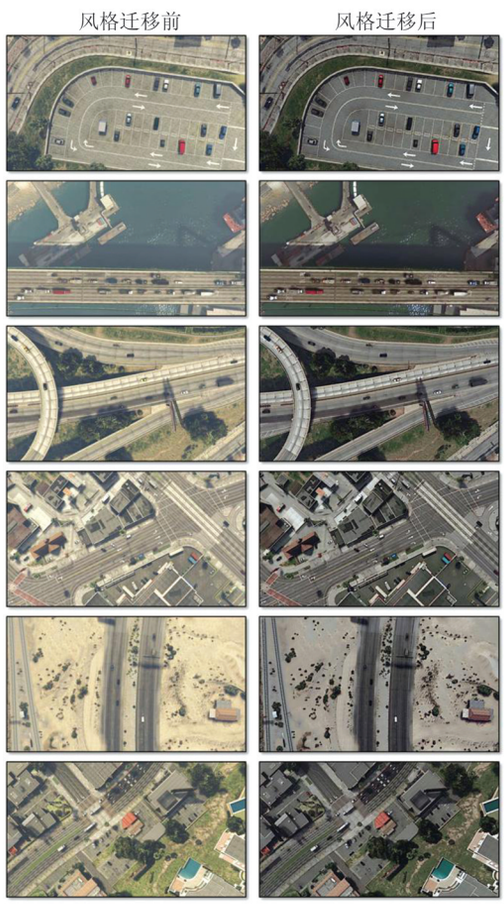


图6 将模拟图像迁移到真实图像的结果

Fig. 6 Results of Synthetic Images Transferred to Real Images

2 基于模拟样本集的遥感图像目标检测

2.1 实验数据和模型介绍

为了验证模拟数据的有效性和泛化性，本文选用了两个公开的真实遥感图像目标检测数据集进行实验，分别是UCAS-AOD和NWPU VHR-10。UCAS-AOD是由中国科学院大学构建的高清遥感图像集，实验从中选取了具有车辆目标的图像共510张，包含公路、停车场、居民区等多个场景。NWPU VHR-10数据集由西北工

业大学发布,共有86张具有车辆目标的图像,包含RGB图像和锐化彩色红外图像。实验选用的模拟数据来自模拟遥感图像车辆检测数据集GTA5-Vehicle,经过风格迁移后生成模拟遥感图像车辆检测数据集Tr-GTA5-Vehicle。由于两个真实数据集的标注里仅限于车辆这一大类目标,不具备更细粒度的子类目标,因此本实验中的模拟数据集也不再对子类目标进行区分。模拟数据集按8∶2随机划分成训练集和验证集,真实数据集按6∶2∶2随机划分成训练集、验证集和测试集。

现有目标检测网络大致划分为一阶段和二阶段两类。二阶段的目标检测网络需要先生成预选框,再进行细粒度的目标检测。而一阶段的目标检测网络则没有产生候选区域的阶段,直接生成目标的类别概率和位置信息。为验证本文方法的适用性,分别从一阶段和二阶段网络中各选一个作为目标检测实验的模型。结合遥感图像目标小而密集、存在噪声等干扰因素的特点,本文选择了基于特征金字塔的Faster R-CNN[11]和YOLOv8[41]进行实验。

2.2　基于模拟样本集的遥感图像目标检测实验

基于遥感图像目标检测实际应用中存在的样本难获取、标注成本高的情况,本文设计了两组实验:(1)在无真实数据标注场景下,分别使用模拟样本集GTA5-Vehicle和风格迁移后的模拟样本集Tr-GTA5-Vehicle训练目标检测网络,均在真实样本集的测试集上进行测试,对比测试结果从而验证风格迁移的有效性,并且评估无真实标注时只使用模拟数据能达到的目标检测效果;(2)在有少量真实数据标注场景下,利用模拟样本集训练出的网络作为预训练网络,引入不同数量的真实数据进行微调,并在真实样本的测试集上进行测试。将该测试结果与不使用模拟样本集作为预训练的测试结果进行对比,验证本文方法能否解决只有少量真实标注数据时的遥感图像目标检测问题。

