



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860, CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目: 利用模拟数据进行遥感图像目标检测模型训练  
作者: 罗诗琦, 罗斌, 苏鑫, 张婧, 刘军  
DOI: 10.13203/j.whugis20230149  
收稿日期: 2023-10-31  
网络首发日期: 2023-12-01  
引用格式: 罗诗琦, 罗斌, 苏鑫, 张婧, 刘军. 利用模拟数据进行遥感图像目标检测模型训练[J/OL]. 武汉大学学报(信息科学版),  
<https://doi.org/10.13203/j.whugis20230149>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

Doi: 10.13203/j.whugis20230149

引用格式:

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), 2023, Doi:10.13203/j.whugis20230149. (LUO Shiqi, LUO Bin, SU Xin, et al. Leveraging Synthetic Data for Object Detection in Remote Sensing Images[J] *Geomatics and Information Science of Wuhan University*, 2023, 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<sup>1</sup> 罗斌<sup>1</sup> 苏鑫<sup>2</sup> 张婧<sup>1</sup> 刘军<sup>1</sup>

1 武汉大学测绘遥感信息工程国家重点实验室, 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, 本文提出将模拟数据应用于遥感图像目标检测模型训练。首先, 基于 GTA5 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 上评估了本文方法的有效性。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 RCNN 和 YOLOv8 上的平均精度最高提升了 27.9% 和 18.5%。证明本文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

## Leveraging Synthetic Data for Object Detection in Remote Sensing Images

LUO Shiqi<sup>1</sup> LUO Bin<sup>1</sup> SU Xin<sup>2</sup> ZHANG Jing<sup>1</sup> LIU Jun<sup>1</sup>

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 image acquisition is limited by weather conditions and policies. However, synthetic data generated by computer rendering is fast and low-cost. Therefore, we propose a method for object detection in remote sensing images by utilizing synthetic data. **Methods** Firstly, a synthetic data collection system was developed based on GTA5(Grand Theft Auto V) to automatically obtain images and their annotations. Using

收稿日期: 2023-10-31

项目资助: 国家自然科学基金(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

通讯作者: 刘军, 博士, 副研究员。liujunand@whu.edu.cn

this system, we construct a large-scale synthetic dataset named GTA5-Vehicle, which includes 29,657 instances of vehicles for remote sensing image object detection. Secondly, we transfer the synthetic image style to real image style by constructing a cycle-consistent adversarial network while preserving their content. Finally, the effectiveness of the proposed approach is evaluated on real datasets, namely UCAS-AOD and NWPU VHR-10. To validate the generalization capability of the approach, both Faster RCNN and YOLOv8 models are utilized for object detection experiments. **Results** The results demonstrate that style transfer reduces the domain discrepancy between synthetic and real data, and leveraging the transferred synthetic dataset for pre-training yields an enhancement in detection accuracy. Specifically, in the absence of real annotated data, the application of style transfer to the Faster RCNN model yields an average precision improvement of 8.7% across both real datasets. Training the YOLOv8 model on transferred synthetic dataset produces remarkable average precision scores, reaching 80.9% and 66.5% on the respective datasets with IoU set to 0.5. Moreover, when a limited amount of real annotated data is available, utilizing the transferred simulated sample set for pre-training enhances detection accuracy, resulting in the highest average precision improvements of 27.9% and 18.5% for Faster RCNN 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/>.

**Keywords:** 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。在民用领域,可以利用其进行交通流量监控<sup>[1]</sup>、自然灾害救援、工程进度监测<sup>[2]</sup>等。在军事领域,该技术能够完成对车辆<sup>[3]</sup>、舰船<sup>[4]</sup>、飞机<sup>[5]</sup>等军事目标的侦察,对战场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。Liu 等人<sup>[6]</sup>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; Xu 等<sup>[7]</sup>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; Cheng 等人<sup>[8]</sup>提出了一个由线性 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,导致这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。一些研究人员采用 R-CNN 来检测遥感图像中的各种目标<sup>[9-10]</sup>,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<sup>[11]</sup>被提出,改进后应用于遥感图像目标检测任务<sup>[12]</sup>。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,例如 Liu 等人<sup>[13]</sup>基于 YOLOv2<sup>[14]</sup>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; Liu 等人<sup>[15]</sup>利用改进的 YOLOv3 网络<sup>[16]</sup>在具有不同背景干扰的遥感图像中检测高压电路绝缘体。

