Vehicle Detection from UAV Remote Sensing Images Using Deep Learning

Luwen Tan, Hasi Bagan
School of Environmental and Geographical Sciences, Shanghai Normal University, Shanghai 200234, China,
Email: 1000495878@smail.shnu.edu.cn
hasibagan@staff.shnu.edu.cn

KEY WORDS: Deep learning; UAV; Remote Sensing; Vehicle identification; Urban management

ABSTRACT: With the development of the city, the number of vehicles in the city is constantly increasing, a large number of vehicles not only increase traffic congestion, but also contribute to the frequent occurrence of traffic accidents, and the public recreational space in the residential area also becomes more and more crowded because of the increase of vehicles. For this problem, the temporal and spatial resolution of satellite remote sensing data are difficult to meet the requirements of urban vehicle information monitoring, while the installation of a large number of fixed cameras is costly and there are many monitoring blind spots. The remote sensing monitoring by UAV can meet this demand with low cost. In this paper, we use UAV to take low-altitude photographs of some stations, highways, neighborhoods in Shanghai and parking lots in Inner Mongolia Autonomous Region to obtain aerial remote sensing images with centimeter-level resolution. In order to obtain higher quality research data, different flight parameters are set during the flight photography process, taking into account the characteristics of the study area and flight management policies. To increase the robustness of the model, different levels of image enhancement are done on the samples before training. Then single target extraction of vehicles in UAV images is performed using the Unet convolutional neural network technique of deep learning. The model training parameters are continuously adjusted during the training process to get the best training results and obtain 99% ultra-high accuracy. Through controlled experiments, it can be seen that the recognition effect of deep learning on vehicles is much better than that of traditional machine learning methods. This experiment shows that the method of this paper is effective and real-time, and can provide a valuable technical means for urban traffic and community management.

1. INTRODUCTION

With the progress of society and the development of science and technology, the car has become a necessary means of transportation for people's daily travel. However, a large number of vehicles not only increase traffic congestion, but also contribute to the frequent occurrence of traffic accidents, and the public recreational space in residential areas becomes more and more crowded because of the increase of vehicles. Therefore, the monitoring and management of vehicles has become a big problem. Quick and accurate access to vehicle distribution information within the city will help city managers to manage the city we live in more efficiently.

Satellite remote sensing can monitor the dynamic changes of the whole earth well, however, for the detection of small target objects like cars, it is difficult to meet the requirements in both spatial and temporal resolution. In near-ground scenes, fixed cameras are generally used for long-term monitoring of vehicle densities, but their drawbacks are also obvious; fixed cameras have limited detection scenes and more monitoring blind spots. UAV remote sensing has the advantages of lightness, speed, low cost, easy operation and less atmospheric interference, which can well make up for the defects of satellite remote sensing and fixed cameras. Nowadays, UAV technology has gradually penetrated into many fields such as transportation, security, logistics, photography, mapping, etc. It has become an indispensable tool for human beings to perceive the world and obtain environmental information.

However, visual inspection of vehicles in UAV remote sensing imagery is a time-consuming and labor-intensive operation. The challenge of remote sensing has always been to automatically, quickly and accurately acquire target information from the imagery. In recent years, machine learning techniques have emerged to provide powerful and intelligent methods to identify ground target objects, but traditional manual feature and machine learning methods for vehicle detection require manual feature screening according to the task purpose and are more susceptible to factors such as lighting, shadows, and image quality, such as support vector machine methods and random forest classification. With the deep development of machine learning, scientists have also proposed the deep learning method, which is an artificial neural network method with multiple hidden layers and deeper combinations, using a large number of positive and negative samples, combined with learning algorithms, to obtain the discriminant function, so as to identify the target in the image with high accuracy, which shows a good performance in target detection.

2. STUDY DATA

2.1 Study Area

The main study area is Shanghai, China, and part of Chifeng City, Inner Mongolia Autonomous Region. The total area of Shanghai is 6340.5 square kilometers. As the largest city in China, Shanghai has a high concentration of population and many vehicles. As of November 1, 2020, at 00:00, the resident population was 24,870,900. The total number of registered motor vehicles in the city as of 2020 is 4.691 million, of which 3.971 million are sedan car. The three scenarios of road, station and district for vehicle inspection in this paper are all in Shanghai. Since many of the parking lots in Shanghai are set up underground and inside buildings, the outdoor open-air parking lots are not representative, and the choice of this scenario is changed to Chifeng City, Inner Mongolia.

