LOVD: Land Vehicle Detection in Complex Scenes of Optical Remote Sensing Image

Puti Yan, Xinrui Liu, Feng Wang, Chengfei Yue, *Member, IEEE*, and Xin Wang, *Member, IEEE*

***Abstract*— Nowadays, there is a growing body of research about object detection in remote sensing data. However, the detection algorithms for small targets in remote sensing areas are inadequate, largely because of the unavailability of high-quality datasets. Most remote sensing datasets are comprehensive, which means that they include bridges, airplanes, and lots of other common categories. Compared with other categories, the number and diversity of weak objects, such as vehicles, are quite insufficient. These limitations greatly affect the detection of small targets in remote sensing images. In order to promote the development of algorithms for the detection of small targets in remote sensing images and also allow access to remote sensing data, we have established a large-scale dataset for the detection of vehicle targets in optical remote sensing images and called it LOVD. It contains 1196 pictures and 541 751 instances, covering 13 categories. For the dataset, we have proposed in this article: 1) it is the largest one in terms of vehicle category and the total number of vehicle instances; 2) it contains images with various backgrounds in different weather and scenarios; and 3) all targets are marked by oriented bounding boxes (OBBs), and two label formats are provided. Finally, we test the state-of-the-art detection algorithms on our dataset and provide a benchmark for OBB detection.**

***Index Terms*— Dataset, remote sensing, vehicle detection.**

I. Introduction

TODAY, with the continuous development of sensors in the aerospace field and the continuous improvement of machine vision algorithms, processing of remote sensing images has gradually focused on the fields of city monitoring [1]–[3], land surveying (including resource exploration and agriculture application) [4]–[12], and environmental-climate monitoring [13]–[18]. Applications in these fields cover remote sensing images' classification [19]–[22], segmentation [23]–[29], object detection [30], [31], and so on.

Considering the lack of application of remote sensing data in land transportation, we have established a remote sensing dataset for vehicle detection to assist research on smart city transportation. Our dataset mainly uses aviation equipment to perform optical imaging of various types and sizes of vehicles in different weather conditions. The shooting scenes include cities, villages, lakes, and so on. Since the proposed dataset is a large-scale optical remote sensing dataset for vehicle detection, we refer to it as LOVD. Our dataset mainly makes up for the following neglected aspects of existing general remote sensing datasets.

1) Although some remote sensing datasets have brought great benefits [32], studies of datasets for the detection of small objects in remote sensing images have been ignored, especially those designed to detect fuzzy and tiny targets (such as cars). The number of such remote sensing datasets is still insufficient.
2) The diversity of scenes in most public vehicle datasets is relatively poor. Most existing vehicle datasets are concentrated on urban roads with a simple background, which easily leads to overfitting and the phenomenon of the case-specific detection algorithm.
3) All existing datasets adopt a rough classification of vehicle types. Most of them are divided into "large vehicles" and "small vehicles" according to their size. However, this classification method is no longer feasible since vehicles of the same size actually have different everyday applications. Therefore, there is a need for a more detailed classification.
4) Furthermore, horizontal bounding boxes (HBBs) are common in existing remote sensing datasets. Only a few datasets have oriented bounding box (OBB) annotations. We found that HBB cannot effectively describe the boundaries of small targets with large aspect ratios and in dense situations. Thus, it is necessary to apply OBB annotations in our dataset.

Large high-quality datasets can greatly improve data-driven algorithms, but the method of obtaining remote sensing data can be quite complicated. Consequently, the proposed dataset also considers the method of obtaining remote sensing data. Most remote sensing datasets are collected in the following ways: 1) images photographed by satellites; 2) data derived from various open-source remote sensing maps, such as Google Earth; and 3) images captured by drones, planes, and high-altitude balloons.

Each of the three methods outlined above has its pros and cons. Most remote sensing datasets adopt the first two methods. However, in those cases, it is very hard to collect pictures with different scenarios, such as rain or fog. Therefore, for our dataset, we adopted the third method, which involves

Manuscript received June 2, 2021; revised October 26, 2021; accepted November 21, 2021. Date of publication December 6, 2021; date of current version February 25, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61833009, Grant 11972130, and Grant 61690212. *(Corresponding authors: Chengfei Yue; Xin Wang.)*

Puti Yan, Xinrui Liu, and Feng Wang are with the Department of Aerospace Engineering, Harbin Institute of Technology, Harbin 150001, China.

Chengfei Yue is with the Institute of Space Science and Applied Technology (ISSAT), Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China (e-mail: yuechengfei@hit.edu.cn).

Xin Wang is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: xin_wang@tsinghua.edu.cn).

LOVD is available at https://loceyi.github.io/LOVD.github.io/dataset.html.

Digital Object Identifier 10.1109/TGRS.2021.3133109

documenting and taking photographs of different places by different aerial photography equipment, such as drones and helicopters. In order to obtain remote sensing data of multiple scenes and multiple weather conditions as much as possible, our team took about 1200 aerial images of multiple regions in China under different weather conditions. The size of all aerial images is 5280 × 3956 pixels. The dataset contains images of vehicles with different scales and orientations. All images have been identified, classified, and annotated carefully.

Compared with other remote sensing datasets for vehicle detection, our dataset has the following advantages.

1) The number of vehicle categories and the number of instances are the largest so far.

2) The scenes have great diversity. Its background covers cities, towns, rural areas, lakes, and other different terrain backgrounds. It also contains images taken in different weather conditions.

3) Our dataset adopts OBBs to label vehicles. OBB lays the foundation for vehicle detection and direction detection in a dense scenario, which is beneficial for the training of detection algorithms in large-scale vehicle scenarios, such as parking lots and urban arterial roads. Our main contributions can be summarized as follows.

1) Creating a large-scale vehicle benchmark dataset. This article proposes a large-scale, publicly available dataset for vehicle detection in optical remote sensing images. As far as we know, the proposed dataset has attained the largest scale in both vehicle categories and the number of instances. It allows for the verification and development of data-driven object detection algorithms.
2) Performance benchmarking on the proposed dataset; training and testing popular existing object detection algorithms on our dataset to determine their performance level.

II. Related Work

After investigation and analysis, it is found that there are currently two mainstream research directions for objects detection in the remote sensing area: one is to improve the quality and size of the dataset; the other is to optimize the detection algorithms.

A. Data

At present, some researchers are interested in establishing remote sensing datasets, and many datasets for objects detection have been proposed. Moreover, compared with the detection of large targets, such as playgrounds and airplanes, research on the detection of smaller targets, such as vehicles, is still inadequate. Since datasets now play an important role in data-driven research, a large amount of data would help in the detection of small and fuzzy targets. First, we will introduce remote sensing datasets that contain vehicles. The corresponding information is shown in Table I.