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<sup>[17-18]</sup>。

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,成本极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<sup>[19-22]</sup>。其中,以 W.Zhang 等人<sup>[22]</sup>提出的新型遥感图像合成管道 (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。例如, S. R. Richter 等人<sup>[23]</sup>基于 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。Johnson-Roberson 等人<sup>[24]</sup>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。S. Bak 等人<sup>[25]</sup>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。G. Ros 等人<sup>[26]</sup>利用 Unity Engine 构建用于自动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。例如，M. F. Reyes 等人<sup>[27]</sup>基于 CityEngine 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。He 等人<sup>[28]</sup>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 (GAN)<sup>[29]</sup>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，它们被用于可见光与红外图像的转换<sup>[30]</sup>，有积雪和无积雪场景的卫星图像的转换<sup>[31]</sup>，光学图像到数字表面模型的转换<sup>[32]</sup>，以及遥感图像到地图的转换<sup>[33]</sup>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。

在这些研究的基础上，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。



(a) 不同天气下的模拟遥感图像

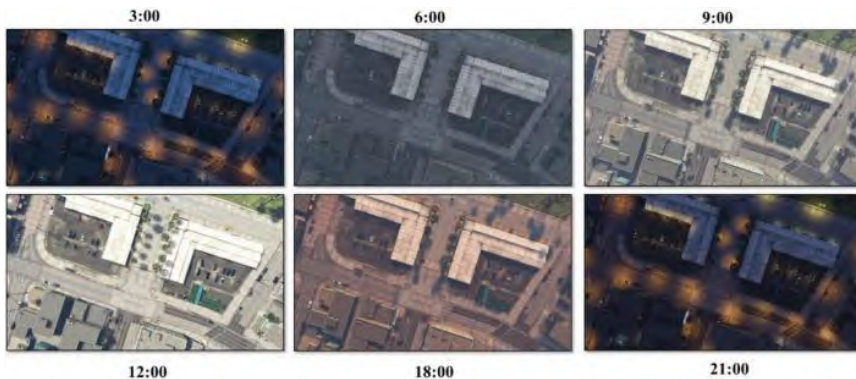
本文方法极大地减小了遥感图像目标检测训练集的构建成本，并在真实数据标注不足的场景中有效提升了检测精度，具有重要的工程应用价值。本文构建的模拟数据自动化采集系统和模拟样本集均已开源，可从以下网址获取：<https://lsq210.github.io/GTAVDataCollection/>。

## 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《侠盗猎车手 5》(Grand Theft Auto V, 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基于 .NET Framework 4.8，使用 C# 进行开发，通过 Script Hook V 1.0.2215.0 和 Script Hook V .NET 3.1.0 两个插件实现对 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。



(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 of 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  $\{(x_i, y_i)\}_{i=1}^8$ 。对  $x_i$  和  $y_i$  分别进行排序，便可得到目标车辆的矩形包围盒左上角坐标  $(x_{\max}, y_{\max})$  和右下角坐标  $(x_{\min}, y_{\min})$ 。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。图像的长宽由使用的计算机屏幕分辨率有关。每个标签第一行是所对应图像的长和宽，第二行是相机的位姿，第三行一直到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。如图 2 所示，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极为准确的标注。

## 1.2 遥感图像车辆目标检测模拟样本集

本文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，将相机设置成俯视视角，同时将时间设置为 8:00-18:00，以尽可能地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

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。最终, 本文构建了一个面向遥感图像目标检测的大规模模拟样本集 **GTA5-Vehicle**。该模拟样本集共有 1 520 张图像, 尺寸为  $2048 \times 1125$  像素, 其中准确地标注了 29 657 个车辆实例。如图 3 所示, 车辆目标包含轿车、跑车、肌肉车、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。

