Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 17(1): 845–861
DOI: 10.3934/mbe.2020045
Received: 30 April 2019
Accepted: 22 October 2019
Published: 05 November 2019

Research article

Convolutional neural network based obstacle detection for unmanned surface vehicle

Liyong Ma*, Wei Xie*and Haibin Huang

School of Information Science and Engineering, Harbin Institute of Technology, Weihai 264209 , China

* **Correspondence:** Email: maly@hitwh.edu.cn, xw1248@163.com; Tel: +866312679199.

Abstract: Unmanned surface vehicles (USV) is the development trend of future ships, and it will be widely used in various kinds of marine tasks. Obstacle avoidance is one key technology for autonomous navigation of USV. Convolutional neural network based obstacle classification and detection method is applied to USV visual images in environment sensing task. To solve the problem of low detection and classification accuracy of obstacles in the visual inspection of USV, a bidirectional feature pyramid networks is proposed combining hybrid network architecture of ResNet and improved DenseNet. The proposed method can further enhance the detection and classification some types of obstacles by using the underlying multi-layer detail features and high-level strong semantic features in the network architecture. The detection and classification performance of the proposed method is evaluated on a self built dataset. Ablation experiments and performance tests on open datasets are also employed. The experimental results show that the proposed algorithm has best performance for obstacles detection, and it is more suitable for autonomous navigation of USV.

Keywords: unmanned surface vehicle; convolutional neural network; obstacle detection; deep learning; feature pyramid network

1. Introduction

In recent years, with the development of computer hardware and information technology, unmanned surface vehicles (USV) has made rapid progress and development, and it is the development trend of future ships. For example, USV can replace persons to perform a task that is dangerous or requiring long-term attendance, thereby reducing casualties and costs. In the future, USV will be widely used in marine tasks, such as environmental monitoring, search and rescue, hydrological mapping, maritime supervision and so on [1, 2].

As a highly intelligent system, USV must have a stable and reliable autonomous navigation system

to finish tasks. Obviously, only with the ability of environment perception, USV can deal with the complex environment. The environmental perception of USV is to obtain environmental information through multi-sensor fusion. The USV uses the highly effective computer vision method to simulate the human vision intelligent behaviour, extract the information from the complex environment to carry on the analysis and the processing, and enhances the environment perception ability. Obstacle avoidance is one of the key technologies for USV. The research on the recognition and detection of obstacle is mainly based on radar, infrared and visual images [3–5]. Compared with radar and infrared method, visible light vision system can obtain obstacle target information, such as texture, shape and so on, so as to improve the obstacle detection ability greatly. This paper studies the obstacle detection and classification method based on visible light for USV.

The early existing methods of detecting marine targets are mostly based on simple image processing and traditional machine learning [3–6]. These methods mainly use image pre-processing to extract edge, shape, texture and other features for target detection, or select target sample on a given image, then extract feature on the sample for training, and finally target detection by classifier. The digital image processing method is simple in operation, but limited in function, easy to be interfered by external factors, it is more suitable for the target detection of simple scene, and the detection accuracy is low in the complex environment. The traditional target detection method based on machine learning needs prior knowledge, manual extraction of features and other processes, its procedures are cumbersome, real-time and not very suitable for complex and multi-target detection tasks. Background subtraction methods were evaluated for object detection in a maritime environment is discussed in [7]. An agglomerative clustering of temporally stable features is applied for object detection in highly dynamic maritime environment in [8]. Adaptive hysteresis threshold method was applied to saliency map for boat detection in [9]. A graphical model is developed for segmentation obstacle in USV [10]. A stereo model instead of single view model is proposed for obstacles extraction in [11]. A more comprehensive review of vision-based maritime object detection and tracking can be found in reference [12].

In recent years, the target detection method based on convolutional neural network (CNN) has emerged as a cutting-edge technology [13–17]. This kind of target detection algorithm based on deep learning has strong intelligence ability and high detection efficiency [18]. For example, dual path network (DPN) is developed to provide better performance for object detection in [19], a modified VGG16 network architecture is proposed for visual object detection of marine surface objects in [20]. CNN was used for surface vehicle detection and tracking in [21], in which Faster R-CNN and YOLO method were employed. Semantic segmentation networks including SegNet, ENet and ESPNet were evaluated for maritime survellance in [22]. An improved Faster R-CNN method is developed for maritime target detection, in which Resnet is used to extract feature and bacth normalization layer is employed to optimize for Faster R-CNN [23]. An object detection method is proposed with fusing region based recognition and regression based location is reported in [24]. Another maritime target detection method based on hierarchical and multi-scale convolutional neural network is proposed in [25]. It used multi-scale strategy to expand region proposal to multi convolutional layers of ResNet. It also extracted target on the fourth layer instead of the last convolutional layer of R-CNN, and added deconvolution operation with bilinear interpolation for small target perception.

Recently, a USV is developed by Harbin Institute of Technology and visual camera is installed on USV for automatic obstacle avoidance [26]. Due to the obstacles are small in image and blurred in

contour, it is the key to improve the accuracy of obstacles for USV. In this paper, the obstacle detection algorithm based on convolutional neural network for USV is developed. A feature pyramid method of hybrid network combining ResNet and DenseNet is proposed, which is more efficient for the obstacle detection by using the underlying multi-detail features and high-level strong semantic features in the network architecture. Experimental results show that this method has the highest detection and classification performance than other CNN based methods and is more suitable for USV.

