

Detection of floating objects in river based on improved HRNet

Qingtai Li

College of Electronic and Information, Southwest Minzu University, Chengdu 610041, China.

Abstract: Human activities and other factors lead to a variety of floating objects in rivers, which seriously affect the quality of human life and natural ecological environment. It has become a widely concerned problem for residents and needs to be solved urgently. Relying on manual inspection to find floating objects is extremely low efficiency, urgent need of intelligent technology to detect and warn in time. To solve the problem of river floating object detection, we proposed an improved HRNet method for river floating object detection. Trunk network HRNet was used to replace the original AlexNet trunk network of Faster-RCNN to detect river floating object, which enhanced the feature extraction ability. A backbone combining DenseNet and HRNet is proposed to optimize the ability to extract details and improve the detection accuracy.

Keywords: River floating object; Target detection; HRNet.

1. Introduction

Fresh water resource is an extremely important natural resource, which is closely related to our human life. With the rapid development of industry, environmental problems have been ignored by people, which leads to the increasingly serious problem of water pollution. As a component of fresh water resources, rivers play a decisive role in providing drinking water and maintaining the ecological environment. However, human activities and other factors lead to a variety of floating objects in rivers, which seriously affect the quality of human life and natural ecological environment. Therefore, it has become a widely concerned problem for residents and needs to be solved urgently. Relying on manual inspection to find floating objects is extremely low efficiency, urgent need of intelligent technology to detect and warn in time.

To solve this problem, this paper carried out a study on river waste detection, and its important work and research results are as follows:

The data set of river floating objects produced by us divides river floating objects into four categories: first, naturally generated garbage, such as plant branches and leaves washed into the river by external forces and collected on the water surface to form floating objects; The second is household waste, such as plastic bottles and other common household waste; Third, watercourse aquatic organisms, water eutrophication leads to the spread of aquatic organisms; Fourth, floating ice, which is common in cold areas, has little impact on the ecological environment but threatens shipping.

To solve the problem of river floating object detection, we proposed an improved HRNet^[1] method for river floating object detection. Trunk network HRNet was used to replace the existing trunk network AlexNet^[3] in Faster-RCNN^[2] to detect river floating object, which enhanced the feature extraction ability. A backbone combining DenseNet^[4] and HRNet is proposed to optimize the ability to extract details and improve the detection accuracy.

2. Research status

Generally speaking, the research on detection and recognition of floating objects on water surface is mainly

divided into two directions. One is image processing and machine learning based on traditional computer vision. This method relies on relatively little data, but has poor robustness and is easily affected by changes in external environment. The other is artificial intelligence algorithm based on deep learning. This kind of algorithm requires relatively sufficient data to train the model. It has been widely used in face recognition, vehicle detection and other fields, and can better cope with visual tasks in complex scenes. This paper also uses artificial intelligence method based on deep learning to conduct experiments.

Zuo Jianjun et al.^[5] used the segmentation method to find the area of floating objects in the image, and then used the back-propagation algorithm to conduct the feature training of floating objects in the river channel and build the classifier of floating objects to identify the area of segmentation. Zhu Dian et al.^[6] created a comprehensive system for collecting and monitoring river floats. Zhu He et al.^[7] combined gray threshold segmentation and river channel contour recognition to achieve accurate extraction of river channel area. The use of channel contour for channel segmentation can reduce the influence of garbage on the detection results, but the influence of floating objects on water bodies with similar forms cannot be ruled out. Li Ning et al.^[8] took plastic bags and plastic bottles as data sets, trained AlexNet network structure, adjusted the network by using gradient descent method, and preprocessed the recognition images by integrating light correction method. Finally, the recognition accuracy of floating objects on water surface was 15% higher than that of the method based on HOG features. Li Changlong^[9] took the images connected to a camera as material, conducted network training based on R-CNN network, adopted histogram equalization and logarithmic transformation to preprocess the images, and proposed an intelligent recognition process, which reached 96% accuracy in its test set. Deng Lei et al.^[10] provided visual guarantee for the garbage salvage process by using the slic method for image preprocessing, then extracting edge, texture and gray features, and using BP neural network for training, which achieved good recognition effect in practice. Lei Liyi et al.^[11] proposed a small data set about floating objects on the water surface based on the self-produced garbage photos on the lake surface of Guangxi

University, and analyzed the performance of several target detection models on the data set, Faster RCNN, R-FCN and SSD. The SSD target detection model has higher accuracy, Faster RCNN model can give more detailed prediction, and the model with rich structural features and considerable depth features has better performance for difficult targets. Zhang et al. [12] divided the model into two modules based on Faster RCNN, which fused lower-layer and higher-layer features respectively. At 13FPS, MAP reached 83.7%.

