Research on UAV Image Classification Based on Deep Learning

Runqi Li[1], Cen Chen[1] and Mengyao Chen[1]
[1]*Institute of Information, Beijing University of Technology, No100 Pingleyuan, Chaoyang District, Beijing, China*
{ lirunqi,cenchen,chenmengyao }@ emails.bjut.edu.cn

Keywords: Deep learning, UAV, Regional classification, Image segmentation.

Abstract: In order to improve the efficiency of urban traffic operation, this paper combines deep learning technology and UAV photography of drones. With the algorithm in this paper, we can classify traffic area and static area in high quality and speed. We make a test in Fengtai District of Beijing to conduct regional traffic identification research on the traffic content of the area. The primary identification area includes the vehicle travel area and coordinates the pedestrian area in a coordinated manner. The main research result of this algorithm is to propose a key frame extraction scheme for UAV image and then combine it with the application of Mask R-CNN in high-altitude image to identify the ground area. The experimental results are similar to the same algorithm (refer to FCNs for this article). Comparative benchmarks) have obvious advantages of high speed and high accuracy, which are of great help to traffic safety and urban planning.

1 INTRODUCTION

With the development of science and technology, basic urban map construction has been realized by means of satellite remote sensing technology, but the details are still in short supply. The emergence of UAV is undoubtedly a new help for the construction of urban maps. Wide range of applications, such as public transportation, news media, aerial photography, agricultural monitoring, etc (Deren, 2014) (W., 2013). With a UAV in a place where people can't reach it, it's easy to get real-time picture information at the best angle. The collected real-time image information plus the deep learning method is trained to determine the category of the circled area, thereby implementing classification. At present, there have been studies on regional classification and identification of environmental pollution areas (Qiong W, 2018), and also there are research about regional terrain classifications (L., 2018), but research on traffic with UAV is still the current pattern recognition and the poor areas of the drone field.

Deep learning is one of the important advances in the development of artificial intelligence, and research in its field has already landed in a wide range of industrial fields (Chen Y, 2017), one of the most used ones is to do image classification. Girshick (R., 2015) has made success in combination of convolutional neural networks and regional algorithms heralds the beginning of a high-speed, high-precision recognition era. Combining the high-speed and accurate recognition and classification technology of UAV's efficient image capture and deep learning, it can be of vital help to the construction of urban maps. The use of targeted algorithms and architecture can also be very accurate. Guarantee.

The main focus of the UAV detection traffic area is the frame selection of the UAV image and the identification of the area. The main methods of frame fetching are the key frame coding frame method and the feature pixel frame method (Narasimha R ., 2003). In view of the continuity of UAV images, this paper adopts the method of taking frames by feature pixels to avoid the repetitive effect of multiple repeated frames on the detection algorithm. In terms of region identification, this paper adopts Mask-RCNN-based image instance segmentation algorithm (Kaiming, 2018) and adds the integrated structure to the whole region. The final output effect is the regional frame selection.

2 RELATED WORK

The current image segmentation application algorithms mainly include supervised segmentation and unsupervised segmentation. The segmentation algorithm in classification based on classification can be divided into pixel-level segmentation and superpixel-based segmentation algorithm.

Li, R., Chen, C. and Chen, M.
Research on UAV Image Classification Based on Deep Learning.
DOI: 10.5220/0008098701480152
In *Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications (CTISC 2019)*, pages 148-152
ISBN: 978-989-758-357-5

Unsupervised segmentation algorithm (Moore, 2008) such as Moorer et al. proposed a superpixel lattice over segmentation method based on greedy algorithm (Shi J, 2000).

2.1 Segmentation Algorithm Based on Superpixel

The concept of superpixel is proposed by Xiaofeng Ren (Shi J, 2000). It is used to describe the pixels of adjacent features to produce "pixel clusters", and then the local clustering of images is classified, which has the advantage of fast operation. However, due to lack of specificity, under-segmentation often occurs result. In terms of application methods, superpixel segmentation mainly includes image segmentation based on graph theory and a method based on statistical classifier, such as Normalized-cuts (Y, 2000)、Graph-cuts (Long, 2014)、Entropy rate superpixel segmentation (Liu M Y, 2010).