2. Materials and methods

2.1. Convolutional neural network

As one of the most popular networks in the deep neural network, convolutional neural network is widely used in many fields, especially in the field of image classification and detection [27–30]. The traditional neural networks are all connected networks, the upper and lower neurons are fully connected. With the increase of the network level, the number of parameters expands, and the computational volume not only makes the network easy to fit, but also easily falls into the local optimum. Convolution networks usually include input layer, convolution layer, pooling layer and fully connected layer. The most important characteristics of convolution networks are weight sharing and sparse connection, which can greatly reduce the number of parameters of training networks and reduce the computational complexity.

Convolution kernel is the core of feature extraction in convolution network. The output pixels x_j^l of convolution layer are calculated as

$$x_j^l = f(u_j^l), \tag{2.1}$$

$$u_j^l = \sum_{i \in M_j} x_i^{l-1} * W_{ij}^l + b_j^l, \tag{2.2}$$

where f is the activation function, x_i^{l-1} is an pixel in the upper feature image layer, W_{ij}^l is the convolution kernel, the symbol $*$ is the convolution operation, b_j^l is the bias item, M_j is the subset feature image of upper layer, l is the layer number. This convolution process is to perform convolution operation of convolution kernel to input layer image, and then the new feature image is obtained by data conversion from activation function. And ReLU function as following is selected as the activation function

$$max(0, x) = \begin{cases} 0, & if \quad x \leq 0 \\ x, & if \quad x > 0 \end{cases} \tag{2.3}$$

Pooling layer samples each input feature map through the following formula and outputs the eigenvalue

$$x_j^l = f(u_j^l), \tag{2.4}$$

$$u_j^l = \beta_j^l down(x_j^{l-1}) + b_j^l, \tag{2.5}$$

where u_j^l is the j-th channel activation of l-th down-sample layer, it is obtained from down-sampling and the weighted calculation of the output feature map x_j^{l-1} from the previous layer, β is the weights.

The *down* symbol represents the down-sample function, it divides the input feature map into several non-overlapping image blocks with size of $n \times n$ by sliding window, and then sums the pixels within each image block to find the mean or maximum value, so that the output image is reduced to $1/n$ by two dimensions.

The convolution kernel parameters can be trained through the following optimization loss function

$$C = \frac{1}{2m} \sum\nolimits_x \|y(x) - a(x)\|^2, \tag{2.6}$$

where C is loss function, m is the sample number, x is samples, y is actual output, a is the network output. The goal of network training learning is to minimize the objective function, which can be accomplished by the reverse gradient descent method.

The most effective methods to improve convolutional neural networks are to improve the performance of models by deepening the network hierarchy. However, when the network layer is increased to a certain extent, the optimization function will fall into the local optimal, deviating from the global optimal, and the deepening of the network accelerates the disappearance of the gradient. In order to solve this problem, the deep residual network ResNet [31] and the densely connected network DenseNet [32] are proposed to mitigate the effect of gradient disappearance. With the deepening of the network, there are a lot of parameters to be trained in the network model. It is difficult to learn better effect from the sample of small dataset, however, large-scale dataset can not get more labeled samples by hand in some tasks. In this paper, transfer learning is used to improve the learning efficiency for the problem of small sample data.

Transfer learning is to apply the knowledge learned in one mode to another related domain for problem solving. The use of transfer learning in image detection means that the feature extraction part of convolutional neural network is trained in another large-scale data set, the corresponding training weight parameter is obtained, and then the training fine tuning of network model is carried out on the basis of the weight of the training in the small scale dataset. Therefore, in order to improve the learning efficiency of the deep neural network, we use the transfer learning in the application of the deep learning neural network for the requirement of large data and the time consuming of training. In this paper, ImageNet dataset is employed for transfer learning.

2.2. Faster R-CNN

Region convolutional neural network (R-CNN) uses deep learning for object recognition and detection employing region search. Fatser R-CNN uses region proposal network (RPN) instead of selective search to speed up the object recognition and detection [13]. RPN reduces the amount of suggested box calculations by sharing the convolution layer and parallel computing. At the same time, the target border is roughly corrected by the border regression in the RPN network, and then corrected again in the final border return of the network, and the two fixes make the target more accurate. The loss function of faster R-CNN is defined as

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_i^*) \tag{2.7}$$

$$L_{cls}(p_i, p_i^*) = -\log\left[p_i^* p_i + (1 - p_i^*)(1 - p_i)\right] \tag{2.8}$$

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \tag{2.9}$$

$$R(x) = \begin{cases} 0.5x^2, & if \quad |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \tag{2.10}$$

where i is the anchor index of batch, p_i is the target prediction probability of the anchor i , if anchor is positive, the true lable probability p_i^* is 1, otherwise 0. t_i is a vector which represents 4 parameters coordinate, t_i^* is the true boundary coordinate, $L_{cls}(p_i, p_i^*)$ is the classification loss, and $L_{reg}(t_i, t_i^*)$ is regression loss. N_{cls} and N_{reg} are normalized, λ is balanced weights. The 4 coordinates are as follows

$$\begin{cases} t_x = (x - x_a)/w_a, \\ t_y = (y - y_a)/h_a, \\ t_w = \log(w/w_a), \\ t_h = \log(h/h_a). \end{cases} \tag{2.11}$$

$$\begin{cases} t_x^* = (x^* - x_a)/w_a, \\ t_y^* = (y^* - y_a)/h_a, \\ t_w^* = \log(w^*/w_a), \\ t_h^* = \log(h^*/h_a), \end{cases} \tag{2.12}$$

where the centre of the boundary is (x, y) , w and h is width and height. Faster R-CNN detection is very accurate, but the detection speed is slow. In our USV application test, it can be performed about 5 frames per second, it cannot meet the real-time requirements of obstacle detection for USV.

