ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

Detection Method of Offshore Droppers Based on UAV and YOLOv3

Chenlin Lin[1, a]

[1]Department of Traffic Information Engineering and Control, Shanghai Maritime University, China

[a]201730110018@stu.shmtu.edu.cn

Abstract

The marine environment is complex and changeable. Once an accident occurs, it is difficult to search and rescue the people on board. The longer the rescue time is, the lower the probability of survival. Aiming at this problem, this paper proposes a method for detecting marine drowning personnel based on UAV and YOLOv3. The image captured by the cameras on the Unmanned aerial vehicles (UAVs) is identified by YOLOv3 algorithm, which verifies that the designed algorithm can quickly detect the drowning personnel. Combining the GPS coordinates of the UAVs can provide accurate and reliable positioning information, which has practical significance for the subsequent rescue work.

Keywords

UAV; man overboard Detection; YOLOv3; maritime search and rescue.

1. INTRODUCTION

With the rapid development of marine tourism and transportation, accidents occur frequently. The drowning personnel show dynamic diversity with the direction of seawater rafting, which increase the difficulty of search and rescue. At the same time, immersion in sea water will cause the body temperature to drop faster. When sea water temperature is 10°C, it can generally only survive in water for 3.5 hours. Therefore, it is especially important to find the drowning personnel and carry out rescue work as soon as possible.

The existing maritime supervision system uses patrol boats and VTS (Vessel Traffic Management System), AIS (Vessel Automatic Identification System), CCTV (Video Monitoring System) and other monitoring systems to cooperate with the mode of cruise supervision, there are some shortcomings: such as short sight range, slow response, VTS radar omission of small carriers and non-intuitive, difficult to grasp the overall state. Potential, illegal ships can not be tracked continuously and effectively, and some illegal acts can not be continued to obtain evidence and deal with, and so on. The traditional mode is characterized by large workload, high cost and low efficiency.

UAVs have the characteristics of being unmanned, portable, small, rapid landing and adaptable to climatic conditions, which make them very suitable for carrying out various search and rescue tasks, including maritime search and rescue. Compared with the traditional marine search and rescue tools such as ships, helicopters and motorboats, UAVs have the advantages of low cost, good mobility and zero casualties. In an environment with poor climatic conditions or serious surface pollution, UAVs can replace rescuers to carry out high-risk search and rescue tasks, and detect and locate the drowning personnel for the follow-up rescue work. When searching and rescuing at sea, the image is collected by UAV airborne camera as the front-end,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

and the video stream information is transmitted to the back-end workstation. The real-time image is assisted by the target detection algorithm, and judges whether there are targets to be rescued, so the performance of detection algorithm used in this process is very important.

Search and rescue task is a long-term and repetitive work, which will weaken people's attention to a certain extent. At the same time, people's visual attention can't focus on the target in the global perspective. The application of computer vision can make up for the shortcomings of the above people in search and rescue task. Therefore, the use of computer vision to assist staff in search and rescue missions is an area of concern.

The task of target detection is to determine whether there is an interested object in the image, and then precisely locate the interested object. At present, target detection mainly includes machine-based learning method and depth-based learning method. Traditional target detection methods based on machine learning first use sliding windows to determine the detection area, and then extract target features. The commonly used features are Haar, SIFT, HOG, etc. Finally, classifiers such as SVM, Adaboost are used for classification. However, such methods require high feature extraction, and the region selection strategy based on sliding window is not targeted, which leads to high time complexity and poor anti-noise performance. Target detection method based on deep learning has been greatly improved the above shortcomings, and has been widely used in recent years. R-CNN, as a milestone of convolution neural network, has entered the field of target detection and recognition. Because of the complexity of the algorithm, the detection time is longer. The detection accuracy of SPP NET, Fast R-CNN [1], Faster R-CNN and other algorithms has been improved continuously, and the speed has been improved significantly, but it is still unable to achieve real-time detection. In order to balance the speed of the algorithm and the accuracy of detection, some researchers have proposed regression-based detection methods, including SSD[2], YOLO[3], YOLOv2[4] and YOLOv3[5]. Among them, YOLOv3 integrates the advantages of deep learning models such as Faster R-CNN, SSDand ResNet on the basis of YOLO algorithm. It is the most balanced target detection algorithm with speed and precision. When the image size is 320 x 320, its mAP value reaches 28.2 and its running speed is 22 Ms. With the same detection accuracy as SSD, YOLOv3 is three times faster and more suitable for real-time application scenarios.