Chifeng City is located in the eastern part of Inner Mongolia Autonomous Region, a total area of 90,021 square kilometers, and a resident population of 4,035,967 as of November 1, 2020. The total resident population is one-sixth of that of Shanghai, but the land area is fourteen times larger than that of Shanghai, so the parking lots are mostly outdoor open-air parking lots.

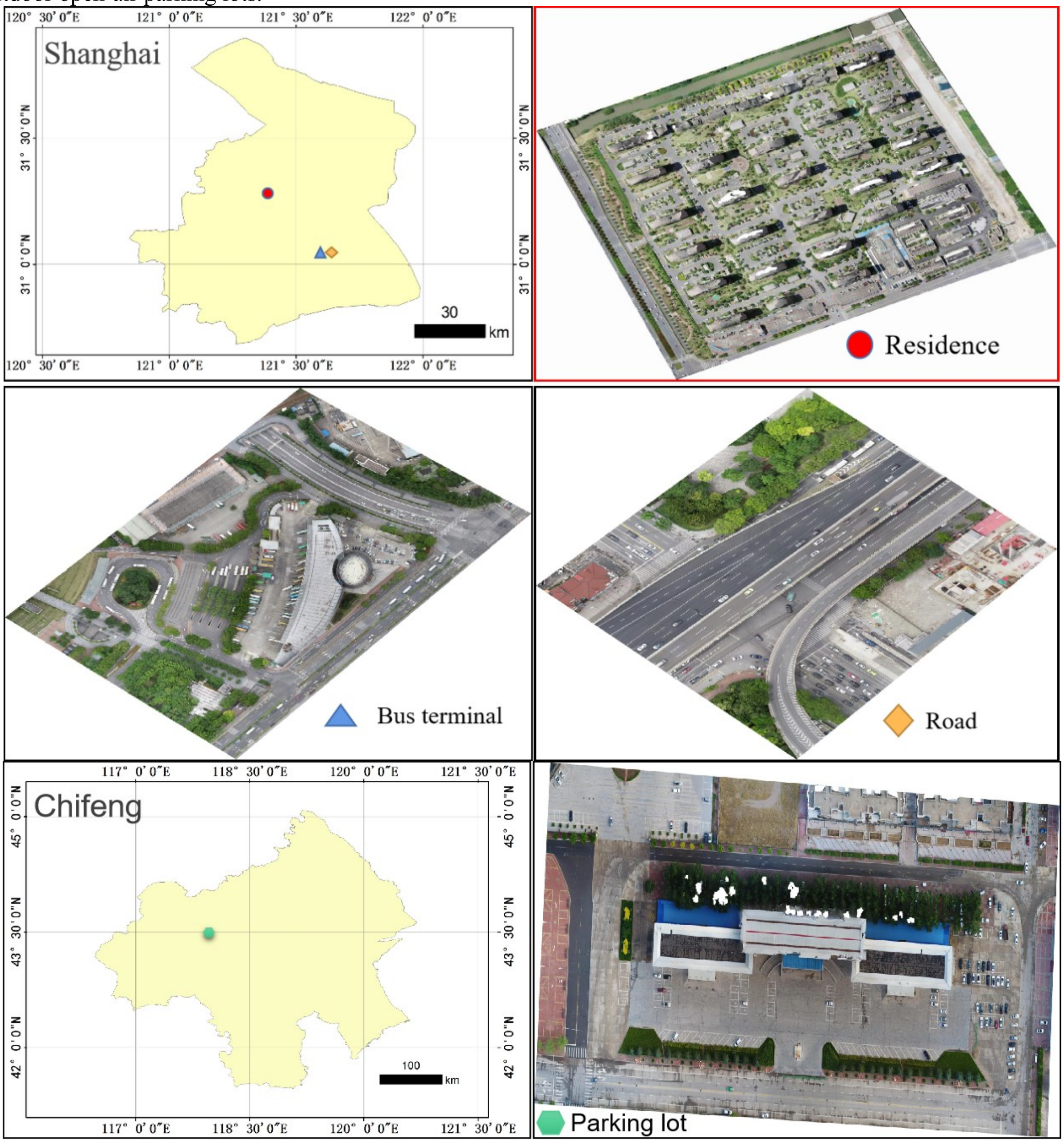


Fig 1.The location of study area

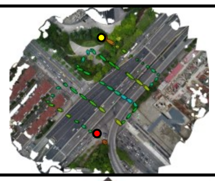
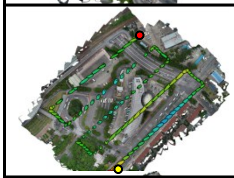
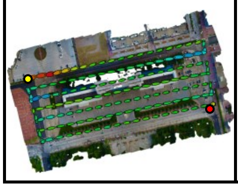
2.2 Data Acquisition