2.3　结果评价指标

在目标检测领域,常用交并比(intersection over union,IoU)来评价预测框与真实框(ground truth,GT)之间的接近程度,其计算式为:

$$\mathrm{IoU}(B_i,B_{\mathrm{GT}})=\frac{B_i\cap B_{\mathrm{GT}i}}{B_i\cup B_{\mathrm{GT}i}} \tag{4}$$

式中,B_i为模型输出的预测框;$B_{\mathrm{GT}i}$为标注的真实框。对于单个目标样本,若IoU值大于设定的阈值,则说明该样本被正确检测。统计所有样本的检测结果,可计算出准确率P和召回率R,计算式分别为:

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}}\times 100\% \tag{5}$$

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}\times 100\% \tag{6}$$

式中,TP(true positive)为真正例,N_{TP}表示正确检测到的样本数量;FP(false positive)为假正例,N_{FP}表示误检的样本数量;FN(false negative)为假负例,N_{FN}表示缺失的真实样本数量。P-R曲线下的面积为平均精度(average precision,AP),R-IoU曲线下面积的两倍为平均召回率(average recall,AR),特定的IoU阈值为0.5时对应的平均精度,记作$\mathrm{AP}^{0.5}$,本文以AP作为主要评价指标,同时参考$\mathrm{AP}^{0.5}$和AR。

2.4　实验结果分析

无真实数据标注场景下的实验结果如表2所示。将使用GTA5-Vehicle和Tr-GTA5-Vehicle作为训练样本集的检测结果进行对比可以看出,风格迁移有效提升了目标检测模型的各项指标。对Faster R-CNN而言,风格迁移在两个真实数据集上的AP、$\mathrm{AP}^{0.5}$和AR分别平均提升了8.7%、17%和9.8%;在YOLOv8上,风格迁移后模型在两个真实测试集上的各项指标均提升了2%~5%。由此可见,对模拟数据集进行风格迁移处理有助于减小模拟数据和真实数据之间的差异。风格迁移对于YOLOv8的提升较Faster R-CNN更小,其原因可能是YOLOv8使用GTA5-Vehicle训练得到的模型已具有较好的目标检测能力,如在UCAS-AOD测试集上,IoU阈值设为0.5时,$\mathrm{AP}^{0.5}$达76.0%,召回率接近70%。实验表明,在没有任何真实数据标注的情况下,本文构建的模拟样本集能在真实遥感图像上进行目标检测,并且风格迁移后的效果更佳,YOLOv8在两个数据集上的$\mathrm{AP}^{0.5}$分别达80.9%和66.5%,AR分别达到72.7%和61.6%。

综上可知,风格迁移对目标检测各项指标都有所提升,因此本文实验直接将Tr-GTA5-Vehicle作为模拟样本。具体地,首先在Tr-GTA5-Vehicle上将训练好的模型作为预训练模型,再使用少量真实样本进行微调,最后在真实数据测试集上进行测试。为探究真实样本数量可能带来的影响,本文先从UCAS-AOD的训练集的306张图像中随机抽取20张、50张、100张图像进行实验,再从NWPU VHR-10的训练集的52张图像中随机抽取10张、20张图像进行实验。

表2 无真实数据标注场景的实验结果/%

Table 2 Experiment Results Without Real Data Annotations/%

测试集	训练集	Faster R-CNN			YOLOv8		
		AP	$AP^{0.5}$	AR	AP	$AP^{0.5}$	AR
UCAS-AOD	GTA5-Vehicle	17.8	49.0	25.9	29.2	76.0	69.2
	Tr-GTA5-Vehicle	26.1	68.2	34.3	32.7	80.9	72.7
NWPU VHR-10	GTA5-Vehicle	20.8	45.0	23.6	29.1	64.3	57.5
	Tr-GTA5-Vehicle	29.8	59.9	34.8	30.6	66.5	61.6
平均提升		8.7	17.0	9.8	2.5	3.6	3.8

