Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 187 (2021) 52–59

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Identification, Information and Knowledge in the internet of Things, 2020

Optimization of Underwater Marker Detection Based on YOLOv3

Ning Jiang, Jinlei Wang, Linghui Kong, Shu Zhang, Junyu Dong*

Ocean University of China

Abstract

The research on the detection and recognition technology of marker is of great significance for some underwater operations, such as marine resource exploration, underwater robot operation and so on. The existing image processing methods can effectively detect and recognize the markers in the air. Nevertheless, in the underwater environment, the complex imaging environment of the ocean leads to serious degradation of underwater images obtained by the optical vision system. Due to the lack of effective information for object recognition, the severely degraded underwater images increases the difficulty of detection and recognition of underwater objects. With the development of high-tech underwater imaging equipment, the quality of underwater images has been improved to a certain extent, but there are still some phenomena such as color fading, low contrast and blurred details. Solutions to overcome these problems are important for the exploration of the ocean. In this paper, we introduce a deep learning model to optimize the performance of detection, and make a unique marker dataset for the application scene of our experiment. We first use the deep learning network to pre-train the marker images in the air. Next, we use the underwater marker images for fine-tuning. Finally, after the target marker is detected, the traditional image processing method is used to recognize the marker. Experimental results show that the optimization method we proposed achieves better performance on the dataset.

Keywords: Computer vision; Underwater marker detection; Deep learning

1. Introduction

The detection, location and recognition of markers are very important for 3D measurement, 3D reconstruction and camera positioning. In the air, after processing the image with the traditional recognition algorithm, the marker can be well recognized.

The research on the detection and recognition technology of marker is of great significance for some underwater operations, such as marine resource exploration, underwater robot operation and so on. The detection and recognition of the marker provides the location information of the underwater object. According to the location information of

* Junyu Dong. Tel.: +86 0532-66782300.
E-mail address: dongjunyu@ouc.edu.cn

10.1016/j.procs.2021.04.106

the object, researchers can monitor and track the object. Since light will be absorbed and scattered when travelling in water, underwater imaging exists three major difficulties, including color cast, under-exposure, and fuzz, resulting in blurred captured images. The traditional image processing method has a low detection rate for underwater markers, which makes it impossible to detect the markers and thus affects the recognition of the markers.