The experimental data in this paper were collected by a DJI commercial drone, phantom 4 Pro. The pan-tilt-zoom camera of this drone can guarantee that the drone can acquire high definition images even during the flight shaking.

Fig 2.DJI Phantom4 Pro

The drone control and route planning software used is Litchi 4.17.0. For the four different scenarios, we set four different flight tracks and parameters. Since Shanghai has two large international airports, Hongqiao and Pudong, half of the city's administrative area is within the restricted flight range, and the maximum flight height cannot exceed 120 meters.The main content of the first scene is the elevated bridge next to Shanghai South Railway Station in Xuhui District and the lane under the bridge perpendicular to the elevated bridge, the cars on the road are in a fast driving state, in order to reduce the impact on traffic conditions, the aircraft flight speed is set to 40 km/h to complete the shooting at a faster speed. In order to reduce the impact on the traffic condition as much as possible, we set the shooting interval to 1 second to meet the demand of sample size for deep learning. The second scene is mainly shot by the bus of Shanghai South Bus Station in Xuhui District, also in order to reduce the impact on the operation of the bus station, the flight speed is set to 40 km/h. In comparison, most of the buses in the station are in a stationary state, so the shooting interval was set to 2 seconds.In the third scene, the main shooting content is the Hejia Xinyuan district in Baoshan District, which is a new district with mostly young residents and a large number of vehicles, the flight speed of the route is set to 30 km/h and the photo interval is set to 2 seconds. The fourth scene of the main shooting content for Chifeng City Daban Town, Balin Youqi Government parking lot, because the scene more vehicles parked in the shade of the trees, so the flight height is set to 60 meters to get more information about the vehicles under the shade of the trees, the flight speed is set to 30 km / h, the photo interval is set to 2 seconds; the number of photos in the four scenes a total of 600, some of the areas overlap too high, labeling The number of photos in the four scenes is 600, among which some areas are overlapped too much and labeled with little significance, so some images are discarded, so the number of images used to make sample labels is 436.

Table 1.Route planning and parameter list

Route Planning	Flight altitude	Flight Speed	Photo interval	Resolution	Number of photos
	100m	40km/h	1s	2.56 cm/pix	52
	100m	40km/h	2s	2.58 cm/pix	111
	100m	30km/h	2s	2.64 cm/pix	236
	60m	30km/h	2s	1.59 cm/pix	201

3. METHODOLOGY

3.1 Image Stitching

Due to the limitation of UAV flight altitude, only a small part of the range of each photographic image is available. To obtain a complete scene of the study area, the photos acquired on each flight track need to be stitched and fused to finally generate RGB three-band orthomosaic images and DSM images containing elevation information. This experiment uses Agisoft Metashape Professional 1.6.6 software to process the aerial images, and the main technical processes are as follows.

