Fine-grained Image Classification of Red Tide Algae Based on Feature Pyramid Networks and Computer Aided Technique

Chunfeng Guo¹*, Hairong Cui², Kun Yu¹

¹Shandong Foreign Trade Vocational College, Qingdao, Shandong, China, 266000
²Qingdao University, Qingdao, Shandong, China, 266000

*E-mail: qdgcf@163.com

Abstract. Based on the demand of fine-grained image classification and recognition technology of red tide algae, the Faster R-CNN algorithm in the deep learning algorithm is studied. At the same time, according to the characteristics of subtle inter-class difference and intra-class difference between red tide algae, FPN algorithm was studied in the detection algorithm by computer aided technology. On this basis, TensorFlow is used to carry out experimental simulation, and the results of classification and recognition of fine-grained image of tidal algae based on different scale feature information are compared. The experimental results show that the Faster R-CNN algorithm performs generally in the classification and recognition of fine-grained image of tidal algae, but the detection results are significantly improved by combining the FPN algorithm.

Keywords: Fine-grained Classification, Red Tide Algae, Feature Pyramid Networks, Computer Aided Technique

1. Introduction
Algae are vitally important for marine and fresh-water ecosystems, and most species of algae are not harmful. However, red tide can occur when certain types of microscopic algae grow quickly in water, forming visible patches that may harm the health of the environment, plants, or animals. Red tide algae deplete oxygen and block sunlight that other organisms need to live on, and some even release toxins that are dangerous to animals and humans. Therefore, identifying the dominant species rapidly and effectively plays an important role in automatic monitoring of red tide\textsuperscript{1}.

Fine-grained image categorization, also known as sub-category recognition, is a very popular research topic in the fields of computer vision and pattern recognition in recent years. The purpose is to make a more refined sub-classification of images. Belonging to the same basic category,
fine-grained image classification is more difficult because the sub-class differences between sub-categories and larger intra-class differences are more complicated, compared to ordinary image classification tasks.

The R-CNN framework has developed for several generations, from R-CNN to Fast R-CNN [2] to Faster R-CNN [3], and the speed of image detection has been improved under the condition of ensuring the accuracy. From Faster R-CNN to Mask R-CNN [4] not only the object position can be marked with the object frame, but also the boundary contour of the object can be traced based on the pixel level. In fact, although the detection model Faster R-CNN has achieved unprecedented image recognition performance, it is still difficult to accurately depict objects in the physical world. Experiments show that fine recognition errors mainly occur at the "Genus" level of biological classification level [5].

Therefore, this paper will study the Faster R-CNN algorithm and its related optimization algorithm FPN (Feature Pyramid Networks) to classify the cell images of red tide algae in fine-grained, so as to achieve more accurate and practical recognition of harmful red tide algae.

2. Faster R-CNN network architecture

The Fast R-CNN algorithm is composed of two modules: RPN candidate box extraction module and Fast R-CNN target detection module. By sharing convolution features, RPN and Fast R-CNN are unified into one network, which solves the problem of consuming time in the process of extracting region proposals.

2.1. Fast R-CNN network

The key technologies of Fast R-CNN network are ROI pooling layer and the multi task loss function of output.

1. ROI pooling layer. ROI is the matrix window of interest in convolutional characteristic graph, which is represented by coordinates (r, c, h, w), with (r, c) represents the coordinates of the upper left corner of the window and (h, w) represents the width and height of the window. The ROI maximum pooling layer uses small size windows (H, W) to segment the original size of windows (h, w), obtains some sub windows with the size of h / H * w / W, and finally obtains the feature matrix. The ROI pooling layer is a special case of SPP [6], which can transform the features in the effective ROI region into the feature map with unified specifications. In this way, the size of the input image is no longer constrained by the ROI pooling layer.

2. The multi task loss function. The multi task loss function integrates the target classification loss and the target frame regression loss to improve the classification performance of the whole network. The definition of the multi task loss function is as follows:

\[ L(p, u, t^i, v) = L_{cls}(p, u) + \lambda \left[ u \geq 1 \right] L_{loc}(t^i, v) \]  

(1)

Thereinto, \( u \) represents the real category of the target corresponding to the candidate box, and \( u \geq 1 \) represents that when \( u \) is greater than or equal to 1, 1 is taken, otherwise 0 is taken. \( P \) is the probability corresponding to the output category \( u \). \( L_{cls}(p, u) \) is the classification loss, \( L_{loc}(t^i, v) \) is the coordinate loss of the candidate box, representing the difference between the class prediction coordinate value \( t^i = (t^i_x, t^i_y, t^i_w, t^i_h) \) and the real coordinate value \( v = (v_x, v_y, v_h, v_w) \).
2.2. RPN network