Raytheon *et al.* developed the OIRDS dataset in 2009, which contains about 1000 annotated images and 1800 vehicles. However, since the dataset is quite old, it was not tested with other algorithms. The number of instances is also relatively small [33]. Gong Cheng created the VHR-10 dataset, which comprised a collection of 800 images for the detection of 12 different types of objects [34], and Kang Liu published the DLR 3K Vehicle dataset in 2015. Both of these datasets are still small scale in nature [35]. In addition, in 2015, VEDAI and UCAS-AOD datasets were published. VEDAI artificially selected 1210 pictures from the Utah AGRC satellite library, which covered nine categories. However, in order to simplify the problem, the author excluded scenes with dense vehicles (such as parking lots) and scenes shot at oblique angles and mainly focused on relatively independent vehicle distribution and simple background image scenes [36]. Thus, this dataset is not suitable for vehicle detection training in urban scenarios, which is much more complicated. The UCAS-AOD dataset [37] collects images from Google Earth. It includes two types of targets: vehicles and airplanes. Among them, there are 2819 vehicles (310 pictures) and 3210 airplanes (600 pictures). The number of targets in the database is still too small, and the target category is largely single. Gradually, other larger datasets have been developed. COWC [38] and ITCVD [39] have a similar number of instances, but the former uses the center point as labels, and the latter uses HBBs. DOTA [32] and DIOR (2018) [31] datasets are also large and comprehensive remote sensing datasets. However, they only have one or two types of vehicles and contain many pictures without vehicles. None of them can meet the requirements of land vehicle detection. Gao *et al.* [40] established the RSOC dataset specifically for target counting in remote sensing images. The targets are divided into four categories: buildings, large vehicles, small vehicles, and ships. Although the dataset has a large number of vehicles, the total number of images is small, its annotation information is not sufficiently detailed, and it is not a public dataset. After doing a thorough survey, it is found that an excellent remote sensing dataset needs to have the following attributes.

1) *Effective Objects and Correct Annotations:* Some instances might be difficult to identify due to low resolution or being obscured by obstacles, such as trees and buildings. If such an object is not correctly labeled or classified, it will greatly affect the training performance of the algorithm. Therefore, effective screening of data is an important requirement for remote sensing datasets.
2) *Diverse Scene:* A high-level remote sensing dataset of vehicles should focus not just on arterial roads or parking lots, which have relatively simple backgrounds; rather, it should include scenes with more diverse backgrounds. Likewise, data collected under different weather conditions are also crucial since sunny days and other weather conditions usually affect the imaging of objects. If not, the dataset will have poor versatility especially for those scenes with complicated backgrounds. Thus, a high-quality remote sensing dataset should have diverse scenarios.
3) *Detailed Classification:* Both the size and application of vehicles should be taken into consideration as a classification standard. A vehicle remote sensing dataset should be classified according to the type, size, and

TABLE I
COMPARISON WITH OTHER DATASETS

Dataset	Scenario	Annotation	main categories	Instances	Images	Image width	Year
OIRDS	aerial	oriented bounding box	4	1800	900	256x256,512x512,640x480	2009
NWPU VHR-10	aerial	horizontal bounding box	10	3775	800	Non-fixed size, around 600	2014
DLR 3K Vehicle	aerial	oriented bounding box	2	14235	20	5616 × 3744	2015
VEDAI	satellite	oriented bounding box	9	3640	1210	1024 × 1024	2015
UCAS-AOD	Google	horizontal bounding box	2	6029	910	1280 × 659	2015
COWC	aerial	center point	1	32716	53	2000 to 19000	2016
DOTA	aerial	oriented bounding box	15	188282	2806	800 to 4000	2017
ITCVD	aerial	horizontal bounding box	1	29088	173	5616 × 3744	2018
DIOR	aerial	horizontal bounding box	20	190288	23464	800 × 800	2018
RSOC	Unknown	center point and HBB	5	286539	3057	512 to 2688	2020
Our dataset	aerial	oriented bounding box	13	541751	1196	5280 × 3956	2021

purpose of vehicles, and the classification should be as detailed as possible.

4) A large number of instances for each category. The amount of data determines the performance of an algorithm to a great extent. It is worth noting that what matters is not just the total amount of data but also the quantity and quality of data within each category.

As has been discussed above, there is currently no public dataset, which meets the requirements of future research for training vehicle detection algorithms on remote sensing images. Therefore, the proposed dataset will be extremely significant for research on the detection of small and blurred targets, such as vehicles, on remote sensing images.

B. Algorithms

Nowadays, most remote sensing target detection algorithms are based on neural networks originating from target detection algorithms for natural images. Target-based detection can be divided into two categories: "two-stage detection" and "one-stage detection." The former defines the detection bounding box as a screening process, while the latter defines it as an "end-to-end one-time completion" event. Furthermore, target detection algorithms based on deep learning are started from the two-stage RCNN. It first extracts a set of object candidate bounding boxes through selective search and then rescaled and input them to a CNN model trained on ImageNet to extract features. Finally, the linear SVM classifier is used to predict the target in each region and identify the target category. However, it brings the problem of low detection speed. To solve the problem of speed, Ren *et al.* proposed the Faster R-CNN detector in 2017. The main contribution of Faster R-CNN was the introduction of a region proposal network (RPN), which makes an almost cost-free region proposal possible. By transiting from RCNN to Faster RCNN, most of the independent blocks in a target detection system, such as proposal detection, feature extraction, and bounding box regression, have been gradually integrated into a unified end-to-end learning framework. In 2017, T.-Y. Lin *et al.* proposed a feature pyramid network (FPN) based on Faster RCNN [41]. Before the development of FPN, most detectors based on deep learning only performed detection at the top layer of the network. Although the deeper features of CNN are conducive for classification and detection, they are not conducive for object positioning. To this end, a top-down architecture design with horizontal connections was developed to build high-level semantics at all levels. Since CNN naturally forms a feature pyramid through its forward propagation, FPN has shown great progress in detecting targets of various scales. The development of FPN greatly improved the detection of remote sensing targets.

You only look once (YOLO) was proposed by Redmon *et al.* [42] in 2016. It is the first single-stage detector based on deep learning, and it is quite fast. As can be seen from its name, the author completely abandoned the previous "proposal detection + verification" detection paradigm. Instead, YOLO follows a completely different scheme: applying a single neural network to the entire image. The network divides the image into multiple regions and predicts the bounding box and probability of each region, concurrently. Later, R. Joseph made a series of improvements on the basis of YOLO, which further improved detection accuracy while maintaining a high detection speed.

Compared with the two-stage detector, YOLO's detection speed has been greatly improved, but its positioning accuracy has declined, particularly for some small targets. The follow-up version of YOLO and the SSD proposed by Liu *et al.* in 2016 [43] pays more attention to this problem. This was the second single-stage detector in the deep learning era. The main contribution of SSD was the introduction of multireference and multiresolution detection technology, which greatly improved the detection accuracy of a single-stage detector, especially for some small targets. SSD has advantages in detection speed and accuracy, and the fast version runs at 59 fps. The main difference between SSD and any previous detector is that the former detects objects of different scales at different layers of the network, while the latter only performs detection on the top layer. Therefore, compared with YOLOv3, SSD is better at remote sensing data detection though the effect is slightly weaker.

Also, in the field of remote sensing, there are also many specific detection and classification algorithms that have been proposed. Zou *et al.* [44] applied deep learning to the classification of remote sensing images and proposed a modified DBN network. In 2016, Cheng *et al.* proposed a method to learn a rotation-invariant CNN (RICNN) detection model, which introduced a new objective function and achieve great performance improvement in detection of remote sensing images [45]. In 2017, Sharma *et al.* [46] proposed a deep patch-based CNN system for the classification in the remote sensing image. In 2018, Cheng *et al.* [47] proposed a method for D-CNN classification model training by imposing a metric learning regularization. They made great progress on the classification of remote sensing images. In 2021, Cheng *et al.* [48] raised FENet for object detection in remote sensing images by applying two feature enhancement modules.