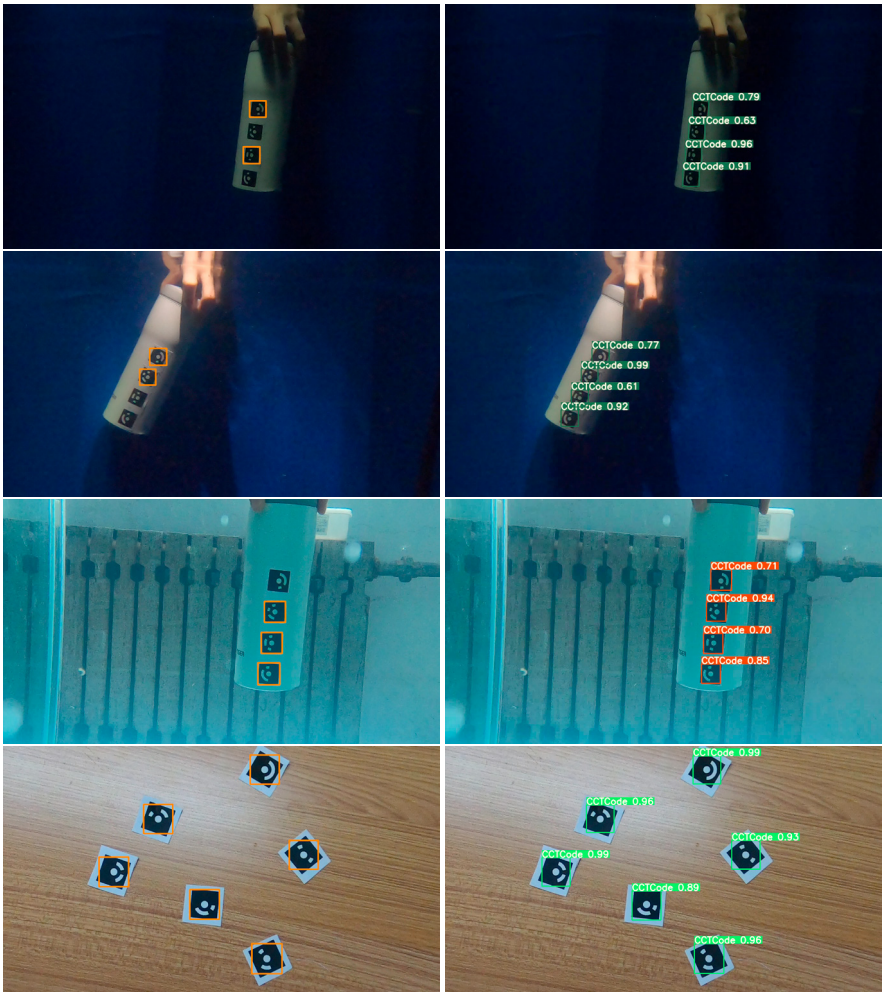

Fig. 1. The images are randomly selected from our test dataset. The left column is the performance of traditional image processing, and the right column is the performance of YOLOv3. The images in the top three rows are taken underwater, the images in the bottom row are taken in the air.

Most of the current image object detection methods are oriented towards clear images in atmospheric environment, which is difficult to adapt to the characteristics of underwater images, resulting in low accuracy of underwater image object detection. In order to solve this problem, most of the current methods are based on pre-processing techniques, such as underwater image enhancement [1], [2] and underwater image restoration [3] to improve the quality of underwater image, or to improve the accuracy of moving object detection by introducing motion information in video sequence [4].

For areas or objects that are relatively easy to detect, image enhancement has little effect on detection; the performance degradation caused by image enhancement occurs in the area with low detection accuracy in the original image. For areas or objects that are difficult to detect, the image quality is usually low (low contrast, strong fog or color deviation). After image enhancement of these regions, the object detector will bring more false positive, thus reducing the recall and therefore reducing the mAP.

Therefore, we introduce a deep learning network to improve the accuracy of underwater object detection, and make a unique marker dataset for the application scene of our experiment.

Based on the YOLOv3 [19] object detection algorithm, we first use the deep learning network to pre-train the marker images in the air. Next, we use the underwater marker images to make fine-tuning, so as to improve the robustness of underwater detection. Finally, after the target marker is detected, the traditional image processing method is used to recognize the marker.

2. PROPOSED METHOD

2.1. Traditional object detection algorithms

Most of the early target detection algorithms are based on manual features. Due to the lack of effective image feature expression methods before the birth of deep learning, people have to design more diversified detection algorithms to make up for the defects of manual feature expression ability. At the same time, due to the lack of computing resources, people have to find more sophisticated computing methods to accelerate the model.

Viola Jones (VJ) detector proposed by Viola P and Jones M [5] and [6] realized real-time face detection for the first time with extremely limited computing resources more than a decade ago. The speed is tens or even hundreds of times faster than that of the detection algorithm at the same time, which greatly promoted the commercialization of face detection applications. The VJ detector uses the most traditional and conservative object detection method −− sliding window detection. This idea seems simple, but in fact it costs a lot of computation.

The hog feature proposed by Dalal N et al. [7] can be regarded as another important improvement on the basis of histogram features of gradient direction, and it is the basis of all object detectors based on gradient feature. Hog detector follows the original idea of multi-scale pyramid and sliding window detection. In order to detect objects of different sizes, the size of detector window is usually fixed and the image is scaled successively to construct a multi-scale image pyramid.

The deformable part based model (DPM) was first proposed by Felzenszwalb P et al. [8], and later improved by his Ph.D. student Girshick R B et al. [9], [10], [11] and [12]. DPM is the peak of the development of detection algorithms based on classic manual features. The main idea of DPM can be simply understood as a process of ”from the whole to the part, and then from the whole to the whole”.

Because there are not many markers types in our marker dataset, and all of them have some common features, we design a unique detection algorithm by using simple image processing methods.

Contour detection is a useful technique for shape analysis and object detection and recognition. A contour is a closed curve joining all the continuous points having some color or intensity, they represent the shapes of objects found in an image. Contours are abstract collections of points and segments corresponding to the shapes of the objects in the image.