在少量真实数据标注场景下,UCAS-AOD数据集和NWPU VHR-10数据集的实验结果如表3所示。由表3可以看出,相较于不使用模拟样本,使用模拟样本进行预训练的平均精度和平均召回率都更高。一方面,样本数量越小,模拟样本的提升效果越显著。以UCAS-AOD数据集为例,真实标注图像为20、50、100、306张时,基于模拟样本的方法使得Faster R-CNN的AP分别提升了27.5%、5.9%、1.8%、0.5%,同时YOLOv8的AP分别提升了18.5%、10.1%、7.4%、2.3%;另一方面,在只有极少真实标注图像的极端情况下,基于模拟样本的方法也能较好地完成遥感图像目标检测任务。在UCAS-AOD数据集上,仅使用20张真实样本即可使得Faster R-CNN和YOLOv8的$AP^{0.5}$分别达到94.4%和84.0%。在NWPU VHR-10数据集上,本文方法仅使用10张真实样本,Faster R-CNN和YOLOv8的$AP^{0.5}$均达到93.7%。

此外,由表3可知,有模拟样本时,使用20张真实标注图像微调得到的模型甚至优于无模拟样本时使用100张真实标注图像训练得到的模型。

表3 真实数据标注场景下UCAS-AOD和NWPU VHR-10测试集的实验结果/%

Table 3 Experiment Results of UCAS-AOD and NWPU VHR-10 Test Datasets with Real Data Annotations/%

测试集	真实标注图像数量	有无模拟数据	Faster R-CNN			YOLOv8		
			AP	$AP^{0.5}$	AR	AP	$AP^{0.5}$	AR
UCAS-AOD	20	无	22.7	65.4	35.3	23.6	51.0	16.4
		有	50.2	94.4	58.3	42.1	84.0	91.6
	50	无	47.3	91.8	55.3	37.7	75.6	59.5
		有	53.2	93.7	61.1	47.8	91.2	82.6
	100	无	52.2	93.6	60.0	41.9	82.9	72.3
		有	54.0	94.8	62.2	49.3	89.7	81.5
	306	无	57.6	97.1	64.8	57.9	97.5	92.8
		有	58.1	97.1	65.4	60.2	98.0	95.7
NWPU VHR-10	10	无	32.0	69.9	52.9	43.4	78.2	26.0
		有	59.5	93.7	67.9	58.8	93.7	90.4
	20	无	50.5	89.4	60.3	48.9	77.5	36.7
		有	62.2	92.9	69.5	63.9	93.9	90.4
	52	无	64.6	92.9	72.7	67.6	95.4	90.7
		有	68.7	95.2	75.2	70.8	97.0	93.2

上述结果证明了在没有或只有少部分真实数据的情况下,使用本文构建的模拟数据和迁移方法训练目标检测模型的可行性,极大地减小了获取和标注真实数据的成本,且接近甚至优于使用更多真实数据所训练出的模型。

结合检测的具体图像分析,若背景较为简单、实例个数少时,有无风格迁移都能将车辆检测出来。但是当车辆实例密度较多时,原始模拟数据存在很多漏检的现象,如图7所示。

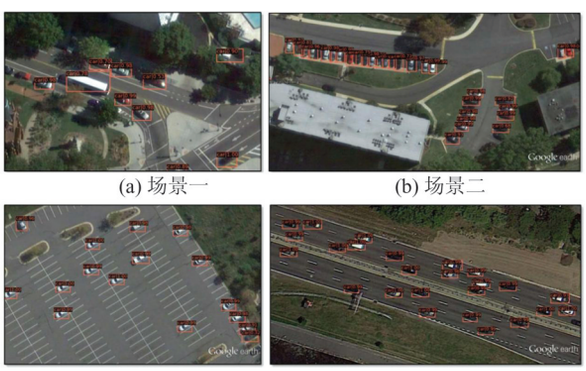
图8是基于模拟数据预训练后,只使用20张真实图像时的Faster R-CNN在UCAS-AOD测试集上的检测结果图。由图8可以看出,在所选的4种不同的场景中,车辆都能被检测出来,在车辆排列密集的图像上也不存在漏检情况,并且部分被树木遮挡的车辆也能被精准地识别出来。