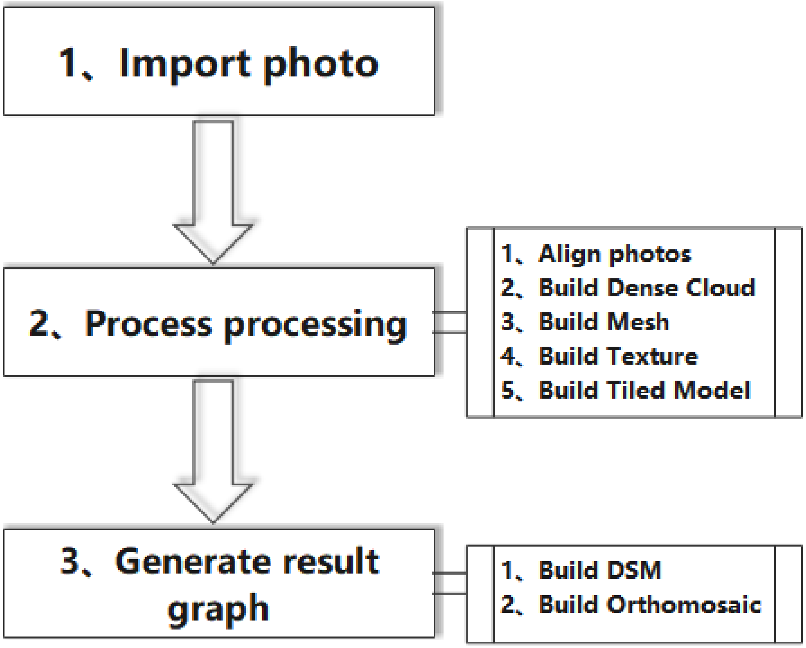


Fig 3.Image Mosaic flow chart

3.2 Image Enhancement

In this paper, the Opencv library based on python is used to enhance the image of UAV images, and then the experiment will get better results. It is well known that acquiring remote sensing images at different times of the day will have different brightness, chromatic aberration, shadows and noise, in addition to the hardware and software condition of the UAV equipment and the flight route planning will all have different effects on the image effect. In the experiments of the thesis, the brightening process is used to simulate the scene when the sunlight intensity is at its maximum; the darkening process is used to simulate the scene when the light is weak; the operation of adding pretzel noise is used to simulate the scene when the light is extremely weak and the hardware or software is obstructed; the reduced resolution process is used to simulate the scene when the route is set too high. The above image enhancement process is used to increase the robustness of the deep learning model.

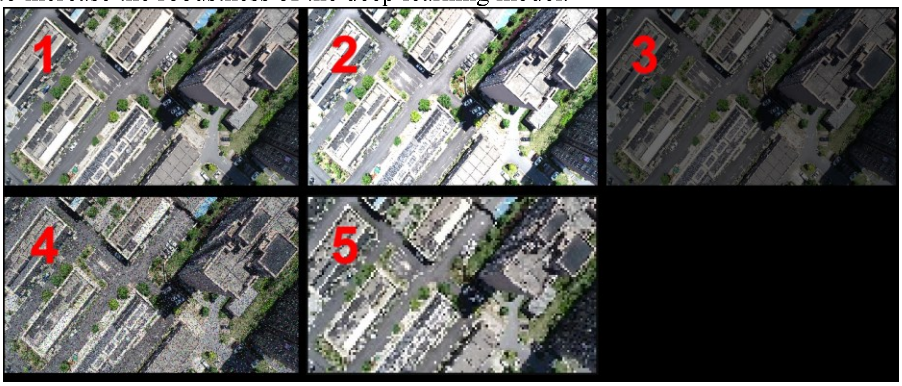


Fig 4.Image enhancement

3.3 Unet convolutional neural network

The Unet network structure is a network structure proposed by Olaf Ronneberger, Philipp Fischer and Thomas Brox in the 2015 ISBI competition, which consists of two parts, the systolic channel and the dilated channel, forming a U-shaped structure, hence the name U-Net. The systolic channel is similar to the traditional convolutional network structure. The result of each convolution is activated using the ReLU function, and the feature map length and width remain unchanged while the number of feature channels becomes twice the original size, and then the maximum pooling operation is performed by 2×2 filtering with a step size of 2. The image length and width were reduced by half after each downsampling, and the number of feature channels remains unchanged. And each step of the expansion channel first goes through a 2×2 upsampling operation, which makes the number of features of the feature map reduced to half of the original, and the length and width of the image expanded to twice of the original. The up-sampled feature map is then connected to the feature map of the corresponding image size at the shrinkage channel, so that the feature map obtained contains both shallow and deep feature information of the image. The number of channels in the connected feature map becomes twice as many as the original one, and then two repeated 3×3 convolution operations are performed to reduce the number of channels in the feature map by half. In the last layer, a 1×1 convolution is used to map the 64 features in each pixel to the two classes to be classified, thus categorizing each image element into either the vehicle or background class, respectively.