As a result, to successfully detect markers in an image, firstly we convert the image to a binary image, it is a common practice for the input image to be a binary image. Then, finding the ellipse contours using findContours() OpenCV function. Next, we can basically find the region where the marker is located through the multiple relationships of the circle radius, and then through affine transformation, the fitted ellipse contour is transformed into a regular circle. Finally, we need to make a judgment on the found area to see whether it is a real marker. In this way, we get the whole marker, and we can get the information carried by the marker through decoding. [21]

In the air, after processing the image with the traditional image processing methods, the marker can be well detected and recognized, as shown in Fig. 2. But, as shown in Fig. 3, the traditional image processing is not ideal for the recognition of underwater markers.

2.2. Deep learning model

Object detection is one of the basic tasks in the field of computer vision, which has been studied for nearly 20 years. With the rapid development of deep learning technology in recent years, the object detection algorithm has also changed from the traditional algorithm based on manual features to the detection technology based on deep neural network.

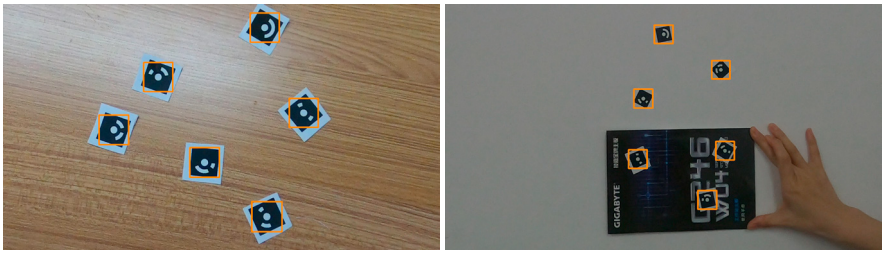

Fig. 2. Performance of traditional image processing methods in air.

Fig. 3. Performance of traditional image processing methods in water.

The task of object detection is to find the objects of interest in the image or video, and detect their positions and sizes at the same time. There are many uncertain factors in the process of object detection. For example, the number of objects in the image is uncertain, the objects have different appearances, shapes, postures, and the interference of illumination and occlusion and other factors in the imaging of objects lead to certain difficulty in the detection algorithm. Since entering the era of deep learning, the development of object detection mainly focuses on two directions.

One is two-stage algorithms based on region proposal, including R-CNN [13], Fast R-CNN [14], Faster R-CNN [15], etc. Firstly, the algorithms need to generate the object candidate frame, that is, the object location. Then they classify and regress the candidate frame. The other type is one-stage algorithms such as YOLO [16], [18], [19] and SSD [17], which only use a convolutional neural network to directly predict the categories and positions of different objects. The first type of method is higher in accuracy, but slow. The second type of algorithm is faster, but the accuracy is lower.

The marker detection task does not have many object categories, so it is easier to detect. In addition, considering the performance reasons, we choose YOLOv3 as the deep learning model.

YOLO [16] frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork. The system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are marked as an $S \times S \times (B * 5 + C)$ tensor.

Compared with v1 version, YOLOv2 [18] has improved from three aspects: better, faster and stronger. YOLO9000 predicts detections for more than 9000 different object categories, all in real-time.

The model of YOLOv3 [19] is much more complex than the previous model, as shown in Fig. 4. The speed and accuracy can be balanced by changing the size of the model structure.

Compared with YOLOv3, YOLOv4 [20] has achieved a very obvious speed and accuracy improvement. But YOLOv4 just combines the methods proposed in recent years for other models with YOLO.

We choose YOLOv3 because YOLOv3 is a more popular network used in industry nowadays, and countless lightweight networks are based on YOLOv3. The main reason for using YOLOv3 is its simple structure and easy to understand.