(a) 仅使用模拟数据训练　　(b) 仅使用风格迁移后的模拟数据训练

图7　UCAS-AOD测试集上的检测结果

Fig. 7　Object Detection Results on UCAS-AOD Test Dataset



(a) 场景一　　(b) 场景二

(c) 场景三　　(d) 场景四

图8　本文方法在仅少量真实标注图像的目标检测结果

Fig. 8　Object Detection Results on Synthetic Data and Few Annotated Real Images

3　结　语

传统方式构建遥感图像目标检测训练样本成本高昂，甚至部分特殊地区或军事目标的遥感图像无法获取。因此，本文提出了一种利用模拟数据进行遥感图像目标检测的方法。首先，创新性地设计了模拟数据自动化采集系统，基于此构建了一个包含29 657个车辆实例的遥感图像目标检测模拟样本集；然后，通过构建CycleGAN将模拟遥感图像风格迁移至真实遥感图像；最后，分场景对本文构建的模拟数据集和方法的有效性进行评估分析。在无真实数据标注场景下，风格迁移能显著提高目标检测模型的精度，迁移后的模拟数据训练得到的YOLOv8在两个数据集上的$AP^{0.5}$分别达80.9%和66.5%；在有少量真实数据标注场景下，将模拟样本集作为预训练可大幅提升检测精度。有模拟样本时使用20张真实标注图像微调得到的模型甚至优于无模拟样本时使用100张真实标注图像训练得到的模型。实验证明本文方法在没有或少量真实标注数据的情况下，能够降低人工标注的成本，通过快速构建模拟样本集的方式对目标检测模型进行训练，从而有效提升目标检测精度，具有一定的现实意义。下一步的研究重点是扩充模拟样本集，并在更多的真实数据上进行测试。

参　考　文　献

[1] BENJDIRA B, KHURSHEED T, KOUBAA A, et al. Car Detection Using Unmanned Aerial Vehicles: Comparison Between Faster R-CNN and YOLOv3[C]//The 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 2019.

[2] 姜万冬，席江波，李振洪，等. 模拟困难样本的Mask R-CNN滑坡分割识别[J]. 武汉大学学报(信息科学版)，2023，48(12)：1931-1942.
JIANG Wandong, XI Jiangbo, LI Zhenhong, et al. Landslide Detection and Segmentation Using Mask R-CNN with Simulated Hard Samples[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(12): 1931-1942.

[3] 刘梓航，刘欣怡，张永军. 联合全局-局部特征和动态错误监督的遥感影像建筑物损伤评估方法[J]. 武汉大学学报(信息科学版)，2025，50(9)：1888-1903.
LIU Zihang, LIU Xinyi, ZHANG Yongjun. Building Damage Assessment from Satellite Images Combining Global-Local Features and Dynamic Error Supervision [J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(9): 1888-1903.

[4] BOUGUETTAYA A, ZARZOUR H, KECHIDA A, et al. Vehicle Detection from UAV Imagery with Deep Learning: A Review[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(11): 6047-6067.

[5] 谢亚坤，陈铭臻，赵耀纪，等. 顾及小目标特征的无人机影像人员检测方法[J]. 武汉大学学报(信息科学版)，2025，50(6)：1077-1087.
XIE Yakun, CHEN Mingzhen, ZHAO Yaoji, et al. A Person Detection Method for UAV Imagery Considering Small Object Features[J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(6): 1077-1087.

[6] LIU Z K, WANG H Z, WENG L B, et al. Ship Rotated Bounding Box Space for Ship Extraction from High-Resolution Optical Satellite Images with Complex Backgrounds [J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(8): 1074-1078.