RPN network is a full convolution network used to extract candidate boxes. It takes an arbitrary size image as input and outputs a series of rectangular target candidate boxes, and each candidate box has a corresponding score. Basic flow of RPN network: firstly, input the characteristic map obtained by sharing the convolution layer into the convolution network, then slide the window on the characteristic map, the position of each slide window maps out a low-dimensional characteristic vector through the convolution layer, and consider k anchor boxes for the position of each slide window. Then, the low-dimensional eigenvectors are input into two parallel convolution layers: one is used to regress the position of the candidate box (save four offsets of the original position: 4 * k), and the other is used to score the foreground background of the candidate area (2 * k).

The objective function of RPN is the sum of classification and regression loss. Softmax classification is used in the classification and Smooth_L1_Loss is used in the regression. The total loss function of RPN is defined as:

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

(2)

In the formula, \( I \) refers to the \( i \)th anchor in each mini-batch, \( L_{cls} \) refers to the logarithmic loss of two categories (foreground and background), and \( p_i \) refers to the prediction probability of each anchor belonging to the target. When \( p_i^* = 1 \), it refers to the positive sample, while when \( p_i^* = 0 \), it refers to the negative sample. \( L_{reg} \) is the regression loss of the target candidate box, \( t_i \) is the coordinate parameter of the predicted target candidate box for each sample, and \( t_i^* \) is the coordinate parameter of the real candidate box corresponding to the sample.

2.3. Faster R-CNN network

The improvement of Faster R-CNN network is that RPN network is proposed to replace the original selective search algorithm to extract region proposals, and to share the characteristics of the convolution layer of the two networks. These two improvements make the quality, speed and accuracy of the candidate frame generated by Faster R-CNN better than that generated by Fast R-CNN. The specific process of Faster R-CNN:

1. The image of arbitrary size is input into Fast R-CNN network, and the feature is extracted through CNN network propagation, and the features are input into RPN network and subsequent convolution layer of Fast R-CNN network respectively.

2. After the feature map is input to RPN network, candidate boxes and their corresponding scores are obtained. NMS algorithm is used to filter scores and input the first \( N \) candidate boxes with high scores into ROI pooling layer.

3. After the feature map is input into the subsequent convolution layer of Fast R-CNN network, the high-dimensional feature map will be obtained. The high-dimensional feature map will be input into the ROI pooling layer together with the region proposals obtained through RPN network.
4. Later, as with Fast R-CNN, two outputs are obtained through the full connection layer, one is the target category obtained by using softmax multi classification for target recognition, the other is the returned bounding box obtained by adjusting the position and size of the target border.

3. FPN network

For the feature information extracted by convolution neural network, the information contained will change with the depth of convolution layer. The feature information extracted by low-level convolution layer contains the detail features of the bottom layer, and the feature information extracted by high-level convolution layer contains the semantic features of the top layer. Due to the subtle inter-class and intra-class differences in the image of red tide algae, in theory, the detection effect of multi-layer feature information is better than that of single-layer feature information. Therefore, in this paper, we will train the Faster R-CNN network model combined with FPN algorithm \[7,8\] to detect and verify the fine-grained classification of red tide algae image.

3.1. The basic idea of FPN

Use the feature information of different scales to predict the targets of different scales. The specific structure is shown in Figure 1.

![Figure 1. FPN network structure](image)

It is mainly divided into two processes: bottom-up and top-down. Bottom-up process is extracting features from common convolution network through feedforward calculation, as shown in Figure 1, assuming that the extracted features are \{C2, C3, C4, C5\}. Top-down process consists of two steps: up sampling and horizontal connection. Firstly, the features \{C2, C3, C4, C5\} extracted from the bottom-up process are sampled, so that the scale of the high-level features can meet the scale size of the horizontal connection with the low-level features, and then the obtained sampling features and the bottom-level features are added to get the features \{P2, P3, P4, P5\}.

3.2. FPN+Faster R-CNN network

Combining the RPN and FPN, we extract the anchor with fixed scale on the feature map of different scale, and get the anchor with different scale according to the different feature scale. In the Faster R-CNN network architecture with FPN structure, according to the following formula, the ROI is determined on the characteristic map of what scale mapping:
\[ k = k_0 + \log_2 \left( \frac{\sqrt{wh}}{224} \right) \]  

(3)

Where \( k \) corresponds to the \( P_k \) layer in FPN, \( w \) and \( h \) correspond to the width and height of ROI.