2.3. Feature pyramid networks

Most deep learning recognition and detection algorithms use the top-level feature map for prediction, but the actual bottom level feature map contains more detailed information and more precise target location. Some other algorithms use multi-scale feature maps for prediction respectively, but mainly use high-level feature map information, such as single shot multi-box detector algorithm [33]. And network feature of feature pyramid networks (FPN) has the advantages of independent forecasts on the feature map, it is due to the fact that different depth corresponds to the different feature information [34].

The underlying high-resolution feature figure contains more details while the high-level low-resolution feature figure contains more semantic information. By fusing feature information of different layers, the efficiency of small target detection and recognition can be improved effectively. It can be seen from the structure that the main network forms different scale and different layers of pyramid feature map when propagating forward [34]. FPN propagates the multi-scale feature map from the side to the back, and the feature map of each layer fuses with the lower layer by upper sampling, and then the fused feature map of each layer is predicted separately. In the forward propagation path of the main network, the scale of feature map decreases gradually, but the semantic feature increases gradually. The reverse top-down path of FPN enhances the underlying semantic features by fused with the underlying feature map through horizontal connection, and at the same time makes better use of the underlying multi-details information. So CNN combined with FPN can further improve the detection accuracy of small targets by using multi-scale feature information. And in this paper, FPN is employed for obstacle detection and recogniton for USV.

2.4. Proposed CNN method for obstacle recogniton of USV

2.4.1. Improved dense block

The gradient descent technique is employed in CNN. When input information and gradient information are transferred layer by layer, the problem of gradient vanishment becomes more and more serious as the number of layers increases. This will lead to the training failure of deep neural network. The most effective way to avoid the disappearance of gradients is to establish direct connections between layers that are not adjacent to each other. DenseNet adopts this approach and achieves great success. In DenseNet, the input for each layer comes from the output of all the preceding layers. Due to each layer is connected to input and loss, it can make gradient vanishment weaken. DenseNet uses dense block to make the transmission of features and gradients more efficient, this network architecture transmit and use features more effectively.

In DenseNet, the output of the l-th layer is

$$y_l = F_l([x_0, x_1, ..., x_{l-1}], W_l), \tag{2.13}$$

where y_l is the output of the l-th layer, F_l is a non-linear transformation function, x_i is the input of the l-th layer ($i = 0, 1, ..., l-1.$), $[x_0, x_1, ..., x_{l-1}]$ refers to the concatenation of the features produced in layer $0, 1, ..., l-1$, W_l is the parameters of F_l in the l-th layer.

In DenseNet, features of the previous layers are concatenated with the same weight, but not all these previous features are useful. An improved architecture is proposed by adding the trainable weight parameters to each skip connection [27, 28]. This is shown in Figure 1.

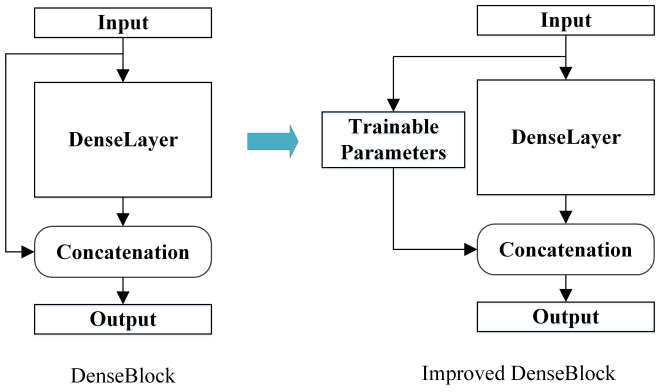


Figure 1. Improved dense block with trainable parameter.

The output of the l-th layer in this improved architecture is modified as

$$y_l = F_l([x_0 k_{l,0}, x_1 k_{l,1}, ..., x_{l-1} k_{l,l-1}], W_l), \tag{2.14}$$

where $k_{l,0}, k_{l,1}, ..., k_{l,l-1}$ is the parameters which determinate weights of $[x_0, x_1, ..., x_{l-1}]$ when they concatenate to the $l-$th layer. The improved dense block is efficient for image classification, and it is employed in this paper for obstacle detection of USV.