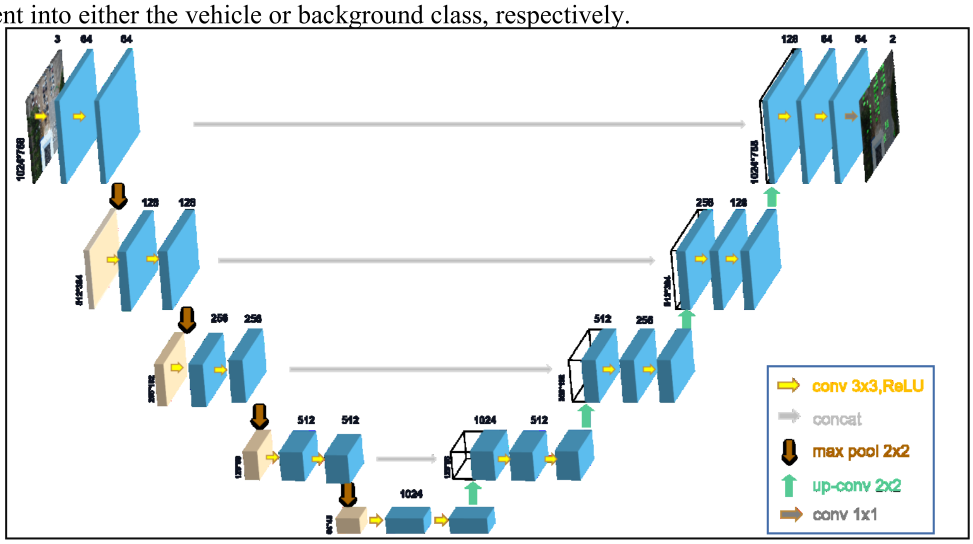


Fig 5.The architecture of Unet

4. RESULTS

4.1 Experimental environment

The vehicle detection experiments designed in this paper mainly consist of three phases: sample annotation, model training and control test, and different processing platforms are used in each phase. The hardware and software configurations relied on for the experiments are shown in the following table.

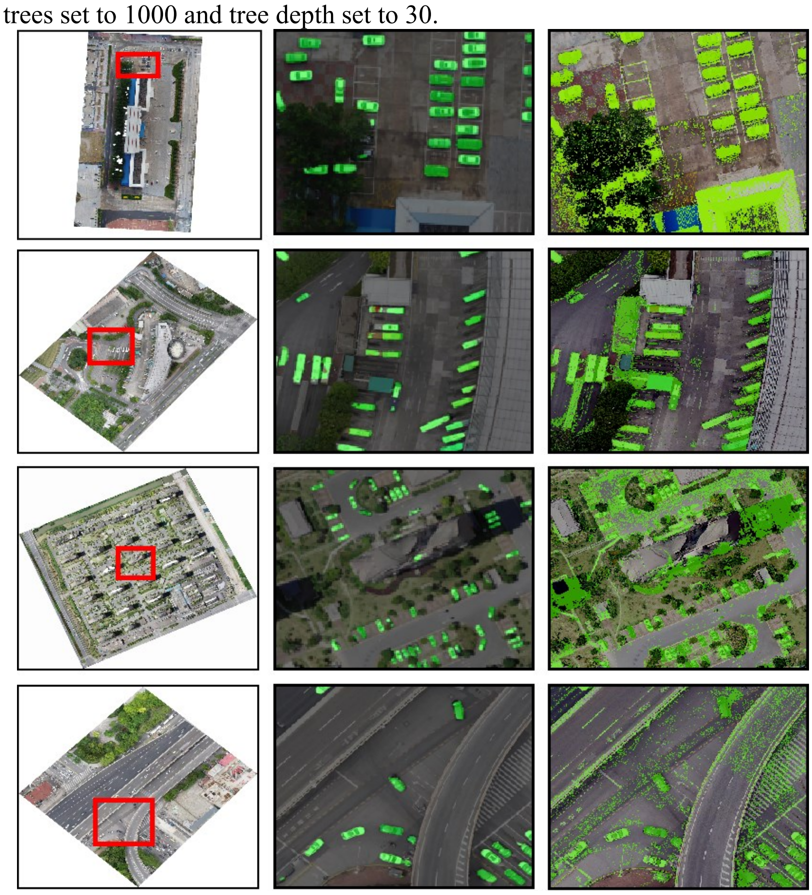
Table 2.The enviroment of expriment

Content	Parameter of Software & Hardware
GPU	NVIDIA Quadro K1200 CUDA CoreS：512
CPU	Intel(R) Xeon(R) E3-1225
Development of Language	Python3.7
Deep learning platform and framework	Tensorflow2.2 Opencv
Annotation tool	Lablme 3.16.7
Control test software	ArcGIS Pro 2.8

4.2 Analysis of results

Since the Unet semantic segmentation model classifies remote sensing images on a pixel-by-pixel basis, the computer's computational workload is huge during the experiment, which requires more computer memory. Batch size is set to 1, and one RGB image is input each time for training. The initial learning rate during training is set to 10^{-3} to speed up the convergence, and when the model accuracy tends to converge, the learning rate is changed to 10^{-4} so that the model can achieve higher accuracy.