However, most of the abovementioned mainstream algorithms use HBB detection, and target detection algorithms using OBBs are still rare. This is because it is very difficult to locate and separate multiangle objects from the background accurately and quickly. Due to the complexity of remote sensing image scenes and a large number of small, messy and oriented targets, two-stage rotated target detectors, such as ROI transformer and SCRDet, are still the most robust choice.

In 2018, Ding and others proposed RoI (Region of Interest) Transformer [49]. The core idea of the RoI transformer is to apply the spatial transformation on RoIs, with the parameters of spatial transformation being learned under the supervision of OBB labels. An RoI transformer is lightweight and can be easily embedded in various rotating target detectors; thus, it is more suitable for transfer learning. In addition, in 2019, Xue Yang and others proposed SCRDet. SCRDet is optimized in terms of small targets, dense arrangement, and arbitrary rotation angles. For small targets: a feature fusion structure is designed from the perspective of feature fusion and anchor sampling. For the dense arrangement problem, a supervised multidimensional attention network is designed to reduce the adverse effects of background noise. For any direction problem: an improved smooth L1 loss is designed by adding an IoU constant factor, which is specifically used to solve the boundary problem of OBB regression.

Today, the best target detection algorithm based on OBB is the R3det dataset proposed by Yang *et al.* [50] in 2019. They designed a feature refinement module that can obtain more accurate features to improve the detection performance of rotating targets. The key idea of the feature refinement module is to reencode the position information of the current refined bounding box into corresponding feature points through feature interpolation to achieve feature reconstruction and alignment. In 2020, Yang and others proposed the CSL method [51]. In the article, they argued that popular existing regression-based angle prediction methods have more or fewer boundary problems. One of the main reasons is that the ideal prediction results exceed ours. The defined range leads to a larger loss value. Therefore, CSL eliminates this problem by converting the angle regression problem into a classification problem and restricting the range of the prediction results. Although these algorithms can identify remote sensing targets better, they still suffer from problems such as extreme imbalance between classes in the dataset or too many angle categories in the joystick target, which weakens the detection effect.

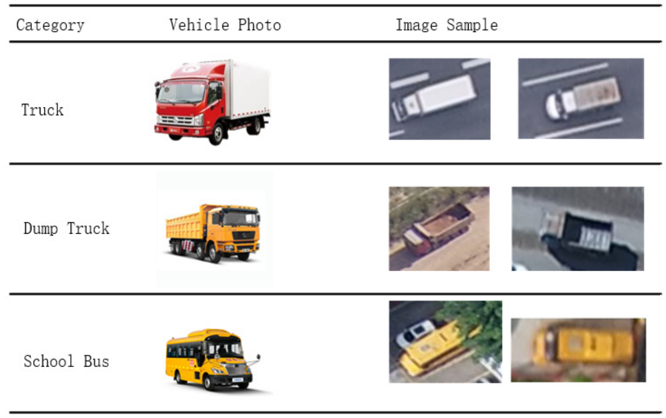

Fig. 1. Similar vehicles from our database.

After investigating the dataset and algorithm separately, it is found that those algorithms are designed based on existing comprehensive datasets, so they cannot effectively solve the problem of identifying weak and small targets, such as vehicles, in remote sensing data. Therefore, in order to solve this problem, we propose the LOVD dataset, whose advantages are given as follows.

1) *More Vehicle Categories:* We take vehicles' application and types as the main classification criteria for vehicles, which mainly includes public transportation, private cars, construction vehicles, and trucks. On the basis of these four categories, a detailed category version is also included.
2) *Sufficient Quantity of Each Category:* We try to enrich the data for each vehicle type, so as to train the algorithms to detect different types of vehicles.
3) *Varied Scenes and Weather Conditions:* Our dataset includes many complex scenes, such as those with dense vehicles and those under shadows. At the same time, the backgrounds of our dataset are as comprehensive as possible and include towns, cities, forests, lakes, and so on, which greatly enriches its diversity. Images under different weather conditions, such as rainy or foggy weather, are included as well.

III. Preparation

A. Images' Collection

Nowadays, most remote sensing images are collected by aircraft or remote sensing satellites. Our team relied on drones, helicopters, and other aerial photography equipment to collect and record data over a long period and, finally, compiled a large aerial dataset.

In order to ensure the legitimacy of the data, all data are collected in an allowable area. Moreover, in order to ensure the geographic diversity of the data, after each flight, the geographic coordinates of aerial images are recorded, and the area will not be revisited anymore. Our data collection task mostly focuses on areas surrounding Shenzhen City and Harbin City in China.

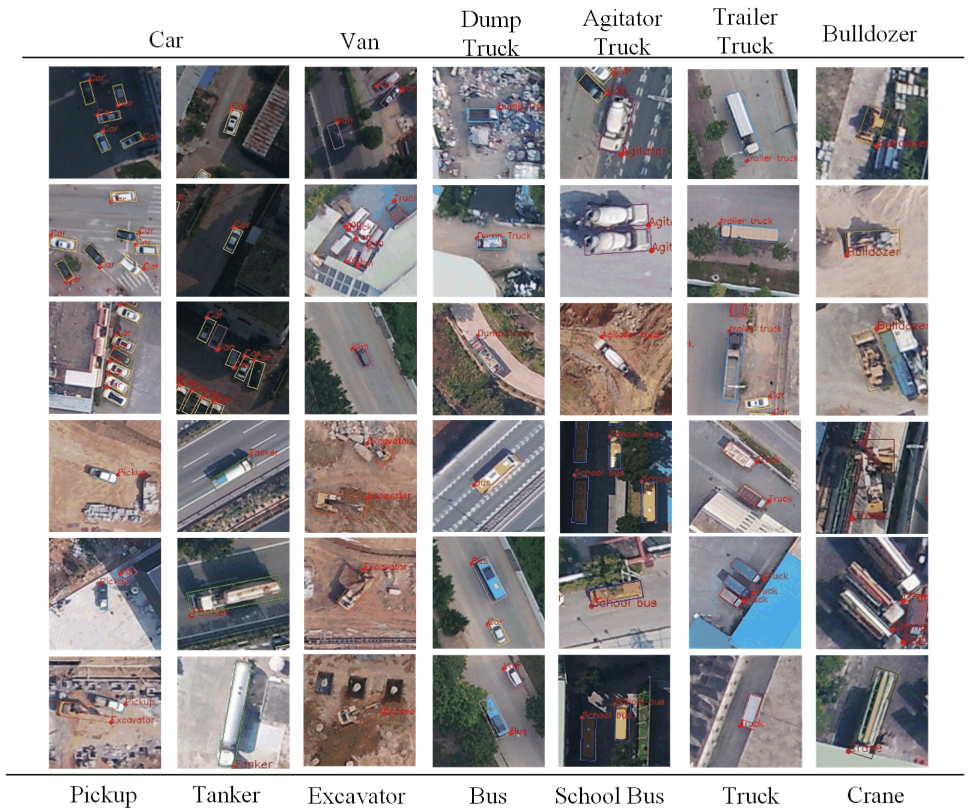

Fig. 2. Some vehicles from our database.