2.4.2. Proposed network architecture

By analyzing the algorithm structure combining Faster R-CNN and FPN network, it can be seen that the network uses the fusion of the feature map of each layer and that of the previous features layer

to enhance the predicted feature information. To a certain extent improve the accuracy of the small target detection and recognition, but it still has improvement space for network structure design. At present, the improvement of deep learning target detection algorithm is mainly considered from two aspects. On one hand, it can attain more detailed information by improving structure of the network. On the other hand, it can assist recogntion of small target using the peripheral object information by combining image context information and algorithm. Being combined with the context information, just like human visual identification, when information such as appearance and color cannot be attained due to far distance, small size and fuzzy contour, it can be conjectured through surrounding large target object information. This idea is similar to the language prediction of recurrent neuron network (RNN) used to establish data sequence and sequence data before and after have relatively stronger correlation. Bidirectional RNN using both front and back information when makes sentence prediction. The bidirectional RNN is shown in Figure 2.

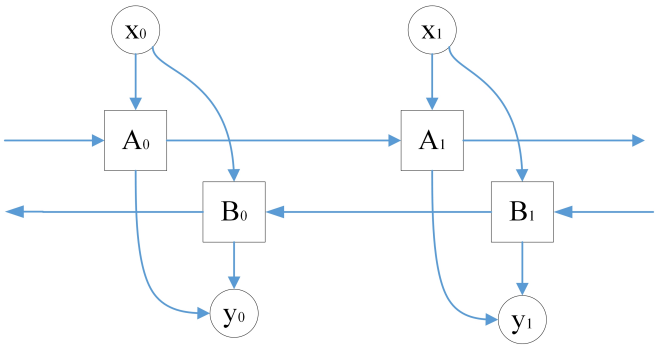


Figure 2. Bidirectional RNN.

Therefore, the idea of improvement in this paper is getting inspiration from the structure of bidirectional RNN, adding a bottom-up path to the top down path in the FPN structure, in order to improve accuracy of marine obstacle detection and recogntion for USV. The structure design is shown in Figure 3.

First, Resnet and DenseNet are used to construct a hybrid network combined with FPN, the output feature map of each layer is improved according to bidirectional architecture. The proposed bidirectional FPN architecture makes the features map of each layer not only be with the aid of the upper high semantic features, but also use the lower level detail features to assist small target classification and recognition of the current layer. Architecture in Figure 3 use the improved DenseNet described before in the hybrid network. The output feature map of each layer in the proposed bidirectional FPN is the sum of P2, P3, P4 and P5 of ResNet and DenseNet.

2.4.3. Network parameters

After reducing and enlarging, the shortest side of the input image is not less than 600 pixels and the largest edge is not more than 1024 pixels, because size setting of image clipping does affect the result of detection. Parameter limitation of 600 and 1024 has a relatively good effect. In RPN network, non-maximum suppression (NMS) is used for border selection. The judgment threshold of RPN target positive sample is set to 0.7 and negative sample is 0.3. In RPN, the proportion of border of foreground target is set to 0.5. That is, the ratio of positive and negative samples of anchor is

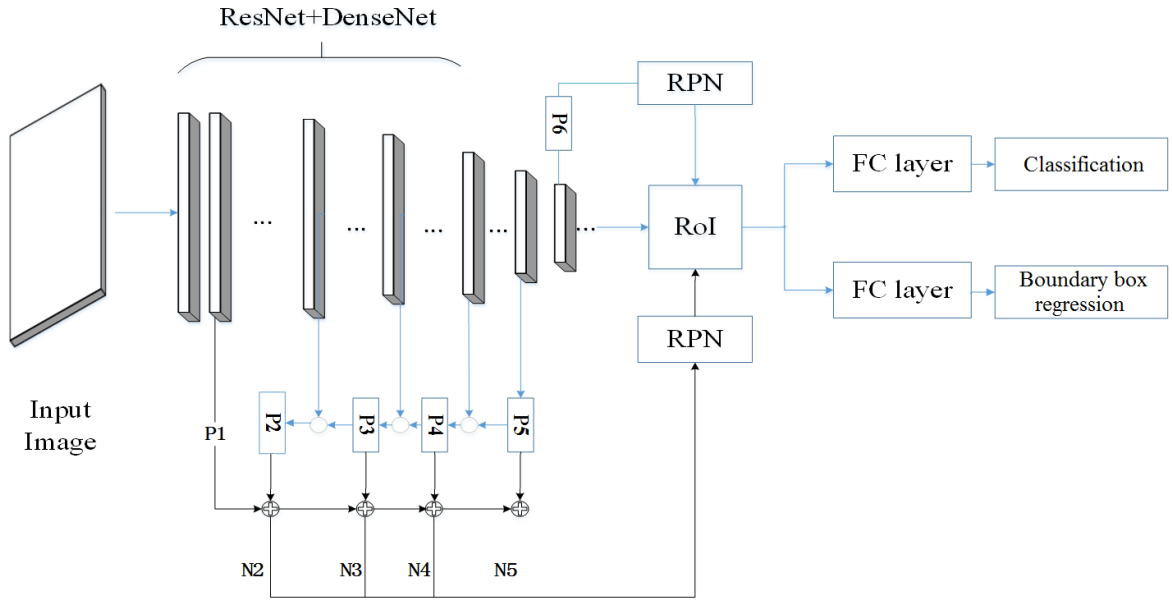


Figure 3. Proposed RFPN with hybrid network.

maintained at 1:1 after selection. There are 256 anchors that are selected here, with 128 positive samples and 128 negative samples.