2.2 Segmentation algorithm based on Deep Learning

In order to solve the problem of segmentation accuracy, the segmentation algorithm begins to focus on extracting image depth features and using the maximum inter-class idea to segment each pixel region (Kaiming, 2018), and uses gradient ascending iteration to update the classification function weights to achieve the optimal solution. The full convolution image segmentation algorithm proposed by Long et al. is the best embodiment for pixel-level segmentation. Later, a large number of improved algorithms have been generated, and w-net (Xia, 2017), u-net (Von Eicken, 1995), and other excellent segmentation algorithms have been evolved until the mask. The advent of Mask-Rcnn has pushed the field of image segmentation research to the forefront of machine vision research.

3 UAV IMAGE ANALYSIS AND PRINCIPLE

In order to identify and identify the image area in a targeted manner, it is first necessary to use image preprocessing and filter feature repetition data for images acquired by the drone. Since the UAV image is a dynamic video format, this paper analyzes the key frame method (Agarwala, 2004), and selects the image with the feature as the data set by combining the image enhancement technology of the key frame image with the big data information analysis method. We choose to select key frames at the first beginning. With the progress of key frames, we can finish other jobs. At present, many techniques for extracting key frames are based on video coding, but this method does not help the aerial image extraction information.

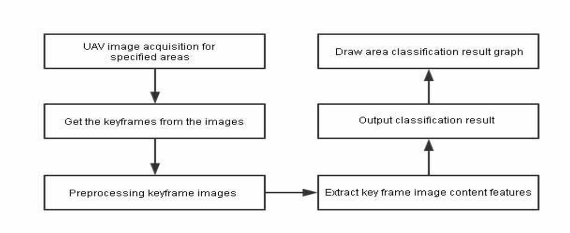
UAV image acquisition for specified areas
Get the keyframes from the images
Preprocessing keyframe images
Extract key frame image content features
Output classification result
Draw area classification result graph

Figure 1: UAV aerial area segmentation design block diagram.

3.1 Principle of Regional Feature Judgment

Since the ground shot at high altitude belongs to the static region, the key frame extraction method adopted in this paper is extracted according to the difference of the feature features. We first extract the I frame in the video coding according to the method of encoding the key frame, and gather the I frame. The eigenvalues are calculated and the eigenvalues are calculated from the constructed system of characteristic equations. This equation describes the eigenvalues by calculating pixel anomalies. The characteristic formulas are as follows:

$$f(x) = \frac{\sum|a_i - a_{i+1}|}{n*w*h} \tag{1}$$

$$g(x) = \begin{cases} f(x) * l_e & ,f(x) \geq 0.5 \\ 0 & ,f(x) < 0.5 \end{cases} \tag{2}$$

Where a_i is the value of the ith image converted to the square matrix determinant, l_e is the key frame roughness degree, calculated by the resolution (w*h) and the pixel variance (σ), and the calculation formula is as follows:

$$l_e = \frac{\sigma}{w*h} \tag{3}$$

We will calculate the result g(x) according to the order from the largest to the smallest, and select the first 50 frames of each sequence for the next study. The results of these 50 frames are our key frames.

3.2 Key Frame Image Pre-processing

In order to deal with the residual factor of the image determined by the feature region, this paper adopts the wavelet threshold demonising based on the wavelet domain, and calculates the hard threshold to eliminate the influence of the noise interference on the image depth information extraction. The wavelet minimax threshold is as Equation 4:

$$\gamma = \begin{cases} 0.3936 + 0.1829\left(\frac{lnN}{\ln 2}\right), N > 32 \\ \quad 0 \qquad\qquad\qquad , N \leq 32 \end{cases} \tag{4}$$

Where N is the signal length and is obtained by image frequency domain transformation. The number0.3936 and number 0.1829 is Sqtwolog threshold coefficient

After the hard threshold is obtained, the noise is demonised by the wavelet hard threshold method. The calculation formula is as follows:

$$w_\gamma = \begin{cases} w \ , |w| \geq \gamma \\ 0 \ , |w| < \gamma \end{cases} \tag{5}$$

4 TRAFFIC DETECTION BASED ON MASK-RCNN

Image segmentation marks and locates objects and backgrounds in an image according to features such as grayscale, color, texture, and shape, and uses these features to show similarity in the same region (Zhang, 2015).