The control experiment uses the traditional machine learning random forest algorithm. The random forest mainly sets two hyperparameters: number of trees and tree depth, and the best results are obtained through several experiments with the number of trees set to 1000 and tree depth set to 30.



(i) Original image (ii) Unet (iii) Random Forest

Fig 6.Comparison of Unet and random forest experiment results

From Fig. 6, we can clearly see that in four scenes, the classification results of UAV remote sensing images based on deep learning can basically identify every vehicle accurately, and there are very few vehicles in scene 2 and scene 4 that are incompletely identified. The random forest method is less effective in vehicle recognition, with more misclassification cases, and it is easy to misclassify the background information such as vegetation and road surface into vehicles. Deep learning is able to learn the feature information of the vehicle target object well and has good prediction performance due to its training process requiring a large number of samples and the extremely large number of neuron parameters of the model. The random forest, on the other hand, has far fewer parameters than the deep learning model and cannot extract the numerous feature information of the target vehicle well, so we need to input images with as many features as possible during data input to ensure the accuracy of vehicle detection. In the control experiment, in addition to the RGB tri-band of the original image input, the DSM band is also added, which enables the classifier to distinguish the vehicle from the background by the elevation information of the features. In scenario two and three, the random forest algorithm well eliminates the houses with large differences in elevation from the vehicle, but in scenario 1, due to the low flight altitude, the quality of the obtained elevation data is poor, and a large number of vegetation and house image elements are wrongly into vehicles, while in scenario 4, since there are vehicles on both the viaduct and the ground road, it cannot distinguish roads and vehicles well by the elevation data, resulting in poor differentiation of roads and vehicles.

4.3 Results evaluation

Usually we use cross-entropy loss functions and accuracy to evaluate the performance of deep learning models. The cross-entropy loss function in the dichotomous case, the model needs to predict the final outcome in only two cases, for each category our prediction gets the probability of p and $1-p$, the expressions are:

$$C = -\frac{1}{n}\sum_{\mathrm{i}}[y_i \cdot log(p_i) + (1-y_i)\cdot \log(1-p_i)]$$

Among them, n is denotes the total number of samples; y_i is denotes the label of sample i, positive class is 1, negative class is 0; p_i is denotes the probability that sample i is predicted to be a positive class.

Accuracy, in layman's terms, is the percentage of correct results predicted by our model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

After obtaining the prediction result map, we calculate the number of TP, FP, TN and FN pixels in the result image, respectively. TP indicates the predicted positive sample image and the actual positive sample image; FP indicates the predicted positive sample image and the actual negative sample image; TN indicates the predicted negative sample image and the actual positive sample image; FN indicates the predicted negative sample image and the actual negative sample image.

Table 3.Model evaluation results

	Accuracy	Cross Entropy Loss
Train Dataset	99.88%	0.3%
Validate Dataset	99.43%	2.62%

In the experiments of this paper, there are 2180 sample images, and the samples are divided into training set and validation set in the ratio of 6:4, there are 1308 sample images in the training set and 872 sample images in the validation set. The accuracy in the validation set is 99.43% and the cross-entropy loss value is 2.62%, which shows that the model performs very well in vehicle detection both from the numerical value and the result image.

5. DISCUSSION AND OUTLOOK

In order to help the construction of a better city and improve the ability of social managers to monitor urban vehicles, this paper adopts a deep learning method to achieve the detection of vehicle targets in UAV remote sensing images, and by comparing it with traditional machine learning methods, it can be found that the deep learning method is currently the best choice to achieve vehicle target information extraction, and the vehicle information extraction can reach 99% super high accuracy. The paper details the complete process from UAV route planning to image enhancement to model training and prediction, proving that the method is realistic and feasible. The simplicity and convenience of UAVs also enable social managers to use them to obtain the latest aerial remote sensing images anytime and anywhere, and then make corresponding spatial distribution maps of vehicles to assist in urban governance.