In this paper, a method for detecting man overboard based on UAV and YOLOv3 is proposed. Firstly, the image captured by camera on UAV is preprocessed, and the preprocessed image is transformed into training data set as input of network. Secondly, it uses the darknet-53 network for iterative training and generates a weight model to extract the characteristics of the drowning personnel. In the final application stage, the UAV as the front-end incoming video stream passes through the back-end workstation, and the weight model is used to perform real-time scanning detection on the falling water area, and the GPS coordinates of the drone are combined to determine the exact location of the target to be rescued.

2. YOLOV3

2.1. YOLOv3 Algorithm Principle

YOLOv3 is an end-to-end target detection algorithm based on deep learning. It uses the whole graph as the input of the network, and directly returns the target borders and categories located in each position of the image. The YOLOv3 algorithm first divides the input image into S*S cells. Each grid contains B detection bounding boxes. It extracts features using convolution neural network and finally outputs the location of the boundary boxes and calculates the confidence score of the target in each boundary box. At the same time, the category information of each grid is predicted.

YOLOv3 continues the YOLOv2 method, which also predicts each boundary box on feature map. Each boundary box contains five values: x, y, w, h and confidence. As shown in Figure 1, (tx,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

ty) is the center point of the prediction boundary box, (tw, th) is the width and height of the prediction boundary box, (cx, cy) represents the distance between the grid and the upper left corner of the image. According to formula (1), the output of the target boundary box is (bx, by, bw, bh).

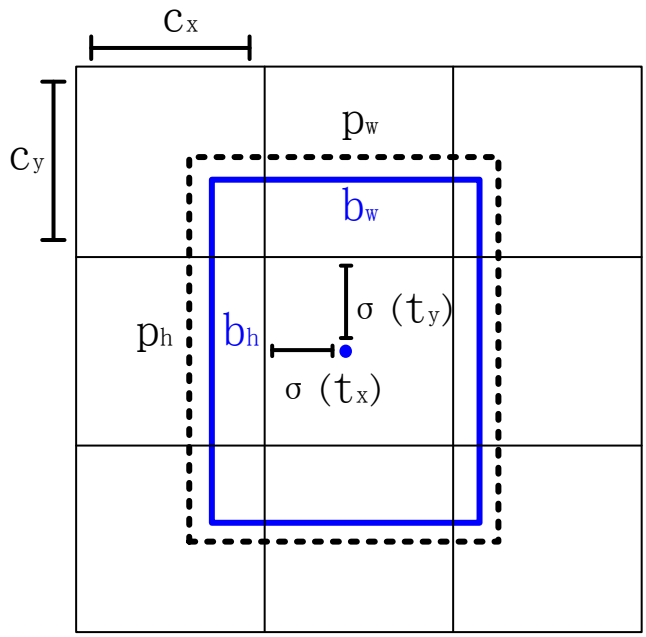


Fig 1. Bounding boxes with dimension priors and location prediction.