4.1 RPN

The RPN network is a full convolution structure inside the mask-rcnn. It can use the sliding window to obtain the candidate region. Compared with the previous Faster R-CNN network, the RPN of the Mask R-CNN uses the bilinear interpolation method for the receptive field. To restore, the experience field here is the default 9 serial ports, corresponding to three areas of 128, 256, 512, the window ratio is 1:1, 1:2, 2:1.

According to the characteristics collected by different windows, the RPN uses the classification layer to calculate the coincidence rate between the target and the window. By setting the threshold (default is 0.7) to judge the difference between the calculated value and the true value, the positioning is completed.

4.2 Output Layer Regression

Mask R-CNN has two output layers, which are the result of detecting coordinates and the score of the detected object category. The difference between the two calculated results and the true value is brought into the loss function to make a gradient drop to obtain the optimal solution. The loss function is shown in Equation (6)-(9):

$$L(p, u, t^u, v) = L_{els}(p, u) + \ \omega[\omega \geq 1]L_{loc}(t^u, v) \tag{6}$$

$$L_{cls}(p, u) = -logp_u \tag{7}$$

$$L_{loc}(t^u, v) = \sum S(t^u - v) \tag{8}$$

$$S_L(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases} \tag{9}$$

Where L_{cls} is used to calculate the classification probability loss, which is a softmax loss function; L_{loc} is the coordinate loss function of the detection frame, where v_x,v_y,v_w,v_h are the actual measured values, respectively, and t_x,t_y,t_w,t_h are the predicted coordinates value.

4.3 Mask R-CNN Experimental Analysis

The image data collected by the drone is separated by key frames to obtain a series of key feature images, and then the data is pre-processed to perform image enhancement operations. The key frames of 2000 sample images are comprehensively selected to determine the traffic condition classification. For the pedestrian area, there is a detection of the direction of pedestrian movement.

The tensorflow framework is used steadily, the VGG16 model is selected with the training model, the training results of VGG16 on ImageNet are used for the pre-training parameters, and the RPN network and the classification network are used for simultaneous training. Under the NVIDIA GTX960 GPU acceleration, the actual detection model is trained and tested. At the same time, the sliding window is also used to intercept the scene image into the Mask R-CNN network for area recognition by the 600*600 area selection window, and finally the results generated by the multiple windows are merged into one picture. The final detection and recognition results are obtained by the method of maximum likelihood estimation. The above process is shown in Figure 2, and final result is shown in Figure 3.

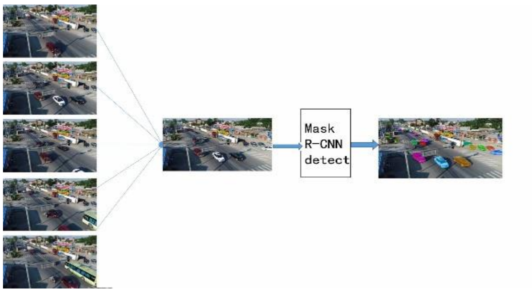

Figure 2: Key frame instance segmentation process

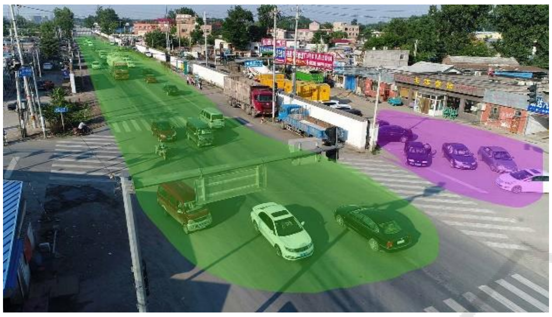

Figure 3: Output result

We classify the results into multiple continuous traffic areas, each of which is an entire category.