The accuracy of deep learning methods in vehicle detection has reached a very high point, and in the future research of UAV vehicle detection, it is no longer significant to improve the detection accuracy again. Our research direction of vehicle detection by UAV remote sensing will be mainly distributed in the following two aspects: First, the detection target will be subdivided, such as vehicles subdivided into sedan cars, buses, trucks, electric vehicles, etc. The impact of different types of vehicles on urban congestion is not the same, and subdividing vehicle types will have further improvement on improving urban life; Second, the vehicle distribution results and other data for statistical analysis research, such as the correlation research between the spatial distribution of vehicles and the layout of catering industry, to ease the traffic congestion situation by optimizing the layout of catering industry, etc.

6. REFERENCES

Ammar, A., Koubaa, A., Ahmed, M., Saad, A., & Benjdira, B., 2021. Vehicle detection from aerial images using deep learning: A comparative study. Electronics, 10(7), 820.

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M., 2017. Deep learning approach for car detection in UAV imagery. Remote Sensing, 9(4), 312.

Audebert, N., Le Saux, B., & Lefèvre, S., 2017. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. Remote Sensing, 9(4), 368.

Bouguettaya, A., Zarzour, H., Kechida, A., & Taberkit, A. M., 2021. Vehicle Detection From UAV Imagery With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems.

Guo, Y., Chen, S., Wu, Z., Wang, S., Robin Bryant, C., Senthilnath, J., ... & Fu, Y. H., 2021. Integrating Spectral and Textural Information for Monitoring the Growth of Pear Trees Using Optical Images from the UAV Platform. Remote Sensing, 13(9), 1795.

Gupta, H., & Verma, O. P., 2021. Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. Multimedia Tools and Applications, 1-21.

Inoue, Y., & Yokoyama, M., 2017. Drone-based remote sensing of crops and soils and its application to smart agriculture. 日本リモートセンシング学会誌, 37(3), 224-235.

Joo, Y. J., 2017. Detection method for road pavement defect of UAV imagery based on computer vision. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 35(6), 599-608.

Khan, N. A., Jhanjhi, N. Z., Brohi, S. N., Usmani, R. S. A., & Nayyar, A., 2020. Smart traffic monitoring system using unmanned aerial vehicles (UAVs). Computer Communications, 157, 434-443.

Kumar, S., Rajan, E. G., & Rani, S., 2021. A Study on Vehicle Detection through Aerial Images: Various Challenges, Issues and Applications. In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 504-509). IEEE.

Li, Y., Melgani, F., & He, B., 2019. Fully Convolutional SVM for Car Detection in Uav Imagery. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 2451-2454). IEEE.

Luo, X., Tian, X., Zhang, H., Hou, W., Leng, G., Xu, W., ... & Zhang, J., 2020. Fast automatic vehicle detection in uav images using convolutional neural networks. Remote Sensing, 12(12), 1994.

Moranduzzo, T., & Melgani, F., 2013. Automatic car counting method for unmanned aerial vehicle images. IEEE Transactions on Geoscience and Remote Sensing, 52(3), 1635-1647.

Osco, L. P., Junior, J. M., Ramos, A. P. M., Jorge, L. A. D. C., Fatholahi, S. N., Silva, J. D. A., ... & Li, J., 2021. A review on deep learning in UAV remote sensing. arXiv preprint arXiv:2101.10861.

Ronneberger, O., Fischer, P., & Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp.234-241). Springer, Cham.

Srivastava, S., Narayan, S., & Mittal, S., 2021. A survey of deep learning techniques for vehicle detection from UAV images. Journal of Systems Architecture, 102152.

Vlahogianni, E. I., Del Ser, J., Kepaptsoglou, K., & Laña, I., 2021. Model Free Identification of Traffic Conditions Using Unmanned Aerial Vehicles and Deep Learning. Journal of Big Data Analytics in Transportation, 3(1), 1-13.

Xu, Y., Yu, G., Wang, Y., Wu, X., & Ma, Y., 2017. Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation, 2017.

Yue, J., Feng, H., Li, Z., Zhou, C., & Xu, K., 2021. Mapping winter-wheat biomass and grain yield based on a crop model and UAV remote sensing. International Journal of Remote Sensing, 42(5), 1577-1601.