Real-time detection method of surface floating objects based on deep learning

ZOU Shanhua[1,2], PENG Li[1,3,4], FANG Ning-sheng[1,2]

（1. Jiangsu Key Construction Laboratory of IOT Application Technology ，Wuxi 214064，China；
2. School of Internet of Things Engineering ，Wuxi Taihu University， Wuxi 214064，China；
3. IOT Application Technology Ministry of Education Engineering Center， Wuxi 214122，China；
4. School of Internet of Things Engineering，Jiangnan University， Wuxi 214122，China）

Abstract: In order to solve the problem that the surface floating object needs manual inspection and time-consuming and labor-intensive problems, and the information of single monitoring means is not comprehensive, this paper proposes a set of integrated monitoring and detection system, which can monitor the video image information in various scenarios. Automatically alert and dispose of. Based on the video surveillance-based surface floating object detection algorithm, the darknet framework is used to establish a deep learning network, and the improved YOLOv3 detection algorithm is designed to solve the problem that the garbage floating on the fast flowing water surface and the algae and other pollutants cannot be quickly identified.

Keywords: YOLOv3; detection; real-time; deep learning

1 Introduction

Nowadays, in the new era of wide spread of data and information, rapid development of science and technology, information products emerge in endlessly in daily life and learning, and traditional life style can not meet the needs of rapid development of human society. As a result, the traditional material resources and human resources are gradually informationized by the changing science and technology, and can be reasonably distributed and used. There are many problems in the method of manual inspection, such as laborious, time-consuming, high cost and low efficiency. However, a single video monitoring or sensor monitoring inspection method can not deal with the complex and changeable actual situation of the site, can not control the actual production in a real and effective way, and it is easy to misjudge or miss the judgment when the pollution exceeds the standard, water conservancy disaster or crop production environment is abnormal, so it is difficult to make an emergency Effective response and disposal.

For example, the environmental protection industry can't monitor the pollution of the garbage, cyanobacteria and other floating objects on the lake surface through the sensor alone. In order to overcome the shortcomings of the existing technology, solve the problem of time-consuming and labor-intensive manual inspection and the lack of comprehensive information of a single monitoring means, a set of video image information that can monitor various scenes is designed, and once there is an exception, it can automatically alert and handle it.

Based on the establishment of deep learning network by using the framework of Darknet, this paper designs an improved yolov3 detection algorithm [1], which solves the problem that the pollutants such as garbage and algae can not be quickly identified on the surface of fast flowing water.

2. Algorithm of yolov3 based on Darknet

Yolov3 algorithm is a target detection algorithm based on deep convolution network. The algorithm has the following advantages:

1) Its detection speed is very fast and meets the requirements of real-time detection;

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00053

2) When using this algorithm, target detection is regarded as a regression problem, and the prediction results can be obtained by dropping the image into a simple convolution neural network, as shown in Figure 1;

3) It can learn the general features of objects in the input image, which are enough to represent the category, so it has strong generalization ability.

Therefore, yolov3 algorithm is used to detect floating objects.

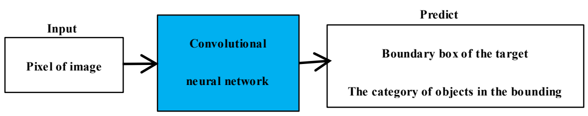

Figure 1 advantages of yolov3 algorithm

Yolov3 algorithm first divides the input $S*S$ image into a grid. If the network recognizes a grid in which the center of an object is located, the grid will be responsible for detecting the object. Each grid also predicts the probability of all possible categories of the target object $P_r\left(Class_i \mid Object\right)$ *by calculating the size, and obtains the confidence of the bounding* B *boxes. The resulting bounding box has five outputs* (x, y, w, h, c), respectively. As the central coordinate of the x, y bounding box, it should be aligned with the grid responsible for predicting the bounding box, that is, it has an offset from the grid to which it belongs, so it will be fixed between 0-1 by the normalization method. w, h is the width and height of the bounding box. In order to maintain the consistency of the data, it is normalized: divide the width and height of the bounding box by the width and height of the input image, and the width and height of the normalized bounding box are between 0-1. In order to predict the confidence degree of the target object in the bounding box, c the calculation formula is as follows:

$$c = P_r(object) * IOU_{pred}^{truth} \quad (1\text{-}1)$$

In the formula, IOU (intersection over union) represents the overlapping degree of predicted boundary box and correctly marked object boundary box [2].

$$IOU = \frac{A \cap B}{A \cup B} \quad (1\text{-}2)$$

Where A is the predicted bounding box, and B is the correct marked bounding box. If a grid does not detect the target object, the confidence c value is 0; if the grid detects the target, the confidence c value is the IOU value of the predicted bounding box and the correct marked bounding box.