The training process uses subsection training method. Namely in every 10,000 times iteration training, every 1,000 times save weight value once time. The maximum primary weight of mAP (mean average precision) in the test results is found as the initial weight of the new training and repeat the replacement of the optimal initial weight. Such segmented training can prevent network over-fitting, speed up network convergence, improve the network performance of training quickly. Replace the weight several times until the detection accuracy is no longer improved. Then the final optimal training weight can be obtained.

3. Results and discussion

3.1. Dataset and methods

To estimate the proposed recognition method, some other methods are employed for test. Standard Faster R-CNN [13] with VGG16 as backbone and Mask R-CNN [15], using ResNet101 as backbone with FPN are employed to compare with the method proposed in this paper. Three state-of-the-art object recognition methods reported in references are selected for comparison, they are improved Faster R-CNN [23], fusion based method which fuses region based recognition and regression based location [24], and CNN method based on multi-scale [25]. So five methods are used to compare with our proposed method, they are Faster R-CNN, Mask R-CNN, improved Faster R-CNN, fusion, and multi-scale method.

We collect 2,800 images using our own USV. As our images are collected by USV on a lake, the content of the images are not rich enough. Some sample images are illustrated in Figure 4. We also collect 9,300 marine images through the network. We select 8,400 images as training set, 2,100 as

validation set and 1,600 as test set. As the purpose of our obstacle detection method is to support the autonomous navigation of USV for the obstacle avoidance, we divide the target into four categories. These four categories are aircraft, bird, ship and people. Then the prepared dataset is labeled with the border and category name of the target object in each image through the image annotation software LabelImg, and the labeling information is saved and transformed into the text format. Python program is written to read text files and convert them into XML files in PASCAL VOC format. Finally, all the labeled dataset is named USVD2018, and it is employed in our test experiments. Some sample images in USVD2018 are illustrated in Figure 5.

Figure 4. USV captured image samples.

Figure 5. Some image samples in USVD2018 dataset.

3.2. Algorithm performance evaluation

According to whether the classification results are correct, TP, TN, FP, and FN can be determined. TP means that the classification result is true and positive. Similarly, TN means true negative, and so on.

Recall, precision, accuracy, specificity and F1-score are employed as classification performance

indicators to evaluate different methods. They are defined as follows.

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3.1}$$

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{3.2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3.3}$$

$$\text{F1-score} = \frac{2TP}{2TP + FP + FN}. \tag{3.4}$$

Recall measures the proportion of actual positives that are correctly identified as such. Accuracy is defined as the proportion of all samples that have been successfully classified. Precision is the ratio of samples correctly classified as positive to all the samples which are classified. F1-score is the harmonic mean of precision and sensitivity. When the above performance index is greater, the classification performance is better.

3.3. Results and discussion

3.3.1. Experiments on USVD2018 dataset

Our proposed method and other methods are tested in our experiments using USVD2018 data set. The performance indicators of each class are listed in Tables 1– 4. These data are also plotted in Figure 6. It is clear that our proposed method obtains the best values in all the evaluation indicators of four categories. It can be seen from Tables 1 to 4 that the detection accuracy of aircraft is the highest in the 4 categories, since the large number of aircraft samples are used in data training. Correspondingly, the detection rate of other categories is relatively low. Thus, it can be seen that the number and quality of samples are the key factors in the training and learning process. It also shows that the performance of our proposed method is the best, even if the number of samples varies.

Table 1. Aircraft class performance comparison.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed
Recall	0.8029	0.8743	0.8257	0.8886	0.8629	0.9257
Precision	0.7337	0.7927	0.7707	0.8141	0.8436	0.8594
Accuracy	0.8931	0.9225	0.9081	0.9313	0.9350	0.9506
F1 Score	0.7667	0.8315	0.7972	0.8497	0.8531	0.8913

Table 2. Bird class performance comparison.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed
Recall	0.7629	0.8114	0.7886	0.8229	0.8514	0.8629
Precision	0.7216	0.8068	0.7340	0.8205	0.8076	0.8603
Accuracy	0.8838	0.9163	0.8913	0.9219	0.9231	0.9393
F1 Score	0.7417	0.8091	0.7603	0.8217	0.8289	0.8616

Table 3. Ship class performance comparison.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed
Recall	0.7620	0.8220	0.7840	0.8440	0.8660	0.8920
Precision	0.8355	0.8726	0.8340	0.8884	0.8819	0.9065
Accuracy	0.8788	0.9069	0.8838	0.9181	0.9219	0.9375
F1 Score	0.7970	0.8466	0.8082	0.8656	0.8739	0.8992

Table 4. People class performance comparison.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed
Recall	0.7800	0.8150	0.7575	0.8300	0.7975	0.8400
Precision	0.7980	0.8338	0.7995	0.8469	0.8351	0.8842
Accuracy	0.8956	0.9131	0.8919	0.9200	0.9100	0.9325
F1 Score	0.7889	0.8243	0.7779	0.8384	0.8159	0.8615