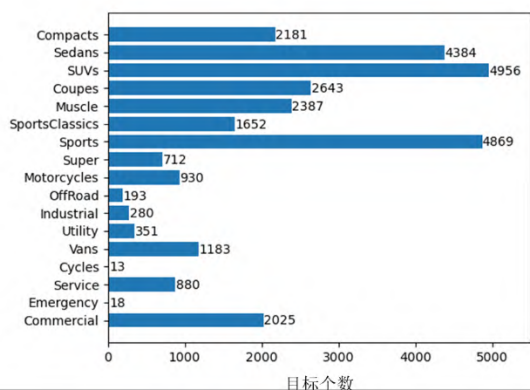


图 3 GTA5-Vehicle 车辆子类目标直方图

Fig.3 Histogram for Vehicle Subclasses in GTA5-Vehicle

为了直观地呈现模拟样本集目标信息, 本文对其目标包围盒面积和目标密度进行了统计, 并且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。图 4 绘制的是目标包围盒面积直方图, 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是很接近的。图 5 中, 目标密度指的是每张图上的目标个数, 两个数据集的目标密度大多都集中在  $[0, 20]$  区间。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。

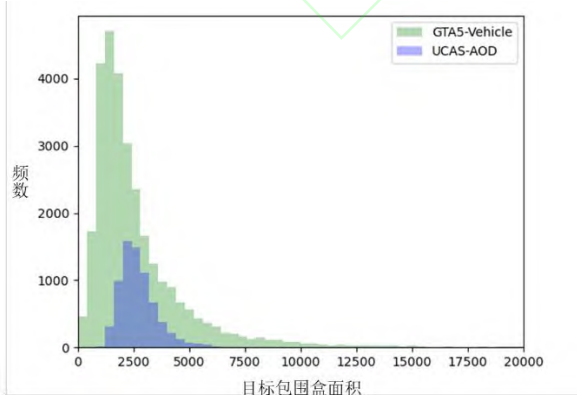


图 4 目标包围盒面积直方图

Fig.4 Histogram of Bounding Box Areas for Objects

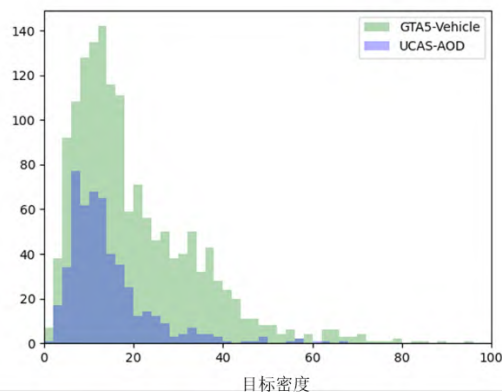


图 5 目标密度直方图

Fig.5 Histogram of Object Density

表 1 列举了现有遥感图像车辆目标检测数据集, 与表中其他真实数据集相比, 本文所构建的 **GTA5-Vehicle** 在图像、车辆实例数量、车辆类别方面都是最多的。

表 1 遥感图像车辆检测数据集对比

Tab.1 Comparison of Vehicle Detection Datasets in Remote Sensing Images

数据集	图像数量	车辆样本	车辆类别	图片宽度
UCAS-AOD(CAR) <sup>[34]</sup>	510	7114	1	~1000
TAS <sup>[35]</sup>	30	1319	1	792
NWPU VHR-10(CAR) <sup>[36]</sup>	86	598	1	~1000
VEDAI <sup>[37]</sup>	1210	3640	9	1024
DLR 3K Vehicle <sup>[38]</sup>	20	14235	2	5616
<b>GTA5-Vehicle</b>	<b>1520</b>	<b>29657</b>	<b>17</b>	<b>2048</b>