[7] XU J, SUN X, ZHANG D B, et al. Automatic Detection of Inshore Ships in High-Resolution Remote Sensing Images Using Robust Invariant Generalized Hough Transform [J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(12): 2070-2074.

[8] CHENG G, HAN J W, ZHOU P C, et al. Multi-Class Geospatial Object Detection and Geographic Image Classification Based on Collection of Part Detectors[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 98: 119-132.

[9] LONG Y, GONG Y P, XIAO Z F, et al. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2486-2498.

[10] 胡佳睿, 潘俊, 赵和鹏, 等. 基于顾及上下文特征的耕地范围建筑物检测[J]. 测绘地理信息, 2025, 50(3): 121-124.
HU Jiarui, PAN Jun, ZHAO Hepeng, et al. Building Detection in the Arable Land Based on Contextual Features[J]. *Journal of Geomatics*, 2025, 50(3): 121-124.

[11] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.

[12] 张晓华, 李小龙, 艾金泉, 等. 复杂背景下基于改进Mask R-CNN的路面裂缝检测算法[J]. 北京测绘, 2024, 38(3): 431-436.
ZHANG Xiaohua, LI Xiaolong, AI Jinquan, et al. Road Crack Detection Algorithm Based on Improved Mask R-CNN in Complex Backgrounds[J]. *Beijing Surveying and Mapping*, 2024, 38(3): 431-436.

[13] LIU W C, MA L, CHEN H. Arbitrary-Oriented Ship Detection Framework in Optical Remote-Sensing Images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(6): 937-941.

[14] LIU C Y, WU Y Q, LIU J J, et al. Improved YOLOv3 Network for Insulator Detection in Aerial Images with Diverse Background Interference[J]. *Electronics*, 2021, 10(7): 771.

[15] 黄丽, 龚志鹏, 刘凡凡, 等. 图像跨尺度特征融合与数据增强的公交客流检测模型[J]. 武汉大学学报(信息科学版), 2024, 49(5): 700-708.
HUANG Li, GONG Zhipeng, LIU Fanfan, et al. Bus Passenger Flow Detection Model Based on Image Cross-Scale Feature Fusion and Data Augmentation [J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(5): 700-708.

[16] YU J, ZENG P, YU Y Y, et al. A Combined Convolutional Neural Network for Urban Land-Use Classification with GIS Data[J]. *Remote Sensing*, 2022, 14(5): 1128..

[17] HAN W, CHEN J, WANG L Z, et al. Methods for Small, Weak Object Detection in Optical High-Resolution Remote Sensing Images: A Survey of Advances and Challenges[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2021, 9(4): 8-34.

[18] 张磊, 张永生, 于英, 等. 遥感图像目标检测的数据增广研究[J]. 测绘科学技术学报, 2019, 36(5): 505-510.
ZHANG Lei, ZHANG Yongsheng, YU Ying, et al. Research on Data Augmentation for Object Detection of Remote Sensing Image [J]. *Journal of Geomatics Science and Technology*, 2019, 36(5): 505-510.

[19] LIN Y, SUZUKI K, TAKEDA H, et al. Generating Synthetic Training Data for Object Detection Using Multi-task Generative Adversarial Networks [J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, V-2-2020: 443-449.

[20] 杨志钢, 杨远兰, 苍思远, 等. 基于GAN的船舶遥感图像数据增广方法[J]. 应用科技, 2022, 49(5): 8-14.
YANG Zhigang, YANG Yuanlan, CANG Siyuan, et al. Data Augmentation Method of Ship Remote Sensing Images Based on GAN[J]. *Applied Science and Technology*, 2022, 49(5): 8-14.

[21] LIU W X, LUO B, LIU J. Synthetic Data Augmentation Using Multiscale Attention CycleGAN for Aircraft Detection in Remote Sensing Images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 4009205.