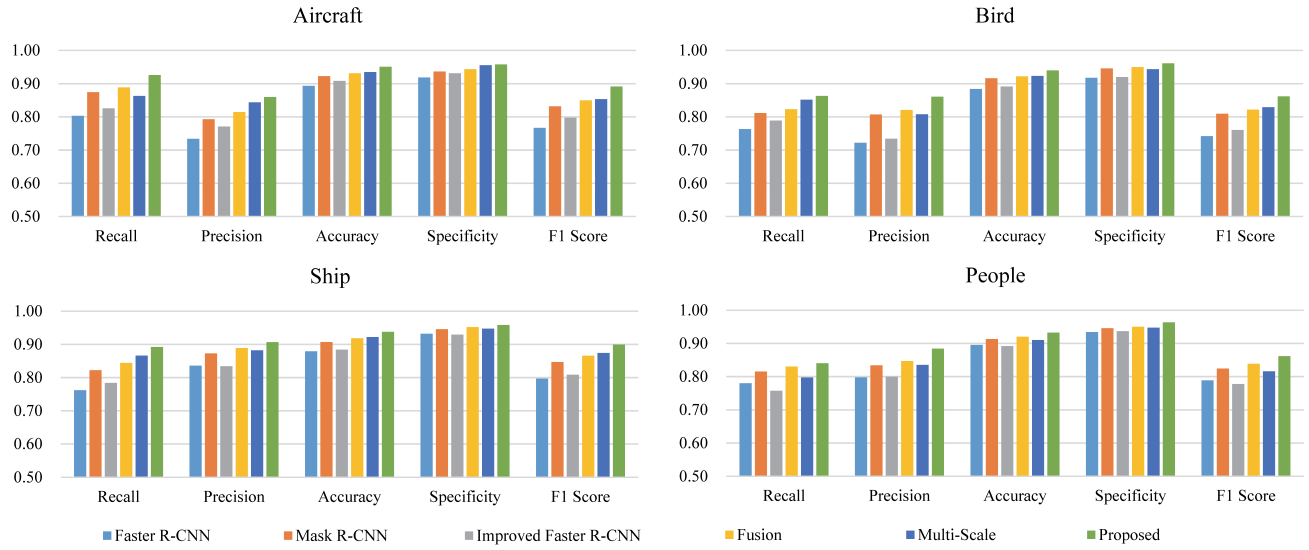


Figure 6. Performance comparison on USVD2018 dataset.

Figure 7 shows ship target detection samples. In the image on the left, there are four boats in the relatively blurry background. Other methods do not correctly detect the smallest boat, while the proposed method in this paper correctly detects all the boats. Similarly, in the image on the right, other methods miss two small boats in a clear background, as well as one not obvious hinder ship in a complex background. However, all the boats are correctly recognized by our proposed method.

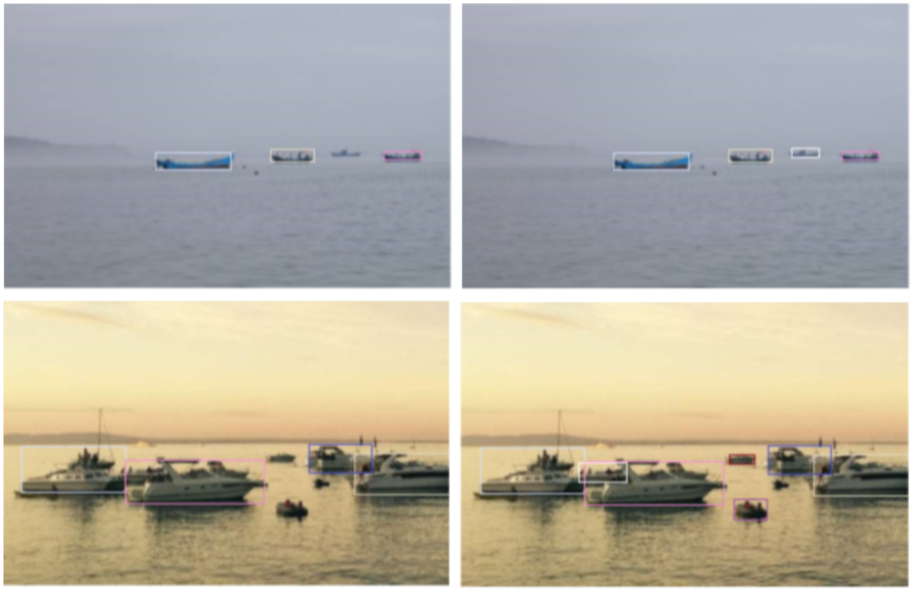


Figure 7. Ship detection result on USVD2018 dataset.

3.3.2. Ablation experiments

In order to verify the effectiveness of each part of the proposed method in this paper, ablation experiments are used in this study.

Firstly, to verify the effectiveness of the proposed biderectional RNN network architecture, experiments are performed without improved dense block. This experiment is performed with the same condition as in above section. mAP (Mean Average Precison) is used to evaluate the obstacle detection performance, and IOU = 0.5 is set as threshold. The results are listed in Table 5. It reveals that the proposed method without improved dense block still has the best performance with the highest mAP value.

Table 5. Performance comparison of the proposed method without improved dense block.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed without improved dense block
mAP	0.8636	0.8926	0.8780	0.8948	0.9040	0.9380

To verify the effectiveness of the proposed improved dense block, the improved dense block is applied to Faster R-CNN and Mask R-CNN respectively. The backbone network of these methods are DenseNet101. Whether improved dense block is used in the proposed method is also shown in Table 6. When the improved dense block is applied to these methods, all the performances is improved. It can be seen that the proposed improved dense block has the ability to improve the obstacle detection performance.

Table 6. Performance comparison of the improved dense block.