3. Method theory

3.1. Network model

We adopted the classic Faster-RCNN as our baseline, replaced the original AlexNet with HRNet, a backbone network with stronger feature extraction ability, and then combined DenseNet with HRNet. We enhance the High-resolution representation (HRNet) convolution by aggregating the (upsampled) representation from all parallel, resulting in a stronger representation. We build multilevel representation from high resolution and apply it to Faster RCNN. Faster RCNN can be divided into four main contents:

1). Conv layers. Faster RCNN first extracts feature maps of images using a set of common conv + relu + pooling layers. The feature maps are shared for the subsequent RPN layer and the full connection layer.

2). Region Proposal Networks. Classic detection methods are time-consuming to generate detection boxes. For example, OpenCV adaboost uses sliding window + image pyramid to generate detection boxes. Selective Search (SS) is selective search for R-CNN. Faster RCNN abandons traditional sliding window and SS methods and directly uses RPN to generate detection frames, which is also a great advantage of Faster RCNN and can greatly improve the generation speed of detection frames. The RPN network is used to generate region proposals. This layer judges whether anchors are positive or negative through SoftMax, and then uses bounding box regression to modify anchors to obtain accurate proposals.

3). Roi Pooling. This layer collects input feature maps and proposals, extracts proposal feature maps after synthesizing this information, and sends them to the subsequent full connection layer to determine the target category. The Roi Pooling layer is responsible for collecting the proposal, calculating the proposal feature maps, and sending them to the subsequent network

4).Classification. Classification section uses the proposal feature maps obtained, calculates the specific category of each proposal through the full connect layer and softmax, and outputs the cls_prob probability vector. bounding box regression was also used to obtain the location offset bbox_pred of each proposal, which could be used to regression more accurate target detection boxes. The model structure is shown in Figure 1.

We replace AlexNet, the original backbone network of Faster-rcnn, into HRNet with more powerful feature extraction function. HRNet[17] keeps high-resolution branches in the whole network propagation process, and gradually adds low-resolution branches at each stage to obtain more global information. HRNet connects high resolution to low resolution subnets in parallel, rather than in series as most existing solutions do. As a result, it is able to maintain high resolution, rather than restoring resolution through a process from low to high, so the predicted heat map may be more spatially accurate. Most of the existing fusion schemes

converge low - and high-level representations. Instead, HRNet enhances the high-resolution representation with low resolution representations of the same depth and similar level, and vice versa, through multiple multi-scale fusions, so that the high-resolution representation also has rich attitude estimation capabilities. As a result, our predicted heat map is likely to be more accurate, and finally, the high-resolution branch is responsible for predicting the final keypoint heatmap. The model structure is shown in the figure2:

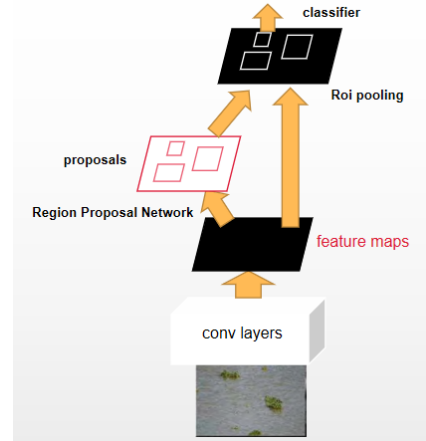


Figure 1. Structure of the Faster-RCNN network

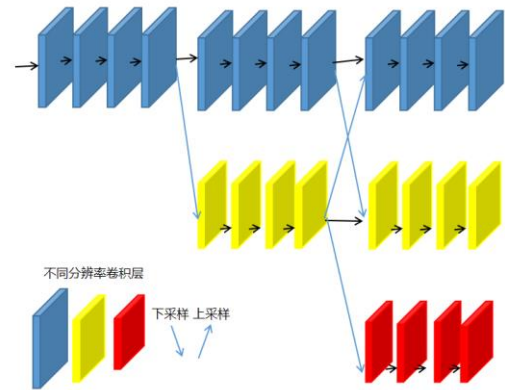


Figure 2. HRNet network structure

3.2. DenseNet

The Dense connection mechanism is added to improve the basic stage in HRNet, which not only obtains clearer feature map but also retains richer details of key points required. DenseNet directly connects all layers under the premise of ensuring maximum information transmission between the middle layer and the layer of the network. In order to ensure feedforward characteristics, each layer concatenates input from all previous layers and then passes the output feature map to all subsequent layers.

DenseBlock is the core of DenseNet, which establishes the dense connection between the previous layer and all the subsequent layers in the network, so that each layer can receive the characteristic input of all the layers before it. The Dense connection mode is shown in Figure 3.

In Figure 3, the thick line connection represents the transmission of features. The connections between the front and back layers are carried out in Concat mode, which enables the information of subtle key points in the shallow network to be retained in the deep network through the repeated reuse of features.