4. Experimental results and analysis

The detection process includes two parts: training stage and testing stage. During the training, the microscopic images of red tide algae cells contained in the three data sets of WHOI-Plankton, OUC-Phytoplankton and ZooScan were used. Then some preprocessing is done to the total data set, including geometric transformation of the image, and random addition of different noises and occlusion to the image. The final number of dataset images is 6000.

Tensorflow deep learning platform was used in the experiment, and the model of GPU was GTX1080ti. Among them, the pre training model is the model of ResNet structure training on the ImageNet data set, using the parameters of the model to initialize the parameters of the network to be trained. In the experiment, we trained the Faster R-CNN network structure combined with FPN idea, and verified the detection accuracy of the network model when extracting the feature information of different layers. The hierarchical distribution of convolution layer feature extraction is shown in Figure 2.

![Convolution layer feature extraction hierarchy](image_url)

**Figure 2.** Convolution layer feature extraction hierarchy

During the test, the same test set is used to test different models respectively, and the results are shown in Table 1. MAP represents the average accuracy when the target corresponds to the fine-grained classification and recognition.
In Table 1, each row represents the detection results of the training model after adding the FPN structure and using the characteristic map of different scales for prediction. The experimental results show that the multi-scale feature is better than the single scale feature for fine-grained classification and recognition of red tide algae image. Among which, the best prediction result can be achieved by combining the characteristic information of p2-p5 scale. Experiments 1-4 and 6 show that the more the number of combined feature scales, the higher the accuracy of prediction results, but when the combined feature information is comprehensive, adding other scale feature information will not have a great impact on the detection results of fine-grained classification. Experimental results 6-7 show that the prediction results with only low-level feature information are not as good as those with only high-level feature information, and the detection model with only low-level detail information is difficult to get more accurate prediction results.

5. Conclusion
Based on the research of Faster R-CNN and FPN algorithm, this paper trains the detection model of different scale feature map, and validates the fine-grained classification and recognition of red tide algae image. According to the comparison of the experimental results of different models, combined with the characteristics of p2-p5 layer, the detection effect of fine-grained classification of red tide algae image is the best, and its accuracy can reach 97.2%. Only using the Faster R-CNN model can detect the red tide algae image, but the model with FPN algorithm is more accurate. In the whole detection process, the target detection method based on deep learning can be applied to fine-grained classification and recognition of red tide algae image. According to the different characteristics of the data set to be detected, the corresponding improvement of the basic detection framework can make the detection results more accurate.

Acknowledgements
This research was financially supported by the Shandong Provincial Higher Education Research Project Foundation under grant number J18KA395.

References

| Index | P2 | P3 | P4 | P5 | P6 | mAP |
|-------|----|----|----|----|----|-----|
| 1     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.970 |
| 2     | ✓  | ✓  | ✓  | ✓  |     | 0.972 |
| 3     | ✓  | ✓  | ✓  |     |     | 0.964 |
| 4     | ✓  |     | ✓  |     |     | 0.951 |
| 5     |     | ✓  |     |     |     | 0.943 |
| 6     |     |     | ✓  |     |     | 0.882 |
| 7     |     |     |     | ✓  |     | 0.755 |
[1] David F. Millie. Detection of harmful algal blooms using photopigments and absorption signatures: A case study of the Florida red tide dinoflagellate, Gymnodinium breve. Limology and Oceanography, 1997, 1240-1251.

[2] Ross Girshick. Fast R-CNN [C]. IEEE International Conference on Computer Vision 2015. arXiv: 1504.08083.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [C]. Conference and Workshop on Neural Information Processing Systems, 2015.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollr, Ross Girshick. Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017. arXiv: 1703.06870.

[5] J. Krause, B. Sapp. A. Howard. H. Zhou, A. Toshev, T. Duerig, J. Philbin, L. fei-fei, The unreasonable effectiveness of noisy data for fine-grained recognition In European Conference on Computer Vision, pages 301-320,2016.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [C]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015.

[7] T. Kong, F. Sun, W. Huang, and H. Liu. Deep feature pyramid reconfiguration for object detection. In ECCV, 2018. 1,2.

[8] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling. M2det: A single-shot object detector based on multi-level feature pyramid network. AAAI, 2019. 2.