B. Category

After investigations, we found that most datasets select the size as the classification criteria for vehicles and tend to ignore other characteristics. Therefore, we chose not to adopt such a single standard. We first classified those vehicles into "large vehicles" and "small vehicles" (such as in DOTA [32]). However, as shown in Fig. 1, we found out that vehicles meant for completely different purposes had the same size. Therefore, we decided to distinguish and classify vehicles based on additional characteristics, such as color, contour, and purpose.

Ultimately, our dataset contains 13 types of vehicles in daily life, as shown in Fig. 2, which includes cars, vans, dump trucks, agitator trucks, trailers, bulldozers, pickup, tankers, excavators, buses, school buses, trucks, and cranes.

C. Scene Selection

When choosing scenes, we focused on conditions that are challenging for existing algorithms and tested popular algorithms, such as YOLO, Faster R-CNN, and Cascade R-CNN on the existing datasets, such as DOTA and DIOR. We found that those algorithms have poor performance in the following scenarios: 1) a dense vehicle area; 2) inconsistent directions of vehicles; 3) junction areas with different terrains; and 4) objects blocked by trees or under the shadow of buildings or covered by clouds and fog.

The areas in which our team collected data from lie approximately 22.32°N, 114.05°E and 44.04°N, 125.42°E, respectively. The area ranges from the coastline to the mainland. The climate and topography of those locations are diverse, which satisfies our requirement for a variety of scenes. In addition, to ensure our dataset contained highly dense vehicle scenes, we focused on collecting images of parking lots, arterial roads, and viaducts because there is a higher probability of dense scenes in these places.

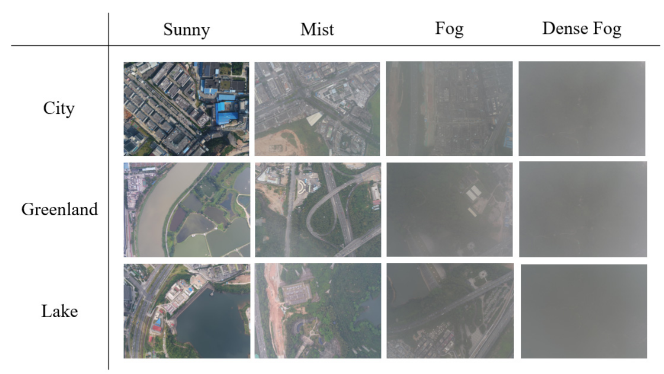

Fig. 3. Weather and background.

D. Weather Selection

Regarding the weather status of the dataset, our team tried to enrich it by including various weather types in the dataset, as shown in Fig. 3. The fog is classified in this dataset according to the degree of ambiguity. The dark channel

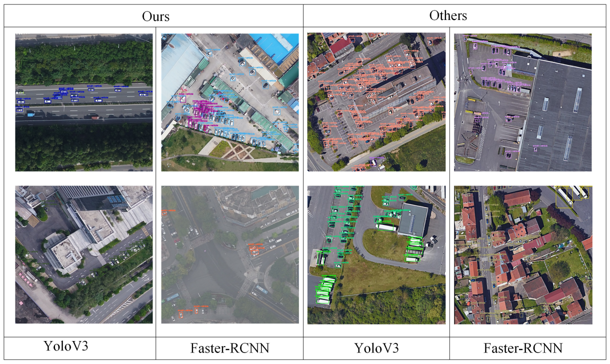

Fig. 4. Results of the horizontal bounding box.

theory [52] is used to calculate the transmittance of image models as described in the following equations:

$$J^{\text{Dark}}(x) = \min_{y\in\Omega(x)} \left(\min_{c\in r,g,b} J^c(y) \right) \tag{1}$$

$$J^{\text{Dark}}(x) \to 0 \tag{2}$$

$$I^c(x) = J^c(x)t(x) + A^c(1 - t(x)) \tag{3}$$

$$\hat{t} = 1 - \min_{y\in\Omega(x)} \left(\min_{c} \frac{I^c(y)}{A^c} \right). \tag{4}$$

J^{Dark} is the dark channel of the image and J^c is the color channel. A^c is the global atmospheric light, which can be considered as a constant. We can deduce transmittance $\hat{t}$ from (4) and utilize it as a criterion to divide images into four levels of fog: sunny, mist, fog, dense fog, and the four levels are marked as 0, 1, 2, and 3, as shown in the following equation:

$$\text{rank} = \begin{cases} 0, & 0.75 < \hat{t}(x) < 1 \\ 1, & 0.5 < \hat{t}(x) < 0.75 \\ 2, & 0.25 < \hat{t}(x) < 0.5 \\ 3, & 0 < \hat{t}(x) < 0.25. \end{cases} \tag{5}$$

E. Annotation Method

Nowadays, there are mainly two annotation methods: HBB and OBB. To evaluate the pros and cons of these two methods, YOLOv3 and Faster R-CNN with FPN were run on our dataset and several others, such as DOTA and DIOR. Based on the results shown in Fig. 4, we took note of the following issues.

1) At places such as parking lots and highways, the density of vehicles is generally high, and the directions are also much inconsistent. Therefore, applying HBBs will result in the loss of pixels for the target and the redundancy of background pixels.
2) The HBB is especially unsuitable to annotate targets with a large aspect ratio.

Consequently, OBBs were adopted to annotate the vehicles, as they can accurately capture the boundary of objects by quadrilateral area coordinates. The four-point coordinates are used to represent the vertices of the bounding box in the image. As shown in Fig. 5, these vertices are arranged in a clockwise direction, the first vertex is emphasized and named "red_point" in the annotation files, and it represents the front left corner of the vehicle. For convenience, we published the two types of

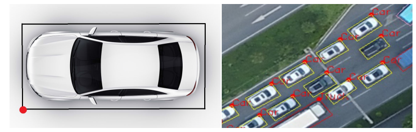

Fig. 5. Annotation demo.

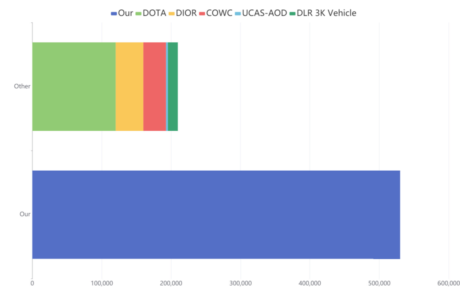

Fig. 6. Comparison of the instance number of vehicles.

annotation formats on Github: one is in XML format contains information about classification labels, center point, height, width, and rotation angle; the other is in TXT format with classification labels and four points' coordinates. In addition, considering that some vehicles might be tough to annotate due to occlusion or blur of images, a flag (named "Difficult") is used to indicate that difficulty. We will take advantage of this feature in our future research on hard-to-detect targets.

IV. Properties

A. Basic

Our dataset contains 1196 pictures, 13 categories, and a total of 541 751 instances. In Fig. 6, it can be seen that it far exceeds the sum of vehicle instances in other existing remote sensing datasets. The resolution of our dataset is about 0.12 m, and the original size of the images is 5280 * 3956 pixels. It is also larger than the size of mainstream remote sensing datasets. In order to keep information of this large number of instances, we did not crop the images when annotating.