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \tag{1}$$

The confidence of the target in the boundary box can be obtained by formula (2), Pr(object)is used to judge whether there is a target in the corresponding grid of the prediction boundary box. If there is a target, the value is 1, otherwise the value is 0. IOU_{pred}^{truth} is used to predict the ratio of the intersection and union area of the boundary box to the real boundary box, see formula (3). If there are targets in the mesh ,and the overlap degree between the predicted boundary box and the real boundary box is greater than any other boundary box, the boundary box fraction is 1. If the overlap degree is greater than 0.5 but not the largest, the prediction is ignored.

$$\text{object}\_conf = P_r(\text{object}) \cdot IOU_{pred}^{truth} \tag{2}$$

$$IOU_{pred}^{truth} = \frac{\text{area}(\text{box}(\text{Pred}) \cap \text{box}(\text{Truth}))}{\text{area}(\text{box}(\text{Pred}) \cup \text{box}(\text{Truth}))} \tag{3}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

2.2. Multi-Scale Prediction

In order to adapt the model to multi-scale input images and improve the detection accuracy, YOLOv3 algorithm uses three different scale feature maps for prediction, and sets three priori boxes for each scale. A priori box is a candidate area box whose size and aspect ratio are fixed. The setting of a priori box will affect the accuracy and speed of recognition. In order to get the appropriate priori box, the K-means [6] clustering method is used to cluster the target boundary box of the training data set. The width and height of each boundary box are clustered with respect to the width and height of the whole picture. The nine cluster centers are (10×13), (16×30), (33×23), (30×61), (62×45), (59×119), (116×90), (156 × 198), (373 × 326), are equally divided into three scale features, each scale contains three cluster centers.

2.3. Network Architecture

Unlike the VGG-16 used in YOLOv1 and darknet-19 used in YOLOv2, YOLOv3 uses darknet-53[7] network architecture to extract image features. As shown in Figure 2, the network includes 52 consecutive 3×3 and 1×1 convolution layers and one fully connected layer. There is no pooling layer. Tensor size transformation is achieved by increasing the step size of the convolution core. Residual connections are widely used in the network to increase the depth of the network while avoiding over-fitting in the training process. The deepening of the network improves the quality of features and thus improves the effect of target detection.

Type	Filters	Size		Output	
Convolutional	32	3	3	256	256
Convolutional	64	3	3/2	128	128
Convolutional	32	1	1		
Convolutional	64	3	3		
Residual				128	128
Convolutional	128	3	3/2	64	64
Convolutional	64	1	1		
Convolutional	128	3	3		
Residual				64	64
Convolutional	256	3	3/2	32	32
Convolutional	128	1	1		
Convolutional	256	3	3		
Residual				32	32
Convolutional	512	3	3/2	16	16
Convolutional	256	1	1		
Convolutional	512	3	3		
Residual				16	16
Convolutional	1024	3	3/2	8	8
Convolutional	512	1	1		
Convolutional	1024	3	3		
Residual				8	8
Avgpool		Global			
Connected		1000			
Softmax					

Fig 2. Darknet-53

2.4. Loss Function

Loss function is a quantitative function to measure the degree of solution error from a statistical point of view. It is an important basis for judging the quality of the algorithm. It can

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

obtain the error between the predicted value and the real value. A good loss function can define a clear quantitative objective for the algorithm, and improve the training speed of the model. The loss function used by YOLOv3 combined with the principle of the algorithm considers the error of boundary frame coordinates, the confidence of objects in the boundary frame and the mesh category, as shown in formula (4). Among them, λ coord is the weight coefficient of coordinate prediction, λ noobj is the penalty coefficient of confidence degree when no object is included. $\prod_{ij}^{obj}$ used to determine whether there is a target in the grid. xi, yi, wi and hi respectively represent the abscissa, ordinate, width and height of the center point of boundary frame in the i-th grid. Ci is the confidence of the presence of the target, and p(c) is the probability of the target category.

$$\begin{aligned} \text{loss} = \lambda_{coord} & \sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{I}_{ij}^{obj}(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\ & + \lambda_{coord} \sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{I}_{ij}^{obj}\left(\sqrt{\omega_i} - \sqrt{\widehat{\omega}_i}\right)^2 + \left(\sqrt{\mathrm{h_i}} - \sqrt{\hat{\mathrm{h}}_\mathrm{i}}\right)^2 \\ & + \sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{I}_{ij}^{obj}\left(C_i - \hat{C}_i\right)^2 \\ & + \lambda_{noobj} \sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{I}_{ij}^{obj}\left(C_i - \hat{C}_i\right)^2 + \sum_{i=0}^{S^2} \mathbb{I}_{i}^{obj} \sum_{c\in classes} \left(p_i(c) - \hat{p}_i(c)\right)^2 \end{aligned} \tag{4}$$