Method	Without improved dense block	With improved dense block
Faster RCNN	0.8761	0.8869
Mask RCNN	0.9002	0.9248
Proposed	0.9380	0.9463

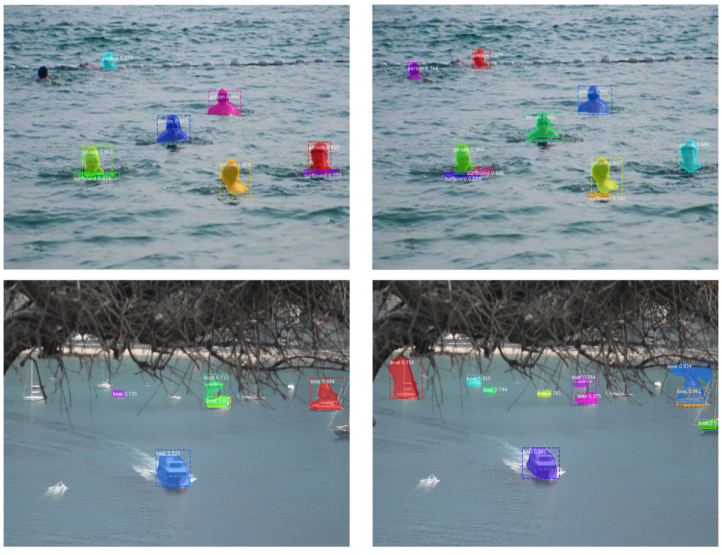
3.3.3. Experiments on COCO dataset

In order to further compare the accuracy of different methods, they are also trained and tested on a large-scale open source data set COCO. After 100 cycles and 1000 iterations per cycle, the test results on COCO data are shown in Table 7. The threshold setting of IOU = 0.5 is selected by mAP corresponding to different methods for comparison.

The experimental results show that the proposed algorithm performs better than other methods in COCO data sets with 80 categories. COCO data sets have many kinds of objects to be recognized. The average accuracy of 80 different types is calculated. The overall mAP is low when the training time is limited, but the actual detection effect of this method is very good. Some test results are shown in Figure 8. It can be seen that the detection effect of this method is better.

Table 7. Performance comparison of different methods on dataset COCO.

Method	Faster RCNN	Mask RCNN	Improved Faster RCNN	Fusion	Multi-Scale	Proposed
mAP	0.3742	0.4169	0.3804	0.4248	0.3740	0.4452



Others Proposed

Figure 8. Detection results on COCO dataset.

4. Conclusion

The obstacle detection method based on CNN for autonomous navigation of USV is discussed. A feature pyramid method of bidirectional feature pyramid networks is developed combining hybrid network architecture of ResNet and improved DenseNet. The results show that the proposed method has the highest performance for obstacle detection and is more suitable for the application of USV.

This paper mainly discusses the detection of water surface obstacles for SUV applications. The number of categories is very limited. The main problem of this method is that it can not detect untrained categories. Therefore, obstacle classification of training is very important for this method. When more classification types are trained, more types of obstacle can be detected and recognized. And more samples are required for training. In the future, we will collect more samples and carry out more detection research.

Acknowledgments

First of all, we would like to thank the reviewers for their valuable comments on this paper, which is very helpful to improve this paper. Secondly, this work was supported by Shandong Provincial Natural Science Foundation of China (ZR2018MF026), Shandong Province Key R&D Program (2019GGX101054, 2019GSF111062, 2018GGX101034), University Co-construction Project at Weihai (ITDAZMZ001708), and the Discipline Construction Foundation in Harbin Institute of Technology, Weihai (WH20160103).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. M. Schiaretti, L. Chen and R. Negenborn, *Survey on autonomous surface vessels: Part I-A new detailed definition of autonomy levels*, International Conference on Computational Logistics, 2017, 219–233. Available from: https://link_springer.xilesou.top/chapter/10.1007/978-3-319-68496-3_15.
2. D. NaÄŚ, N. MiÅąkoviÄĞ and F. MandiÄĞ, Navigation, guidance and control of an overactuated marine surface vehicle, *Annu. Rev. Control*, **40** (2015), 172–181.
3. M. Schuster, M. Blaich and J. Reuter, Collision avoidance for vessels using a low-cost radar sensor *IFAC Proc. Vol.*, **2014** (2014), 9673–9678.
4. S. Kim and J. Lee, Small infrared target detection by region-adaptive clutter rejection for sea-based infrared search and track, *Sensors*, **14** (2014), 13210–13242.
5. H. Wang, X. Mou, W. Mou, et al., *Vision based long range object detection and tracking for unmanned surface vehicle*, 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2015, 101–105. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/7274604/.