B. Various Orientations of Vehicles

Since our dataset chooses to apply OBBs, the distribution of the directions of objects needs to be taken into consideration. We summarized the angles of vehicles in the dataset and visualized them through a histogram. As shown in Fig. 7, there are denser distributions at 0° (360°), 90°, 180°, and 270°. This is a normal phenomenon since most streets are aligned east–west or north–south. However, in general, the angles of vehicles in our dataset are still distributed evenly.

C. Various Categories of Vehicles

As shown in Table II, most existing datasets classify vehicles according to size, such as DOTA (divides vehicles into "large vehicles" and "small vehicles"), or as only one category, such as DIOR.

However, if the intelligent transportation service is used for the purpose of scientific research, it is evident that the

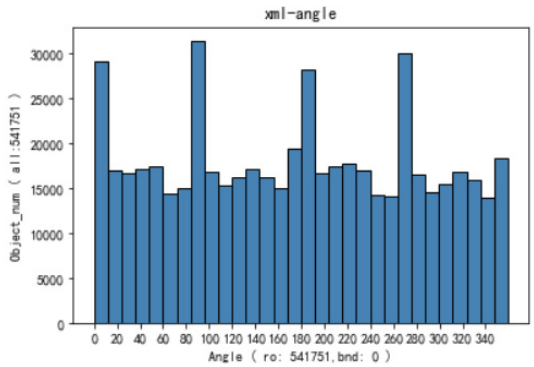

Fig. 7. Distribution of orientations.

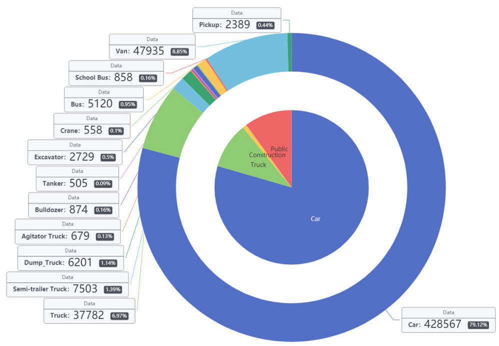

Fig. 8. Distribution of the vehicle type.

TABLE II
COMPARISON OF TOTAL VEHICLE CATEGORIES

Name	No. of vehicle targets	No. of vehicle type
OIRDS	1800	4
NWPU VHR-10	477	1
DLR 3K Vehicle	14235	2
VEDAI	3470	7
UCAS-AOD	2819	1
COWC	32716	1
DOTA	120000	2
ITCVD	29088	1
DIOR	40370	1
RSOC	165432	2

characteristic information of many vehicles is inappropriate if the size is adopted as the only criterion. For example, school buses and trucks are similar in size, but, because they actually perform different tasks, there are certain differences in both their outline and color.

Therefore, in accordance with the application of vehicles in real life, our dataset is divided into 14 categories. Since there are many similar types of vehicles in the dataset, in order to accurately annotate the types of vehicles in the figure, each type of car is strictly defined according to its unique characteristics. At the same time, we have performed multiple calibrations. For example, the tanker has an oval oil tank; the

TABLE III
COMPARISON OF TOTAL VEHICLE CATEGORIES

Name	No. of vehicle targets
Car	428567
Van	47935
Truck	37782
Trailer Truck	7503
Dump Truck	6201
Bus	5120
Excavator	2729
Pickup	2389
Bulldozer	874
School Bus	858
Agitator Truck	679
Crane	558
Tanker	505

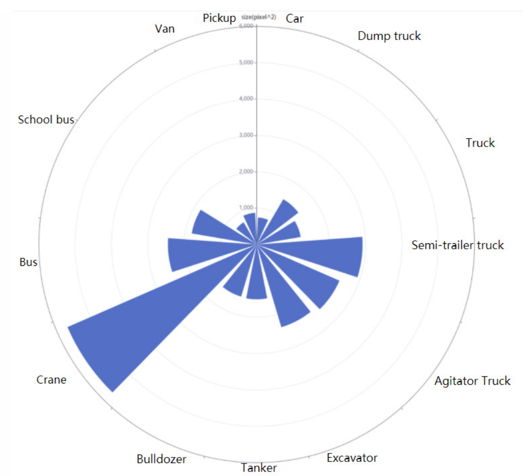

Fig. 9. Average pixel size of 13 different vehicles.

trailer truck has a very long square compartment; compared with the front of the vehicle, the van is a square and complete body; the truck is composed of the front and the compartment; and the school bus is in yellow. We divide them according to the most obvious difference. Although there may be some wrong classifications, we have reduced the error to a very small amount.

D. Statistics of Targets' Area

The pixel area is used as a measurement method of the size of the instance, and the area of the vehicle ranges from 600 to 6000 pixels. Based on the size of the instance area, the characteristics of each category, and the number of instances in each category, targets in the dataset are divided into four categories: cars, trucks, construction, and the public to improve the detection accuracy. For details, see Section V. From Figs. 9 and 10, the area range of targets of different sizes can be obtained. As shown in pictures, except for the Crane category, which has a larger area, other categories have

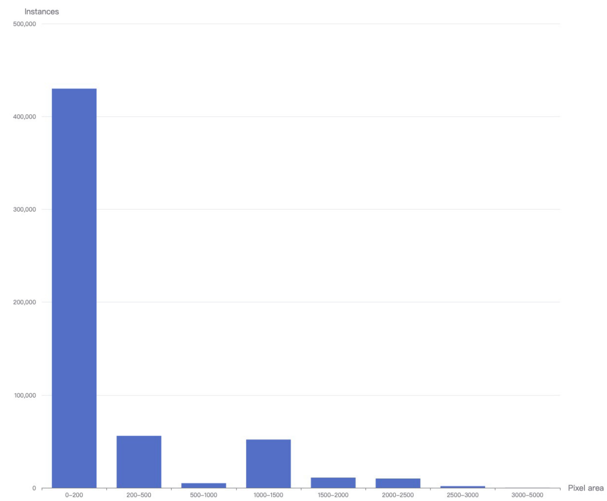

Fig. 10. Number of instances.

relatively small areas. Within the range of 600–3000, the area is relatively balanced, and the aspect ratio is maintained at relatively stable figures. After classification, the area of the four categories was in the range of 600–2500, and the aspect ratio was more stable, which is more conducive for the training of vehicle detection algorithms.

V. Evaluations

A. Evaluation Prototypes

OBB-based prediction is more difficult because the most advanced detection methods are not designed for oriented objects. Therefore, we will choose MMdetection for its accuracy and efficiency, and modify it to predict the OBB. The image size of our dataset is too large to be directly input to the CNN-based detector. Therefore, we crop the original image into a series of 600*600 patches with a stride of 200. Note that, during the cropping process, a complete object may be cut into two parts. There will be incomplete instances on our images. Thus, we set the area of the original object as A_0, the area of the remained part P_r, and as a_r, and then calculate

$$U_r = a_r / A_0.$$

If $U_r < 0.7$, we delete the target; otherwise, we keep it the same as the original annotation.

B. Different Illuminations

Some of the pictures in this dataset are imaging the same area at different time periods, i.e., imaging the same area under distinct illuminations. In Fig. 12, we can see that, under different illuminations, insufficient illumination and shadows influence the detection effect, leading to missed detection or false detection.