3. EXPERIMENTAL PROCESS

3.1. Experimental Platform

This experiment is completed in Linux environment. The operating system is Ubuntu 16.04 64 bits, and Darknet's deep learning open source framework is used as the algorithm implementation platform. Therefore, CUDA9.0, cudnn 7.1, Opencv3.4, Python 3.5 and other third-party libraries are installed to support Darknet's operation. The computer has 16.0GB of memory and Intel (R) Core (TM) i7-8750H CPU@2.20Ghz processor. The graphics card uses NVIDIA GeForceGTX1060, the display memory type is GDDR5, the capacity is 6GB, and the core frequency is 1620-1847 MHz.

3.2. Data Set

The data set mainly chooses a large number of images taken by UAVs in different scenarios. The shooting angles and shooting backgrounds are different, which provides support for good generalization of the model. Firstly, image enhancement method is used to pre-process the samples. Because the attitude of UAV is not fixed during flight, some samples are rolled and rotated, which accords with the actual situation and increases the sample size. Then the images are labeled with the image labeling tool called"Wizard labeling assistant" and stored as an XML file in pascal_voc format. The XML file is transformed into a TXT file in < label, X, Y, W, H> format by format conversion script. The experiment produced about 2380 data, totaling 3400 labeling features. The main objects of the labeling were landing personnel, life jackets and life buoys, which were divided into three categories: training, validation and testing, with a ratio of 8:1:1.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

3.3. Yolov3 Network Training

When training, the learning rate adopts a step-by-step strategy. The initial value is set to 0.0001 because too much learning rate will lead to divergence. The optimization algorithm uses Momentum gradient descent algorithm, the momentum parameter is set to 0.9, the sample size of each iteration is 128, that is, the parameters of each 128 samples are updated once, and the descending parameter of learning rate with the number of iterations is 0.0005. After 42,000 iterations, the network finally converges and the loss value is 0.060790. Figure 3 shows the change curve of the loss value during the network training process. It can be seen from the graph that the loss value decreases sharply at the beginning. With the continuous iteration, the decline decreases gradually and eventually converges. When the number of iterations is 10 000, the parameter changes are very stable. Figure 4 shows the average IOU change curve. Contrary to the loss value, the IOU rises sharply at the beginning, as the iteration progresses, the gain decreases and eventually converges to a value close to 1.