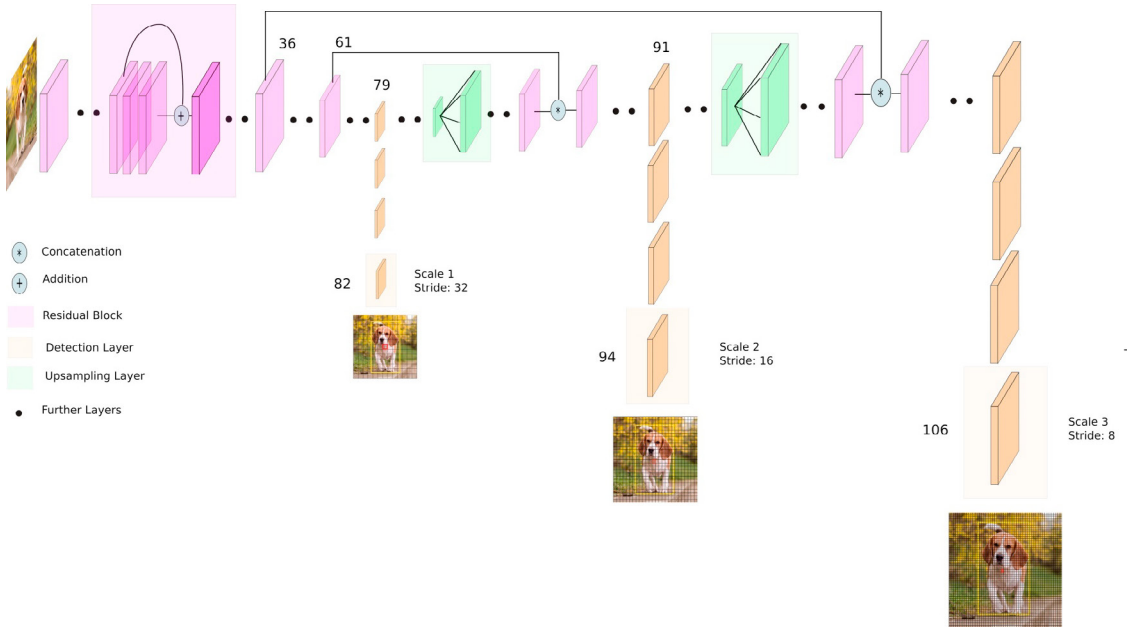

Fig. 4. This is the architecture of YOLOv3.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Dataset

There are many application scenarios for markers, and different markers will also have different application scenarios. For example, researchers can use markers to locate and identify the joints of robot arm, so as to know its motion state.

There are many kinds of markers. We design a kind of ingenious marker for our experimental scene. As shown in Fig. 5, it uses simple black and white color to increase the robustness of contour extraction, and uses simple graphic design to facilitate the identification of marker. Specifically, based on the radius (r) of the innermost white solid circle, respectively construct three circles with r, 2r, and 3r as the radius. The first circle is the innermost white solid circle, the second circle is the circle inside the white ring, and the third circle is the circle outside the white ring. It can be seen that this marker has some common features, which provide a basis for subsequent detection and recognition.

When drawing the marker, we draw based on the angle and the corresponding binary code. We use the outermost ring to perform binary coding. For example, the ring is divided into 12 parts, that is, each unit angle is $\frac{2\pi}{12}$, and our binary code bit is 12 bits. Because of the symmetry of circles, we only keep the only markers. On the contrary, we can use the inverse process to decode and get the encoded information contained in the mark. [21]

Therefore, we produce a unique dataset, including training dataset and test dataset. Our dataset has been carefully constructed manually. The data and labels are all artificial, and the pictures are diverse and innovative. There are a total of 4372 images in the training dataset, 2480 in the air and 1892 in the water. The test dataset consists of 407 images to test the accuracy of the detection algorithm based on traditional image processing and the detection algorithm based on deep learning in different environments in air and water respectively.

3.2. Implementation Details

This experiment is implemented on a linux Ubuntu 16.04 with two 2080Ti GPU, running on Pycharm. The images resolution is 1920×1080.

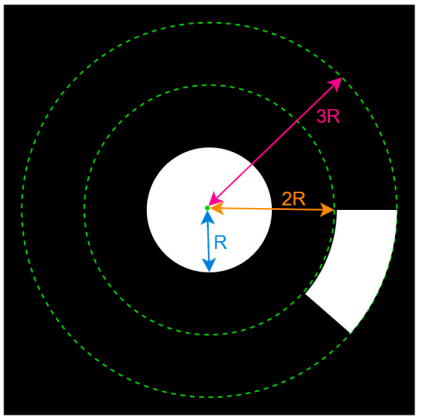

Fig. 5. The design of marker.

3.2.1. The first stage

We use 2480 pictures captured in the air for pre-training. Fig. 6 is part of the dataset.

Because the traditional recognition algorithm has a good effect on the recognition of the marker in the air, we first used the traditional algorithm to detect the marker, then filter and label the data to get the dataset we need for training. Trained 50 epochs.