C. Experimental Analysis

First, we conducted an experiment on extracting features for all categories in the LOVD dataset, as shown in Fig. 11. We analyzed all categories through the strategy of visualizing the convolutional layer, and we summarized all small categories into four major categories: Car, Truck, Construction, and Public. For the "Car" category, they only correspond to objects with the label "Car." This label has the largest number of instances, and it includes SUVs and cars. Moreover, its characteristics are quite different from those of all other subcategories, so it is taken as a separate category. As for "Truck," by observing from the categories of Dump, Truck, Trailer Truck, and Agitator Truck in Fig. 11, it can be found that these four subcategories have similar characteristics. Their vehicle bodies are composed of two parts: front and carriage, which can also be captured from the feature images. Thus, considering their application scene and contour feature, these four categories can be classified as one big category "Truck" for identification. For "Construction" vehicles, they can further be classified as Excavator, Tanker, Bulldozer, and Crane, as these are the types of vehicles, which mostly appear in construction sites, so they can be classified as one major category. The remaining four categories of Bus, School Bus, Pickup, and Van are classified as "Public" because most of these types of vehicles are used for public transportation and generally can carry six or more passengers.

For the analysis of algorithm performance, we adopt mAP and F1-score as the performance metric, which are the two most common methods. We evaluate those state-of-art algorithms with both four classification labels and 13 classification labels. Results are shown in Table VI. According to the results, we further analyze the performance of each algorithm.

1) YOLOv3: The YOLOv3 algorithm [53] applies Darknet and a large number of residual modules. Convolutional layers are used for downsampling. In order to achieve better detection performance on small objects, YOLOv3 adopts a fusion of multiple scale feature maps. Although YOLOv3's training speed and running time are better than those of other methods, its detection accuracy is not as good as that of the existing R3det method; especially, for the detection of BC3 (Construction), as shown in Table VI and VII, the YOLO algorithm has the lowest detection accuracy. YOLOv3 is based on an anchor box and needs to cluster the target labels in advance. Due to the particularity of the OBB, there is no specific way to use the angle parameter when setting the anchor box. The model can be trained with labels of four point coordinates. The center point can also be used for training with the height and angle information. In addition, there is no standard for calculating the IOU for the OBB, so the application of YOLOv3 in the case of the OBB does not have the same advantages as the HBB. Also, in order to evaluate the detection performance and classification performance of the algorithm on the dataset separately, we first treat all targets as one category for detection and then calculate the classification rate of YOLOv3 among the detected targets. As shown in Table IV, YOLOv3 has a higher score in the classification of targets in LOVD, which shows the goodness of the target classification features in LOVD.

2) Faster R-CNN With FPN: Small objects have less pixel information and so are easily lost during the downsampling process. In order to deal with this type of detection problem with obvious object size differences, the classic method is to

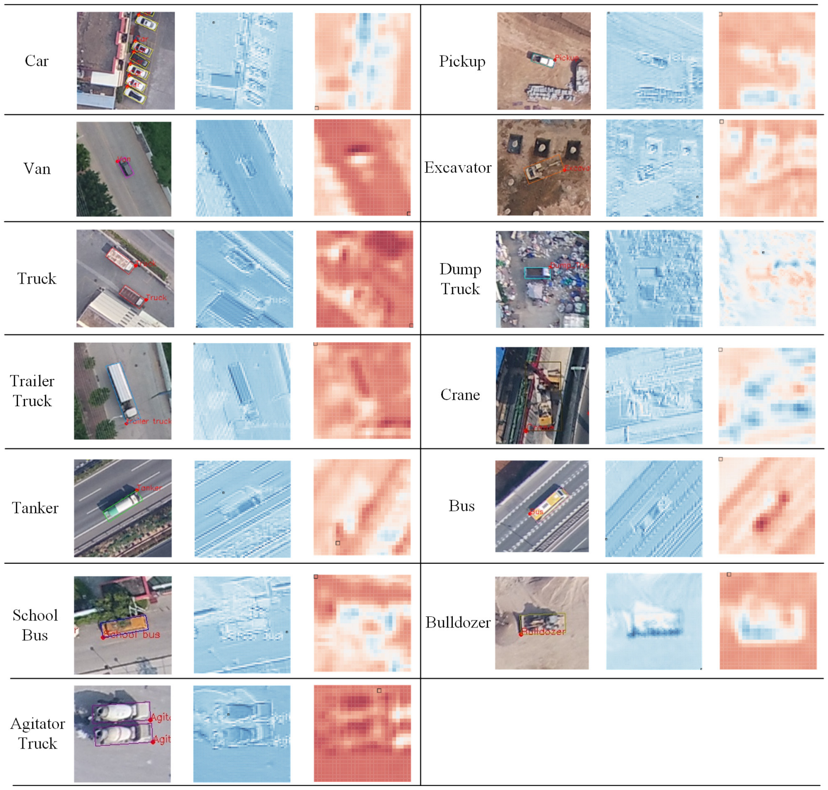

Fig. 11. Feature classification.

(a) (b)

Fig. 12. Test results in the same place under different illuminations. (a) In dim conditions. (b) In light conditions.

TABLE IV
DETECTION AND CLASSIFICATION RESULTS

	Detection rate	Classification rate
YOLOv3	82.4	90.1

use image pyramids to enhance multiscale changes, but this will lead to much more calculations. Therefore, FPN adopted an FPN structure, which can handle the problem of multiscale changes in object detection with a very small amount of calculation. As shown in Tables VI and VII, it can be observed that Faster RCNN with FPN is much worse than YOLO in detecting small vehicles in group BC1. However, it is much better than YOLO when it comes to detecting transportation vehicles with large lengths and widths or more complicated shapes in groups BC2 and BC3. The analysis may be due to excessive downsampling results in the loss of the target feature information with different rules for various vehicle categories. FPN adopts prediction in each layer; its efficiency is much worse than that of YOLO [54].

3) R3det: R3det [50] improves the problem of dense distribution and extremely unbalanced categories in oriented target detection. The oriented anchor avoids noisy areas by adding angle parameters and has better detection performance in dense scenes. However, the number of anchors increases exponentially, reducing the efficiency of the model. Researchers use the method of superimposing feature maps to obtain new features through two-way convolution to improve the detection effect. We found that, although this greatly improves the detection performance of the OBB, the training speed is several times slower than that of the HBB. Thus, how to improve the efficiency of OBB training is still an urgent problem.

4) CSl-FPN: From the analysis of experimental data, such as CSL-FPN-based, compared with other mainstream networks, the detector based on the CSL can learn the direction information of the target very well. However, in the case of more vehicle rotation angles in the test image, its performance is still not that good, such as group BC3 in Table VI.

5) SCRDet and SCRDet++: After experiments, we found that SCRDET++ is very good at detecting small and cluttered objects. The instance-level denoising module for suppressing instance noise in its structure has significantly improved the detection performance of various types of vehicles.