### 1.3 模拟遥感图像风格迁移

尽管我们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, 本文构建将模拟遥感图像风格迁移至真实遥感图像的循环生成对抗网络。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经取得了瞩目的成果<sup>[39]</sup>, 但无论是将照片转换画作, 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。而我们的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, 我们在生成对抗网络的基础上加入循环一致性损失<sup>[40]</sup>, 构建的循环生成对抗网络如图 6 所示。

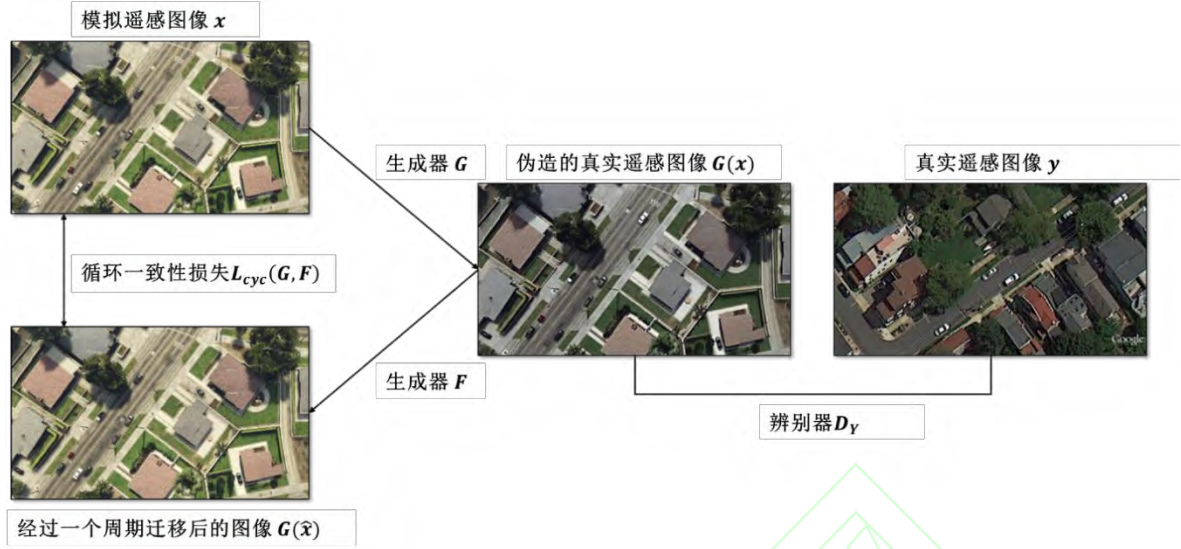


图 6 循环生成对抗网络框架

Fig.6 Cycle Generation Adversarial Network Framework

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 $Y$ ，并且给定了训练样本 $\{x_i\}_{i=1}^N$ 和 $\{y_i\}_{i=1}^M$ ，其中 $x_i \in X, y_i \in Y$ 。目标是实现模拟遥感图像 $\{x_i\}_{i=1}^N$ 到真实遥感图像 $\{y_i\}_{i=1}^M$ 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。经过生成器生成的图像统称为“伪造的”，例如将模拟遥感图像输入生成器后得到的图像称“伪造的模拟遥感图像”。与此同时，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 $\{F(y)\}$ ；同样地， $D_Y$ 用来辨别真实遥感图像集 $\{y_i\}$ 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，该部分的对抗损失计算公式为：

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_y(y)}[\log D_Y(y)] + E_{x \sim p_x(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

式中， $E_{y \sim p_y(y)}[\log D_Y(y)]$ 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 $D_Y$ 能越准确地将真实遥感图像识别为真实遥感图像。对于 $E_{x \sim p_x(x)}[\log(1 - D_Y(G(x)))]$ ， $x$ 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 $E_{x \sim p_x(x)}[\log(1 - D_Y(G(x)))]$ 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 $\min_G$ 。综合来看该生成对抗网络训练的目标就是 $\min_G \max_{D_Y} L_{GAN}(G, D_Y, X, Y)$ 。同样地，对于生成器 $F$ 和它的判别器 $D_X$ 的训练目标是 $\min_F \max_{D_X} L_{GAN}(F, D_X, Y, X)$ 。