6. Y. Liu, L. Ma, W. Xie, et al., Parallel GPU computation model for block matching of speckle tracing, *J. Nonlinear Convex Anal.*, **20** (2019), 827–833.
7. D. Prasad, C. Prasath, D. Rajan, et al., Object detection in a maritime environment: Performance evaluation of background subtraction methods, *IEEE Trans. Intell. Transp. Syst.*, **20** (2019), 1787–1802.
8. C. Osborne, T. Cane, T. Nawaz, et al., *Temporally stable feature clusters for maritime object tracking in visible and thermal imagery*, 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2015, 1–6. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/7301769.
9. T. Cane and J. Ferryman, *Saliency-based detection for maritime object tracking*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, 18–25. Available from: https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w20/html/Cane_Saliency-Based_Detection_for_CVPR_2016_paper.html.
10. M. Kristan, V. Kenk, S. KovaÄŊiÄŊ, et al., Fast image-based obstacle detection from unmanned surface vehicles, *IEEE Trans. Cybern.*, **46** (2016), 641–654.
11. B. Bovcon and M. Kristan, *Obstacle detection for USVs by joint stereo-view semantic segmentation*, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, 5807-5812. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8594238.
12. D. K. Prasad, D. Rajan, L. Rachmawati, et al., Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey, *IEEE Trans. Intell. Transp. Syst.*, **18** (2017), 1993–2016.
13. S. Ren, K. He, R. Girshick, et al., *Faster R-CNN: Towards real-time object detection with region proposal networks*, Advances in neural information processing systems, 2017, 1137–1149. Available from: http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.
14. J. Redmon and F. Ali, *YOLO9000: Better, faster, stronger*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 6517–6525. Available from: http://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_ 2017_paper.html.
15. K. He, G. Gkioxari, P. Dollar, et al., *Mask R-CNN*, The IEEE International Conference on Computer Vision (ICCV), 2017, 2961–2969. Available from: http://openaccess.thecvf.com/content_iccv _2017/html/He_Mask_R-CNN_ICCV_2017_paper.html.
16. S. Pang, J. Coz, Z. Yu, et al., Deep learning to frame objects for visual target tracking, *Eng. Appl. Artif. Intell.*, **65** (2017), 406–420.
17. Y. Long, Y. Gong, Z. Xiao, et al., Accurate object localization in remote sensing images based on convolutional neural networks, *IEEE Trans. Geosci. Remote Sens.*, **55** (2017), 2486–2498.
18. Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, **512** (2015), 336–444.

19. Y. Chen, J. Li and H. Xiao, et al, *Dual path network*, Advanced in Neural Information Processing Systems, 2017, 4468–4476. Available from: http://papers.nips.cc/paper/7033-dual-path-networks.
20. A. Kumar and E. Sherly, *A convolutional neural network for visual object recognition in marine sector*, 2017 2nd International Conference for Convergence in Technology (I2CT), 2017, 304–307. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8226141.
21. J. Yang, Y. Li, Q. Zhang, et al., *Surface vehicle detection and tracking with deep learning and appearance feature*, 2019 5th International Conference on Control, Automation and Robotics (ICCAR), 2019, 276–280. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8813345.
22. T. Cane and J. Ferryman, *Evaluating deep semantic segmentation networks for object detection in maritime surveillance*, 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, 1–6. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8639077.
23. H. Fu, Y. Li, Y. Wang, et al., *Maritime target detection method based on deep learning*, 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, 878–883. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8484727.
24. L. Qu, S. Wang, N. Yang, et al., *Improving object detection accuracy with region and regression based deep CNNs*, 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2017, 318–323. Available from: https://ieeexplore_ieee.xilesou.top/abstract/document/8304297.
25. W. Chen, J. Li, J. Xing, et al., *A maritime targets detection method based on hierarchical and multi-scale deep convolutional neural network*, Tenth International Conference on Digital Image Processing (ICDIP 2018). International Society for Optics and Photonics, 2018, 1080616. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10806/1080616/A-maritime-targets-detection-method-based-on-hierarchical-and-multi/10.1117/12.2503030.short?SSO=1.
26. S. Jia, L. Ma and S. Zhang, *Big data prototype practice for unmanned surface vehicle*, ICCIP '18 Proceedings of the 4th International Conference on Communication and Information Processing, 2018, 43–47. Available from: https://dl_acm.xilesou.top/citation.cfm?id=3290466.
27. L. Y. Ma, C. K. Ma, Y. J. Liu, et al., Thyroid diagnosis from SPECT images using convolutional neural network with optimization, *Comput. Intell. Neurosci.*, **2019** (2019), 6212759.
28. L. Y. Ma, W. Xie and Y. Zhang, Blister defect detection based on convolutional neural network for polymer lithium-ion battery, *Appl. Sci.*, **9** (2019), 1085.
29. S. Pouyanfar, S. Sadiq, Y. Yan, et al., A survey on deep learning: Algorithms, techniques, and applications, *ACM Comput. Surv.*, **51** (2019), 92.
30. W. Rawat and Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, *Neural Comput.*, **29** (2017), 2352–2449.
31. K. He, X. Zhang, S. Ren, et al., *Deep residual learning for image recognition*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770–778. Available from: http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR _2016_paper.html.

32. G. Huang, Z. Liu, L. Van Der Maater, et al., *Densely connected convolutional networks*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 4700–4708. Available from: http://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR _2017_paper.html.

33. W. Liu, D. Auguelov, D. Erhan, et al., *SSD: Single shot multibox detector*, European Conference on Computer Vision, 2016, 21–37. Available from: https://link_springer.xilesou.top/chapter/10.1007/978-3-319-46448-0_2.

34. T-Y. Lin, P. Dollar, R. Girshick, et al, *Feature pyramid network for object detection*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 2117–2125. Available from: http://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR _2017_paper.html.

AIMS Press

© 2020 the authors, licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)