The loss function of Yolo algorithm is defined as:

$$\begin{aligned}
&\lambda_{\mathbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&+ \lambda_{\mathbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}$$

(1-3)

Mean square and error are used to integrate coordinate error and classification error of bounding box. If the proportion of coordinate error and classification error is equal, the training will be divergent. Because a large part of the mesh does not contain the target object, the confidence c value of the prediction boundary box of these meshes will be 0. Therefore, different weights must be set to increase the proportion of position error, that is $\lambda_{coord} = 5$; At the same time, reduce the proportion of confidence loss of bounding box without target, that is $\lambda_{noobj} = 0.5$ .The weight of confidence loss of the bounding box containing the target is normally 1.

In the loss function represented by formula (1-1), the first two lines are coordinate errors, the first line is the error of the central coordinate (x, y) of the bounding box, and the second line is the error of predicting the width and height of the bounding box.In this paper, the quadratic root of width and height is used, and the reason why height and width are not used directly is that under the condition of the same error value of height and width, the accuracy of small target is affected much more than that of large target, which can be reduced by using the quadratic root. The third behavior is the loss of confidence in the bounding box, including two cases: the grid contains and does not contain the target object. The loss is not always to be calculated, because each mesh contains two bounding boxes, so this item is only calculated when the intersection ratio of the bounding box of the object correctly marked and one of the bounding boxes predicted by the mesh is the largest. The last behavior predicts the error of the category to which it belongs.

In the test phase, the probability that the predicted bounding box belongs to each category and the information of the accuracy of the bounding box position are obtained. The formula is as follows:

$$P_r(Class \mid Object) * P_r(object) * IOU_{pred}^{truth} = P_r(Class_j) * IOU_{pred}^{truth} \tag{1-4}$$

Yolov3 adopts the network structure of darknet-53, uses the residual module to build the depth model and alleviates the problem of gradient disappearance. It uses three different scale feature maps to improve the detection ability of different size targets [3], and the network structure is shown in Figure 2.

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

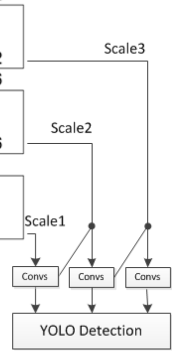

Figure 2 yolov3 network framework

3. The realization of floating object detection algorithm

3.1 algorithm implementation

This paper presents a real-time detection method based on yolov3. This method is implemented from the following points:

1) Image data preprocessing: collect the detection pictures of floating objects on the water surface, at the same time, in order to improve the generalization ability of the network, enhance the data of the collected pictures, including the image turning, translation, zooming and other processing, increase the data amount, and then manually label each picture, including box selection of the target to be tested and confirm the target category [4].

2) Darknet-53 network training: input the picture into the darknet-53 deep convolution framework, and the main parameters include the learning rate of 0.0001 and the maximum number of batches (max_ Batches) is 10000, batch size_ Size) is 16, momentum is 0.7, and finally a network structure with good fitting ability can be obtained.

3) Detection of floating objects on water: read the video to be detected, input the image into the trained network structure frame by frame [5], because yolov3 algorithm directly outputs the target category and coordinate position information by regression,so each frame image also directly displays whether there is floating objects on water and selects the position of floating objects in time frame, so as to realize end-to-end detection. The confidence parameter set during detection is 0.5, and the parameter used for maximum suppression filter target box is 0.4. The overall detection process is shown in Figure 3:

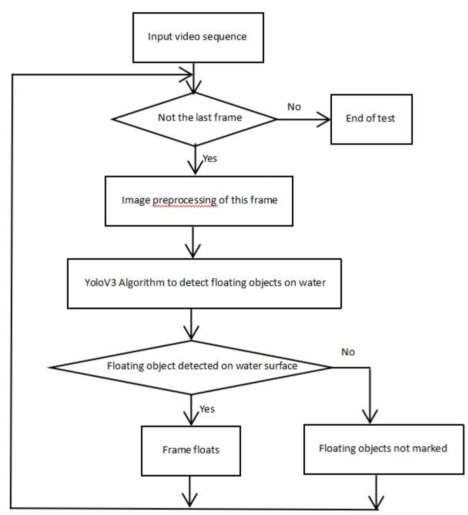

Figure 3 detection flow chart

4 Conclusion

In order to solve the problem that the waste floating on the fast flowing water surface, algae and other pollutants can not be quickly identified. The proposed algorithm based on Darknet and yolov3 can effectively overcome the influence of weather light and other complex environment on detection, and can quickly identify and detect floating objects on the water surface, frame them and give early warning. The algorithm is fast and robust. The above work verifies the feasibility and efficiency of using yolov3 to detect floating objects on the water surface, and the follow-up work can be optimized for data collection.

reference

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

[2] Benjdira B, Khursheed T, Koubaa A, et al. Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3[J]. 2018.

[3] Cui Z, Heng L, Yeo Y C, et al. Real-Time Dense Mapping for Self-driving Vehicles using Fisheye Cameras[J]. 2018.

[4] Research on image target recognition based on deep learning [D]. Harbin Engineering University, 2018

[5] Qiu Peng, Deng Xiuhui, Huo Ying. An anomaly detection method based on deep learning in complex environment [J]. Science and technology and engineering, 2018, 18 (10): 231-234