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。比如， $G$ 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，也能达到骗过判别器 $D_Y$ 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，该损失定义为：

$$L_{cyc}(G, F) = E_{x \sim p_x(x)}[\|F(G(x)) - x\|] + E_{y \sim p_y(y)}[\|G(F(y)) - y\|] \quad (2)$$

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 $E_{x \sim p_x(x)}[\|F(G(x)) - x\|]$ 越小。同理， $E_{y \sim p_y(y)}[\|G(F(y)) - y\|]$ 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。最后，将对抗损失和循环一致性损失相加得到全部的损失，如式(3)所示：

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

式中， $\lambda$ 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，输入循环生成对抗网络。结果如图 7 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，例如峰值信噪比、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7 将模拟图像迁移到真实图像的结果  
Fig.7 Results on Transferring Synthetic Images 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 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，经过风格迁移后的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。模拟数据集按照 8: 2 随机划分成训练集和验证集，真实数据集按照 6: 2: 2 随机划分成训练集、验证集和测试集，将标签组织成 COCO 格式。

现有目标检测网络大致划分为一阶段（one-stage）和二阶段（two-stage）两类。二阶段的目标检测网络需要先生成预选框，然后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 RCNN 和 YOLOv8<sup>[41]</sup>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 GTA-V 和风格迁移后的模拟样本集 Tr-GTA-V 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，最后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，人们用  $IoU$  值来评价预测框与真实框之间的接近程度，其计算公式如下：

$$IoU(B_i, B_{GTi}) = \frac{B_i \cap B_{GTi}}{B_i \cup B_{GTi}} \quad (4)$$

式中  $B_i$  为模型输出的预测框， $B_{GTi}$  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地检测出来了。统计所有样本的检测结果，可计算出准确率  $P$  和召回率  $R$ ：

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$



式中  $TP$  为真正样本,  $TN$  为真负样本,  $FP$  为假正样本,  $FN$  为假负样本。Precision-Recall 曲线下的面积为平均精度  $AP$ ,  $Recall - IoU$  曲线下面积的两倍即为平均召回率  $AR$ , 特定的  $IoU$  阈值下对应的平均精度被记作  $AP^{IoU\_threshold}$ 。本文以  $AP$  作为主要评价指标, 同时参考  $AP^{0.5}$  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 和使用 Tr-GTA5-Vehicle 作为训练样本集的检测结果进行对比, 可以得知风格迁移有效提升了目标检测模型的各项指标。对 Faster RCNN 而言, 风格迁移在两个真实数据集上的  $AP$ 、 $AP^{0.5}$  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 的提升较于 Faster RCNN 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, 例如在 UCAS-AOD 测试集上,  $IoU$  阈值设为 0.5 时的平均精度达 76%, 召回率接近 70%。该组实验表明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  $AP^{0.5}$  分别达 80.9% 和 66.5%,  $AR$  分别达到 72.7% 和 61.6%。

表 2 无真实数据标注场景的实验结果

Fig. 2. Experiment Results Without Real Data Annotations

测试集	训练集	Faster-RCNN			YOLOv8		
		$AP$	$AP^{0.5}$	$AR$	$AP$	$AP^{0.5}$	$AR$
UCAS-AOD	GTA5-Vehicle	0.178	0.490	0.259	0.292	0.760	0.692
	Tr-GTA5-Vehicle	0.261	0.682	0.343	0.327	0.809	0.727
NWPU VHR-10	GTA5-Vehicle	0.208	0.450	0.236	0.291	0.643	0.575
	Tr-GTA5-Vehicle	0.298	0.599	0.348	0.306	0.665	0.616
平均提升		<b>0.087</b>	<b>0.170</b>	<b>0.098</b>	<b>0.025</b>	<b>0.036</b>	<b>0.038</b>