TABLE V
CLASSIFICATION

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Car	Dump	Truck	Trailer truck	Agitator truck	Excavator	Tanker	Bulldozer	Crane	Bus	School bus	Van	Pickup
BC1	BC2				BC3				BC4			
Car	Truck				Construction				Public			

TABLE VI
RESULTS OF FOUR CLASSES

	Backbone	BC1(AP,F1)	BC2(AP,F1)	BC3(AP,F1)	BC4(AP,F1)	(mAP,F1)
Faster R-CNN	ResNet101	(64.9,67.2)	(39.8,49.6)	(15.6,22.7)	(34.3,43.4)	(38.15,45.73)
Faster R-CNN with FPN	ResNet101	(71.6,74.9)	(45.6,50.4)	(24.7,28.1)	(45.1,48.3)	(46.8,49.4)
R3det	ResNet101	(81.1,82.9)	(75.0,73.6)	(45.4,43.1)	(65.0,60.7)	(66.7,65.1)
YOLOv3	darknet	(75.6,81.8)	(37.0,50.7)	(12.2,16.6)	(32.4,44.8)	(39.3,48.3)
CSL-FPN	ResNet101	(77.3,73.2)	(71.3,68)	(40.1,38)	(59.9,59.8)	(61.4,59.8)
R^2CNN	ResNet101	(75.8, 72.4)	(58.7,57.8)	(35.9,31.8)	(54.6,52)	(54.7,53.5)
SCRDet	ResNet101	(75.2,71.8)	(63.7,59.1)	(38.9.36.1)	(56.9,54.2)	(58.6,55.3)
SCRDet++ MS(FPN)	ResNet101	(77.5,73.9)	(72.6,74.1)	(43.4,40.8)	(62.9,59.4)	(64.1,62.0)

TABLE VII
MAP RESULTS OF 13 CLASSES

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Faster R-CNN	Resnet101	64.9	10.1	19.4	27.4	3.6	27.4	3.2	5.4	48.9	3.2	4.7	7.5	14.9
Faster R-CNN with FPN	Resnet101	71.6	19.6	49.8	39.8	10.5	37.8	9.4	15.7	59.5	21.6	12.7	15.2	22.5
YOLOv3	darknet	74.8	8.5	9.1	28.1	6.6	31.9	0	8.1	51.2	18.7	9.1	9.1	13.5
R3det	Resnet101	81.1	63.6	69.6	54.7	48.3	42.7	31.5	34.9	64.3	39.4	38.6	28.9	28.4
CSL-FPN	Resnet101	79.3	69.2	75.2	69.6	65.2	59.1	39.7	42.8	59.3	66.7	56.7	77.4	65.9

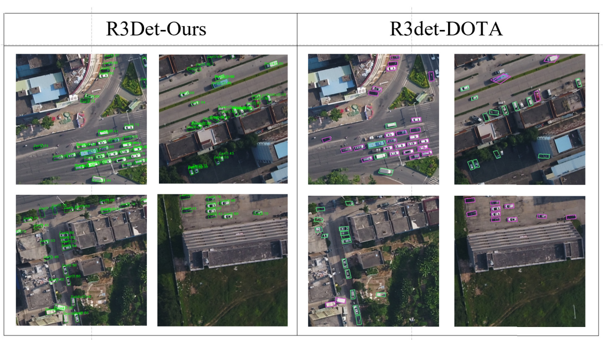

Fig. 13. Universality test of LOVD and DOTA.

Fig. 14. Future work.

Overall, compared with performance in DOTA's SV and LV [32], most of the algorithms have improved vehicle detection results to varying degrees, and the overall accuracy rate has increased by 3.8%. At the same time, in terms of recall rate, we use F1-measure for further experiments. In fact, SCRDET (based on FPN) and R3DET have the best performance, especially for target detection in areas where the direction of the dense and dense cars is changeable.

D. Universality Test

We also perform universality experiments to verify whether our dataset will improve the data-driven algorithms. We choose the DOTA dataset for universality testing because, compared with other aerial object detection datasets, DOTA vehicles' data are relatively abundant. Our strategy is to first adopt R3det for this experiment, then train the R3det algorithm on the

DOTA and LOVD datasets, respectively, and randomly select 100 images from other datasets for testing. In the case where the input picture and other settings are the same, the result is shown in Fig. 13. Although the R3det algorithm trained on the DOTA dataset works very well, compared with the LOVD dataset, there will still be missed targets. We believe that this is because the LOVD dataset has more comprehensive categories and rich tags for various types of vehicles.

VI. Conclusion and Future Work

This article proposes a large-scale, publicly available vehicle dataset that is much larger than any existing vehicle dataset in the remote sensing field. Different from the normal vehicle remote sensing dataset, the proposed LOVD dataset applies OBBs to annotate targets and is more excellent on the diversity of scenarios, weather, and vehicle category. Moreover, evaluation of the performance for those modified mainstream detection algorithms and other bounding-box-specific algorithms is also done on LOVD, and the experimental results can be a useful performance benchmark for future research.

Based on the proposed remote sensing dataset, we will continue to conduct research on vehicle detection algorithms for remote sensing images. Although existing algorithms for detecting OBBs in remote sensing images have been continuously proposed, there has not yet been an algorithm that can truly achieve excellent detection results. As shown in Fig. 14, we still find challenges in the following areas.

1) Detection on dense scenes. When we conducted a universal test, we found that, in low-resolution remote sensing images, vehicles in dense scenes are more difficult to detect.
2) Detection of targets near the edge of the picture and occluded targets. Although the R3det alleviated some of the problems, it still did not achieve relatively excellent results. If most of the features for a vehicle are occluded, the detection performance will be far from satisfactory.
3) Detection of targets in dense fog and shadow. In dense foggy weather, vehicle target features are almost completely occluded, making it difficult to identify such targets.
4) Extraction and identification of various parts of vehicles. The separate parts of the vehicle are also very important for detection, which can make the algorithm learn the characteristics of the target better.

In view of the above, we are going to seek to update the proposed dataset, add more instances for those categories with relatively fewer samples, and optimize the application of the dataset in the multicategory vehicle classification algorithm in the future. We also hope that we will receive valuable suggestions and feedback for our proposed dataset.

References

[1] K. Su, J. Li, and H. Fu, "Smart city and the applications," in *Proc. Int. Conf. Electron., Commun. Control*, Sep. 2011, pp. 1028–1031.

[2] B. Kucherov, O. Pribyl, and V. Artyushenko, "Increasing efficiency of getting results of satellite remote sensing for smart cities," in *Proc. Smart City Symp. Prague (SCSP)*, May 2017, pp. 1–6.

[3] L. Hao, X. Lei, Z. Yan, and Y. ChunLi, "The application and implementation research of smart city in China," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, Jun. 2012, pp. 288–292.

[4] M. C. Hansen, R. S. DeFries, J. R. Townshend, and R. Sohlberg, "Global land cover classification at 1 km spatial resolution using a classification tree approach," *Int. J. Remote Sens.*, vol. 21, nos. 6–7, pp. 1331–1364, 2000, doi: 10.1080/014311600210209.

[5] E. L. Hestir *et al.*, "Identification of invasive vegetation using hyperspectral remote sensing in the California delta ecosystem," *Remote Sens. Environ.*, vol. 112, no. 11, pp. 4034–4047, Nov. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425708002046

[6] R. Sugiura, N. Noguchi, and K. Ishii, "Remote-sensing technology for vegetation monitoring using an unmanned helicopter," *Biosyst. Eng.*, vol. 90, no. 4, pp. 369–379, Apr. 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511004002296

[7] A. Ahmadalipour, H. Moradkhani, H. Yan, and M. Zarekarizi, "Remote sensing of drought: Vegetation, soil moisture, and data assimilation," in *Remote Sensing of Hydrological Extremes* (Remote Sensing/Photogrammetry). Cham, Switzerland: Springer, 2017, pp. 121–149, doi: 10.1007/978-3-319-43744-6_7.