We compare our method and simply use the traditional method of instance segmentation and semantic segmentation as Table 1 and Table 2.

Table 1: Comparison between pixels.

Method	Pixel acc.	f.w. IU
FCNS	60.3	49.5
MASK-RCNN	75.6	48.0

Table 2: Comparison between average coefficients.

Method	Mean acc.	Mean IU
FCNS	42.2	34.0
MASK-RCNN	60.9	36.0

The test video is based on 1080p and when we choose 720p to test the work, the results is lower than 1080p and the result is shown in table 3.

Table 3: Comparison between 1080p & 720p in Mask-RCNN.

Resolution	Mean acc.	Mean IU
720p	57.9	33.0
1080p	60.9	36.0

5 CONCLUSIONS

In this paper, we have used the video stream keyframe extraction technique and the instance segmentation algorithm to identify the traffic area. The application of the work can be used in making traffic map and real-time traffic monitoring. The comparison from the pixel level is completely higher than the traditional semantic segmentation, and there are key frame extraction techniques. The support is reduced by nearly 2000 times the detection range per frame. From the comparison of the results of pixel mean, the robustness of the proposed algorithm can be reflected. Combining the above analysis, the algorithm proposed in this paper combined with the high-altitude shooting and pattern recognition of UAV is feasible and with accuracy.

REFERENCES

Deren, L. I., & Ming, L. I. (2014). Research advance and application prospect of unmanned aerial vehicle remote sensing system. Geomatics & Information Science of Wuhan University, 39(5), 505-513.

Ziyan W. (2013). Based on the UAV Orthophotos to analyse Land Use/Land Cover. Inner Mongolia Normal University.

Qiong W, & Songze L.(2018). Research on Identification Technology of UAV's Environmental Pollution Target based on Deep Learning. Environmental Science and Management, 43(07):91-94.

Ximei L. (2018). Using the Topographic Map for UAV Images Classification. Journal of Surveying and Mapping and Spatial Geographic Information, 41(11):46-48+54.

Chen Y, & Fan R. (2017). Segmentation of High-Resolution Remote Sensing Image Combining Phase Consistency with Watershed Transformation. Laser & Optoelectronics Progress, 54(9):092803.

Girshick R. (2015). Fast R-CNN. IEEE International Conference on Computer Vision.

Narasimha R ., & Savakis A E. (2003). A neural network approach to key frame extraction.

Kaiming, H. , Georgia, G. , Piotr, D. , & Ross, G. . (2018). Mask r-cnn. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1.

Chunyao Wang , & Junzhou Chen. (2014). Review on superpixel segmentation algorithms. Application Research of Computers, 31(1):6-12.

Moore, A. P., Prince, S. J. D., Warrell, J., Mohammed, U., & Jones, G. (2008). Superpixel lattices. IEEE Conference on Computer Vision & Pattern Recognition.

Shi J , Malik J . Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):888-905.

Boykov Y . (2000). Interactive Organ Segmentation using Graph Cuts. Proceedings of "MICCAI"-2000

Liu M Y , Tuzel O , & Ramalingam S. (2010). Entropy rate superpixel segmentation. Computer Vision & Pattern Recognition. IEEE.

Long, J. , Shelhamer, E. , & Darrell, T. . (2014). Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(4), 640-651.

Xia, X. , & Kulis, B. . (2017). W-net: a deep model for fully unsupervised image segmentation.

Von Eicken, T. , Basu, A. , Buch, V. , & Vogels, W. . (1995). U-Net: a user-level network interface for parallel and distributed computing. Fifteenth Acm Symposium on Operating Systems Principles. ACM.

Agarwala, A. , Hertzmann, A. , & Salesin, D. H. . (2004). Keyframe-based tracking for rotoscoping and animation. Acm Siggraph. ACM.

Yujin Zhang. (2015). A Survey on Transition Region-Based Techniques for Image Segmentation. Journal of Computer-Aided Design & Computer Graphics, 27(03):379-387.