由上可知, 风格迁移对目标检测各项指标都有所提升, 因此本实验直接将 Tr-GTA5-Vehicle 作为模拟样本。具体地, 首先在 Tr-GTA5-Vehicle 上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 (306 张) 中随机抽取了 20 张、50 张、100 张图像进行实验, 再从 NWPU VHR-10 的训练集 (52 张) 中随机抽取了 10 张、20 张图像进行实验。有少量真实数据标注场景下的实验结果分别如表 3 和表 4 所示, 与不使用模拟样本相比, 使用模拟样本预训练的平均精度和平均召回率都更高。

一方面, 样本数量越小时, 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-RCNN 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 RCNN 和 YOLOv8 的  $AP^{IoU=0.5}$  分别达到 94.4% 和 84%。在 NWPU VHR-10 数据集上, 本方法仅使用 10 张真实样本时, Faster RCNN 和 YOLOv8 的  $AP^{IoU=0.5}$  达到 93.7%。

此外, 由表 3 可得, 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 有真实数据标注场景的 UCAS-AOD 测试集上的实验结果

Fig. 3 Experiment Results on UCAS-AOD Test Set with Real Data Annotations

测试集	真实标注 图像数量	有无模拟 数据	Faster-RCNN			YOLOv8		
			$AP$	$AP^{0.5}$	$AR$	$AP$	$AP^{0.5}$	$AR$
UCAS-AOD	20	无	0.227	0.654	0.353	0.236	0.510	0.164
		有	0.502	0.944	0.583	0.421	0.840	0.916
	50	无	0.473	0.918	0.553	0.377	0.756	0.595
		有	0.532	0.937	0.611	0.478	0.912	0.826

100	无	0.522	0.936	0.600	0.419	0.829	0.723
	有	0.540	0.948	0.622	0.493	0.897	0.815
306	无	0.576	0.971	0.648	0.579	0.975	0.928
	有	0.581	0.971	0.654	0.602	0.980	0.957

表 4 有真实数据标注场景的 NWPU VHR-10 测试集上的实验结果

Fig.4 Experiment Results on NWPU VHR-10 Test Set with Real Data Annotations

测试集	真实标注 图像数量	有无模拟 数据	Faster-RCNN			YOLOv8		
			AP	AP <sup>0.5</sup>	AR	AP	AP <sup>0.5</sup>	AR
NWPU VHR-10	10	无	0.320	0.699	0.529	0.434	0.782	0.260
		有	0.595	0.937	0.679	0.588	0.937	0.904
	20	无	0.505	0.894	0.603	0.489	0.775	0.367
		有	0.622	0.929	0.695	0.639	0.939	0.904
	52	无	0.646	0.929	0.727	0.676	0.954	0.907
		有	0.687	0.952	0.752	0.708	0.970	0.932

综上，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。既极大地减小了获取和标注真实数据的成本，又能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，实例个数少的时候，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的时候，如图 8

所示，原始模拟数据存在很多漏检的现象。图 9 是基于模拟数据预训练后，只使用 20 张真实图像时的 Faster-RCNN 在 UCAS-AOD 测试集上的检测结果图，从图中可以看到各种场景的车辆都能被检测出来，针对车辆排列密集的图像也不存在漏检情况，并且部分被树遮挡了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 8 UCAS-AOD 测试集上的检测结果

Fig. 8 Object Detection Results on UCAS-AOD



图9 本文方法在只有少量真实标注图像下的目标检测结果（由 Tr-GTA5-Vehilce 预训练和 20 张 UCAS-AOD 微调得到的 Faster R-CNN，在 UCAS-AOD 测试集上测试）

Fig.9 Object Detection Results on Synthetic data and Few Annotated Real Images (The Faster R-CNN model pre-trained with Tr-GTA5-Vehicle and fine-tuned using 20 UCAS-AOD images, tested on UCAS-AOD)

### 3 结论

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 29657 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，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  $AP^{IoU=0.5}$  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。实验证明了本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]. 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 2019.

[2] Xiao Liyang, Li Wei, Yuan Bo, et al. Pavement Crack Automatic Identification Method Based on Improved Mask R-CNN Model[J]. *Geomatics and Information Science of Wuhan University*. 2022: 1-16.(肖力场, 李伟, 袁博, 等. 一种改进型Mask R-CNN模型的路面裂缝识别方法[J]. 武汉大学学报(信息科学版). 2022: 1-16.)