[22] ZHANG W C, ZHANG R, WANG G Q, et al. Physics Guided Remote Sensing Image Synthesis Network for Ship Detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4700814.

[23] RICHTER S R, VINEET V, ROTH S, et al. Playing for Data: Ground Truth from Computer Games [M]//Computer Vision — ECCV 2016. Cham: Springer, 2016: 102-118.

[24] JOHNSON-ROBERSON M, BARTO C, MEHTA R, et al. Driving in the Matrix: Can Virtual Worlds Replace Human-Generated Annotations for Real World Tasks? [C]//IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017.

[25] BĄK S, CARR P, LALONDE J F. Domain Adaptation Through Synthesis for Unsupervised Person Re-Identification[M]//Computer Vision — ECCV 2018. Cham: Springer, 2018: 193-209.

[26] ROS G, SELLART L, MATERZYNSKA J, et

al. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.

[27] REYES M F, D'ANGELO P, FRAUNDORFER F. SyntCities: A Large Synthetic Remote Sensing Dataset for Disparity Estimation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 10087-10098.

[28] HE B Y, LI X J, HUANG B, et al. UnityShip: A Large-Scale Synthetic Dataset for Ship Recognition in Aerial Images[J]. *Remote Sensing*, 2021, 13(24): 4999.

[29] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[J]. *Advances in Neural Information Processing Systems*, 2014, 27.

[30] 马得草, 鲜勇, 苏娟, 等. 基于改进的条件生成对抗网络的可见光红外图像转换算法[J]. 光子学报, 2023, 52(4): 0410003.
MA Decao, XIAN Yong, SU Juan, et al. Visible-to-Infrared Image Translation Based on an Improved Conditional Generative Adversarial Nets[J]. *Acta Photonica Sinica*, 2023, 52(4): 0410003.

[31] MOHAJERANI S, ASAD R, ABHISHEK K, et al. CloudMaskGAN: A Content-Aware Unpaired Image-to-Image Translation Algorithm for Remote Sensing Imagery[C]//IEEE International Conference on Image Processing (ICIP), Taipei, China, 2019.

[32] PAOLETTI M E, HAUT J M, GHAMISI P, et al. U-IMG2DSM: Unpaired Simulation of Digital Surface Models with Generative Adversarial Networks[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(7): 1288-1292.

[33] SONG J Q, LI J, CHEN H, et al. RSMT: A Remote Sensing Image-to-Map Translation Model *via* Adversarial Deep Transfer Learning[J]. *Remote Sensing*, 2022, 14(4): 919.

[34] ZHU H G, CHEN X G, DAI W Q, et al. Orientation Robust Object Detection in Aerial Images Using Deep Convolutional Neural Network[C]//IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 2015.

[35] HEITZ G, KOLLER D. Learning Spatial Context: Using Stuff to Find Things[M]//Computer Vision — ECCV 2008. Berlin, Heidelberg: Springer, 2008: 30-43.

[36] CHENG G, HAN J W. A Survey on Object Detection in Optical Remote Sensing Images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 117: 11-28.

[37] RAZAKARIVONY S, JURIE F. Vehicle Detection in Aerial Imagery: A Small Target Detection Benchmark[J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 187-203.

[38] LIU K, MATTYUS G. Fast Multiclass Vehicle Detection on Aerial Images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(9): 1938-1942.

[39] ZHONG L, LI N, SUN Y. Research and Application of Image Style Transfer Method[C]//IEEE 8th Intl Conference on Big Data Security on Cloud, High Performance and Smart Computing, (HPSC) and Intelligent Data and Security (IDS), Jinan, China, 2022.

[40] ZHU J Y, PARK T, ISOLA P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks [C]// IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[41] 赵铁锋. 一种轻型无人机遥感影像目标检测方法[J]. 北京测绘, 2025, 39(6): 913-920.
ZHAO Tiefeng. A Lightweight Object Detection Method Based on UAV Remote Sensing Images[J]. *Beijing Surveying and Mapping*, 2025, 39(6): 913-920.