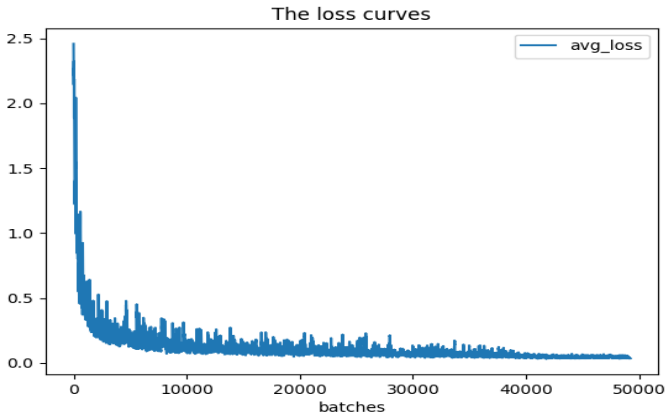


Fig 3. Average loss value curve

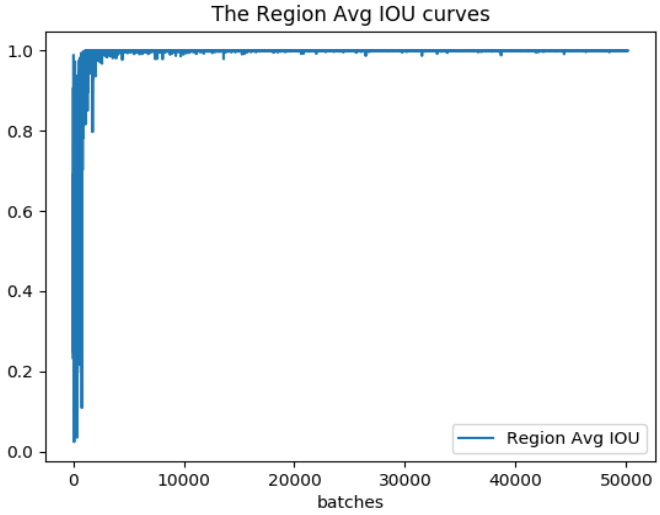
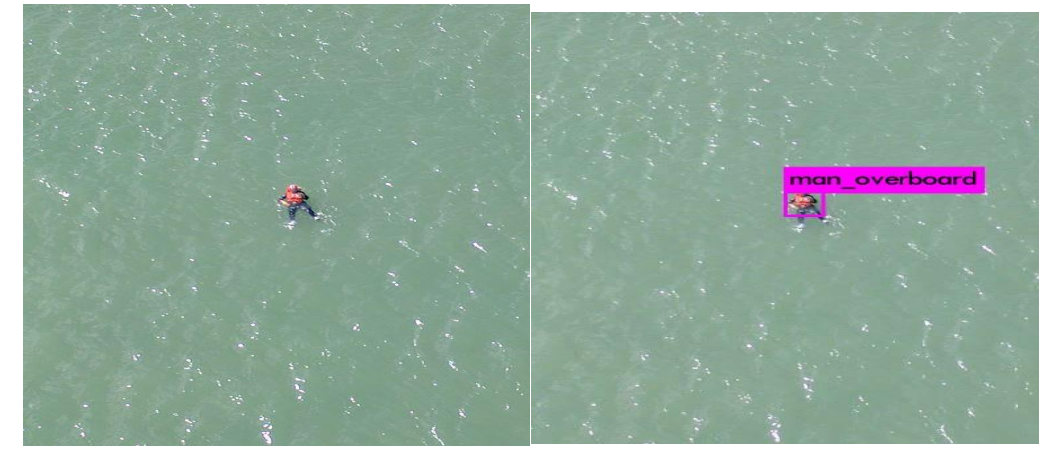


Fig 4. Average IOU change curve

4. RESULT ANALYSIS

The test results are shown in Fig. 4, where (a), (c), (e), (g) are the original images taken by UAV camera, (b), (d), (f), (h) are the results of the method proposed in this paper. It can be found that YOLOv3 algorithm can accurately identify the people falling into the water in the image, and detect the non-people falling into the water who have some interference in the background.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig 5. Man overboard test results

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0032

5. CONCLUSION

This paper proposes a method of inspecting man overboard based on UAV and YOLOv3. The method combines deep learning with inspecting man overboard, extracts image features by darknet-53 network, sets the best priori frame by K-means clustering method, and directly regresses the location of the target in the graph, and determines the category. The experimental results show that YOLOv3 algorithm has high recognition accuracy and detection speed of 30 fps, which meets the real-time requirements. The method proposed in this paper has good performance and can be used for UAV detection and rescue of people falling into water at sea, but there is still much room for improvement. At present, the related work of using UAV to search and rescue at sea has not been popularized, the data collected on the network is less, and most of the images are taken in a good environment. In real life, the environment of the people falling into the water is more complex, so the detection under bad weather conditions or poor lighting conditions needs to be studied in depth. And verify it.

REFERENCES

[1] Ren S, He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137-1149.

[2] Liu W, Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. 2016.

[3] Redmon J, Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection [J]. 2015.

[4] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. 2016.

[5] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018.

[6] Antic B, Letic D, Culibrk D, et al. K-means based segmentation for real-time zenithal people counting[C]// IEEE International Conference on Image Processing. 2009.

[7] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.