[3] 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*. 2021.

[4] Song Zhina, Sui Haigang, Li Yongcheng. A Survey on Ship Detection Technology in High-Resolution Optical Remote Sensing Images[J]. *Geomatics and Information Science of Wuhan University*. 2021, 46(11): 1703-1715.(宋志娜, 眭海刚, 李永成. 高分辨率可见光遥感图像舰船目标检测综述[J]. 武汉大学学报(信息科学版). 2021, 46(11): 1703-1715.)

[5] Yu Donghang, Guo Haitao, Zhang Baoming, et al. Aircraft detection in remote sensing images using cascade convolutional neural networks[J]. *Acta Geodaetica et Cartographica Sinica*. 2019, 48(8): 1046-1058.(余东行, 郭海涛, 张保明, 等. 级联卷积神经网络的遥感影像飞机目标检测[J]. 测绘学报. 2019, 48(8): 1046-1058.)

- [6] Liu Z, Wang H, Weng L, 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, 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, Zhou P, 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] Salberg A. Detection of seals in remote sensing images using features extracted from deep convolutional neural networks[C]. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 2015.
- [10] Long Y, Gong Y, Xiao Z, 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.
- [11] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. *Advances in neural information processing systems*. 2015, 28.
- [12] Yin X, Yang Y, Xu H, et al. Enhanced faster-RCNN algorithm for object detection in aerial images[C]. 2020 IEEE 9th joint international information technology and artificial intelligence conference (ITAIC), Chongqing, China, 2020.
- [13] Liu W, 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] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
- [15] Liu C, Wu Y, Liu J, et al. Improved YOLOv3 network for insulator detection in aerial images with diverse background interference[J]. *Electronics*. 2021, 10(7): 771.
- [16] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. *arXiv preprint arXiv:1804.02767*. 2018.
- [17] Han W, Chen J, Wang L, 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] 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.(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.)
- [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] 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, 23.(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, 23.)
- [21] Liu W, 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: 1-5.
- [22] Zhang W, Zhang R, Wang G, et al. Physics Guided Remote Sensing Image Synthesis Network for Ship Detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*. 2023, 61: 1-14.
- [23] Richter S R, Vineet V, Roth S, et al. Playing for data: Ground truth from computer games[C]. Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016.
- [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]. 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, Singapore, 2017.
- [25] Bak S, Carr P, Lalonde J. Domain adaptation through synthesis for unsupervised person re-identification[C]. Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 2018.
- [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]. Proceedings of the IEEE conference on computer vision and pattern

recognition, Las Vegas, Nevada, 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, Li X, 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] 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): 232-246.(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): 232-246.)

[31] S. M, R. A, K. A, et al. Cloudmaskgan: A Content-Aware Unpaired Image-To-Image Translation Algorithm for Remote Sensing Imagery[C]. 2019 IEEE International Conference on Image Processing (ICIP), 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, 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, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 2015.

[35] Heitz G, Koller D. Learning spatial context: Using stuff to find things[C]. *Computer Vision - ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, 2008*.

[36] Cheng G, Han J. 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]. 2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2022.

[40] Zhu J, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. *Proceedings of the IEEE international conference on computer vision, Venice, Italy, 2017*.

[41] Glenn J. Ultralytics YOLOv8[DB/CD]. 2023.

#### 网络首发:

**标题:** 利用模拟数据进行遥感图像目标检测模型训练

**作者:** 罗诗琦, 罗斌, 苏鑫, 张婧, 刘军

**Doi:** 10.13203/j.whugis20230149

**收稿日期:** 2023-10-31

#### 引用格式:

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)*, 2023, Doi:10.13203/j.whugis20230149. (LUO Shiqi, LUO Bin, SU Xin, et al. Leveraging Synthetic Data for Object Detection in Remote Sensing Images[J] *Geomatics and Information Science of Wuhan University*, 2023, Doi:10.13203/j.whugis20230149.)