Fig. 6. Data in different scenes in air.

3.2.2. The second stage

Add 1892 pictures taken underwater to fine-tune the model. There are 4372 pictures in total. Fig. 7 is part of the dataset.

The traditional recognition algorithm is difficult to identify underwater marker in a complex environment. We have photographed the images with diverse environment and background. First, we preprocess the data with traditional detection methods, and then filter the data. Next, we manually label the non-conforming data to form the underwater dataset. Trained 50 epochs.

Fig. 7. Data in different scenes in water.

3.3. Quantitative and Qualitative Results

3.3.1. Quantitative Results

As shown in Table 1, we quantitatively compare the YOLOv3 with traditional image processing methods. It can be seen that the accuracy of YOLOv3 underwater detection performance in complex environments is much better than the traditional image processing methods, and the detection performance in the air is comparable to the traditional image processing methods.

Table 1. Accuracy of YOLOv3 and traditional image processing methods

	Traditional image processing methods	YOLOv3
Dark underwater	75.4%	92.9%
Dark underwater + light	79.2%	99.5%
Light underwater	89%	99%
In the air	99.5%	99.5%

3.3.2. Qualitative Results

We randomly selected four images containing markers from the test dataset for qualitative evaluation. As shown in Fig. 1, YOLOv3 achieves excellent performance for solving light and shade, complex environments.

4. CONCLUSIONS

In this paper, we propose an effective optimization method for detecting marker in many kinds of environments. Especially underwater, the detection algorithm based on deep learning is far superior to the traditional image processing methods. And we make a unique marker dataset for the application scene of our experiment, which also can be applied to other scenarios.

Acknowledgements

The authors would like to thank the members of the vision laboratory of Ocean University of China who participated in this work. This work was supported by School of Information Science and Engineering, Ocean University of

China. This work was supported by National Natural Science Foundation of China (NSFC) (No.U1706218,41927805), the Fundamental Research Funds for the Central Universities (201964022) and Discipline Development Strategy Research of Academic Divisions of the Chinese Academy of Sciences XK2018DXC002 and NSFC L1824025.

References

[1] S. Zhang, T. Wang, J. Dong, and H. Yu. Underwater image enhancement via extended multi-scale Retinex[J]. Neurocomputing, 2017.
[2] P. Guo, D. Zeng, Y. Tian, et al. Multi-scale enhancement fusion for underwater sea cucumber images based on human visual system modelling[J]. Computers and Electronics in Agriculture, 175.
[3] Foresti G L , Gentili S . A Vision Based System for Object Detection in Underwater Images[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2000, 14(2):167-188.
[4] Walther D,Edgington D R,Koch C. Detection and tracking of objects in underwater video[C]. CVPR2004,2004,1:I-544-I-549Vol.1.
[5] Viola P, Jones M. Rapid Object Detection Using a Boosted Cascade of SimpleFeatures [C]. IEEE Conference on Computer Vision and Pattern Recognitino, 2001,1:511.
[6] Viola P, Jones M J. Robust Real-Time Face Detection [J]. IEEEInternational Conference on Computer Vision 2004:747-747.
[7] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.IEEE Computer Society, 2005:886-893.
[8] Felzenszwalb P, Mcallester D, Ramanan D. A Discriminatively Trained, Multiscale,Deformable Part Model [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2008, 8: 1-8.
[9] Felzenszwalb P F , Girshick R B , Mcallester D A . Cascade object detection with deformable part models[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.
[10] Felzenszwalb P F , Girshick R B , Mcallester D , et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9):1627-1645.
[11] Girshick R B, Felzenszwalb P F, Mcallester D. Object Detection withGrammar Models[J]. Nips, 2010, 33:442-450.
[12] Girshick R B. From Rigid Templates to Grammars: Object Detection with StructuredModels [J].A Dissertation Submitted to the Faculty of the Division of the PhysicalSciences, 2012.
[13] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. computer vision and pattern recognition, 2014: 580-587.
[14] Girshick R . Fast R-CNN[J]. Computer ence, 2015.
[15] 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 and Machine Intelligence, 2015, 39(6).
[16] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2015.
[17] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. 2016.
[18] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. 2016.
[19] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018.
[20] Bochkovskiy A , Wang C Y , Liao H Y M . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020
[21] https://github.com/poxiao2/CCTDecode