[8] D. Phiri and J. Morgenroth, "Developments in landsat land cover classification methods: A review," *Remote Sens.*, vol. 9, no. 9, p. 967, Sep. 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/9/967

[9] T. A. Carrino, A. P. Crósta, C. L. B. Toledo, and A. M. Silva, "Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 64, pp. 287–300, Feb. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243417301071

[10] M. Fingas and C. E. Brown, "A review of oil spill remote sensing," *Sensors*, vol. 18, no. 1, p. 91, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/1/91

[11] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosyst. Eng.*, vol. 114, no. 4, pp. 358–371, Apr. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511012001419

[12] M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: A meta-review," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111402. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719304213

[13] Y. Zhong, A. Ma, Y. S. Ong, Z. Zhu, and L. Zhang, "Computational intelligence in optical remote sensing image processing," *Appl. Soft Comput.*, vol. 64, pp. 75–93, Mar. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494617307081

[14] Y. Chebud, G. M. Naja, R. G. Rivero, and A. M. Melesse, "Water quality monitoring using remote sensing and an artificial neural network," *Water, Air, Soil Pollut.*, vol. 223, no. 8, pp. 4875–4887, Oct. 2012, doi: 10.1007/s11270-012-1243-0.

[15] Q. Weng, "Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 4, pp. 335–344, Jul. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092427160900046X

[16] L. Giglio, J. Descloitres, C. O. Justice, and Y. J. Kaufman, "An enhanced contextual fire detection algorithm for MODIS," *Remote Sens. Environ.*, vol. 87, pp. 273–282, Oct. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425703001846

[17] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, p. 6442, Nov. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/22/6442

[18] E. Jang, Y. Kang, J. Im, D.-W. Lee, J. Yoon, and S.-K. Kim, "Detection and monitoring of forest fires using Himawari-8 geostationary satellite data in South Korea," *Remote Sens.*, vol. 11, no. 3, p. 271, Jan. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/3/271

[19] R. Neware and A. Khan, "Survey on classification techniques used in remote sensing for satellite images," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 1860–1863.

[20] D. R. Sowmya and P. D. Shenoy, "Remote sensing satellite image processing techniques for image classification: A comprehensive survey," *Int. J. Comput. Appl.*, vol. 161, no. 11, pp. 24–37, Mar. 2017.

[21] Y. Li, H. Zhang, X. Xue, Y. Jiang, and Q. Shen, "Deep learning for remote sensing image classification: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 6, p. e1264, Nov. 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1264

[22] Y. Gu, Y. Wang, and Y. Li, "A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection," *Appl. Sci.*, vol. 9, no. 10, p. 2110, May 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/10/2110

[23] L. Drăguţ, D. Tiede, and S. R. Levick, "ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 6, pp. 859–871, Jan. 2010, doi: 10.1080/13658810903174803.

[24] T. Su and S. Zhang, "Local and global evaluation for remote sensing image segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 256–276, Aug. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271617302654

[25] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271618301229

[26] S. Yin, Y. Zhang, and S. Karim, "Large scale remote sensing image segmentation based on fuzzy region competition and Gaussian mixture model," *IEEE Access*, vol. 6, pp. 26069–26080, 2018.

[27] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective," *ISPRS J. Photogram. Remote Sens.*, vol. 150, pp. 115–134, Apr. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619300425

[28] Y. Yi, Z. Zhang, W. Zhang, C. Zhang, W. Li, and T. Zhao, "Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network," *Remote Sens.*, vol. 11, no. 15, p. 1774, Jul. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/15/1774

[29] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 165, pp. 108–119, Jul. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301295

[30] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616300144

[31] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271619302825

[32] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," 2017, *arXiv:1711.10398*.

[33] F. Tanner *et al.*, "Overhead imagery research data set—An annotated data library tools to aid in the development of computer vision algorithms," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2009, pp. 1–8.

[34] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271614002524

[35] K. Liu and G. Máttyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.

[36] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery : A small target detection benchmark," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 187–203, Jan. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320315002187

[37] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3735–3739.

[38] T. Nathan Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," 2016, *arXiv:1609.04453*.

[39] M. Y. Yang, W. Liao, X. Li, and B. Rosenhahn, "Deep learning for vehicle detection in aerial images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3079–3083.

[40] G. Gao, Q. Liu, and Y. Wang, "Counting dense objects in remote sensing images," 2020, *arXiv:2002.05928*.

[41] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," 2016, *arXiv:1612.03144*.

[42] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, *arXiv:1506.02640*.

[43] W. Liu *et al.*, "SSD: Single shot MultiBox detector," 2015, *arXiv:1512.02325*.

[44] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[45] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[46] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Netw.*, vol. 95, pp. 19–28, Nov. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608017301806

[47] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[48] G. Cheng, C. Lang, M. Wu, X. Xie, X. Yao, and J. Han, "Feature enhancement network for object detection in optical remote sensing images," *J. Remote Sens.*, vol. 2021, pp. 1–14, Jul. 2021. [Online]. Available: https://spj.sciencemag.org/journals/remotesensing/2021/9805389/

[49] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning RoI transformer for detecting oriented objects in aerial images," 2018, *arXiv:1812.00155*.

[50] X. Yang, J. Yan, Z. Feng, and T. He, "R3Det: Refined single-stage detector with feature refinement for rotating object," 2019, *arXiv:1908.05612*.

[51] X. Yang and J. Yan, "CSL: Arbitrary-oriented object detection with circular smooth label," in *Computer Vision* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, Aug. 2020, pp. 677–694.

[52] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[53] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[54] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

Puti Yan received the B.E. degree in electronical information science and technology and the M.S. degree in aerospace science and technology from the Harbin Institute of Technology, Harbin, China, in 2017 and 2019, respectively, where he is pursuing the Ph.D. degree in aerospace science and technology.

Xinrui Liu received the B.E. degree in aerospace science and technology from the Harbin Institute of Technology, Harbin, China, in 2019, where he is pursuing the M.S. degree of aerospace engineering.

His research interests include object detection and remote sensing.

Feng Wang received the Ph.D. degree in aerospace science and technology from the Harbin Institute of Technology, Harbin, China, in 2008.

Since 2014, he has been a Professor with the School of Astronautics, Harbin Institute of Technology. His research fields include dynamics and control of distributed satellites systems, on-orbit service for spacecraft, the general design of small satellites, and spacecraft attitude control.

Chengfei Yue (Member, IEEE) received the B.E. degree in spacecraft design and engineering from the Honor School, Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree from the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, in 2019.

He is an Associate Professor with the Institute of Space Technology and Applied Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. His research interests include spacecraft design, on-orbit service, and spacecraft attitude control.

Xin Wang (Member, IEEE) received the B.E. and Ph.D. degrees in computer science and technology from Zhejiang University, Hangzhou, China, in 2011 and 2017, respectively, and the Ph.D. degree in computing science from Simon Fraser University, Burnaby, BC, Canada, in 2016.

He is an Assistant Professor with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He has published several high-quality research papers in top conferences, including ICML, KDD, WWW, and SIGIR ACM Multimedia. His research interests include relational media big data analysis, multimedia intelligence, and recommendation in social media.

Dr. Wang was a recipient of the 2017 China Postdoctoral Innovative Talents Supporting Program. He also received the ACM China Rising Star Award in 2020.