On the basis of HRNet high resolution, Dense connection

is introduced to improve the stage module in HRNet. This is because shallow feature maps contain rich details but lack semantic information, while deep feature maps are on the contrary [16]. Dense connection mode helps to complement the advantages between shallow and deep features. The dense connection mechanism extracts the features of each layer repeatedly and transmits the information of multiple layers together, which makes the network have richer feature graph representation.

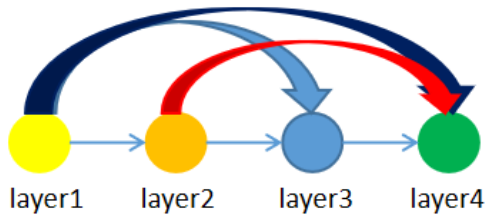


Figure 3. Schematic diagram of DenseNet connection mode

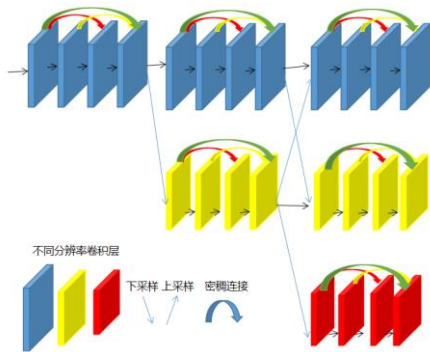


Figure 4. Dense-HRNet implementation with multiple resolutions

4. Experiment

Experimental environment and parameters

Model training and testing in this paper were carried out on MMDetection platform, which is an open source project launched by SenseTime and the Chinese University of Hong Kong for target detection tasks. MMDetection has implemented a large number of target detection algorithms based on Pytorch. The process of data set construction, model building, training strategy and so on are encapsulated into one module. Through module call, we can realize a new algorithm with a small amount of code, which greatly improves the code reuse rate. The server is installed with Ubuntu18.01 OS and three NVIDIA 2070TI graphics cards. We scaled the input image size uniformly to 256 x 192, the epoch number was 36, the batch size was 15, and the learning rate was 0.001 as our training strategy.

4.1. Data set

In recent years, with the rapid development of deep learning, the number of layers of neural network keeps increasing. Compared with traditional models, deep learning models are more and more capable of expressing, but at the same time, they are more likely to sacrifice the generalization ability of unknown data and focus on interpreting or memorizing training data, which is the so-called overfitting phenomenon. One of the effective means to make up for this shortcoming is to use high-quality and large-scale data sets for training.

The data set we produced was collected in Chengdu, Sichuan Province, and a total of 1236 images of floating

objects on the water surface were collected. lableme annotation tool was used to convert the data into VOC data set format. In order to reduce the overfitting problem caused by too small data set and the influence of various external factors, image processing operations such as cropping, horizontal flipping and adding Gaussian noise were carried out. The data set was expanded to 3525 pieces to improve the data diversity, and the annotated images were divided into training set and verification set according to 8:2.

4.2. Evaluation index

The object detection task usually contains multiple classes of object, so it is used to measure the performance of the model in all classes

mAP (mean Average Precision) Indicates the mean average precision of all aps. If there is only one detection category, the mAP value has the same meaning as the AP value. The formula for calculating mAP with n categories is as follows:

4.3. Ablation experiment

In order to verify the effectiveness of the improved algorithm, we respectively compared the algorithm accuracy of Faster-RCNN, the method of replacing AlexNet with trunk network HRNet, and the algorithm accuracy after DenseNet was introduced into HRNet. In figure 5, The experimental results are shown in the figure. According to the results, Compared with the original Faster-RCNN, the accuracy of faster-rcnn-hrnet is improved by 4.3%, because after AlexNet is replaced by HRNet, the accuracy is greatly improved, which proves the powerful feature extraction ability of HRNet. Fence-rcnn-hr-dense-Net increased by 1.36% compared with fence-rcnn-hrnet, which proves that the combination of HRNet and DenseNet enhances the feature extraction ability of details. Compared with the other three categories, the detection accuracy of aquatic organisms is lower, which may be because the color of aquatic organisms is similar to that of river channels, and it is difficult to extract features.

model	average ap				mAP
	Branches and leaves	Household garbage	Aquatic organism	floating ice	
Faster-RCNN	71.06	70.73	68.56	72.34	71.25
Faster-RCNN-HRNet	76.51	75.43	73.15	77.05	75.55
Faster-RCNN-HR-Dense-Net	77.46	76.49	74.71	78.91	76.91

Figure 5. Experimental results

5. Conclusion

In this paper, an improved HRNet algorithm based on the river floating object detection method is proposed. In the river floating object data set, the feasibility of the algorithm is verified by ablation, which can accurately identify and locate river floating objects, greatly reduce the manual workload, and is conducive to the optimization of ecological environment. This algorithm can be applied not only to the detection of floating objects in river course, but also to the detection of similar objects.

Acknowledgments

This project is supported by the graduate Innovative Research Project of Southwest University for Nationalities (Project No. YB2022400)

References

- [1] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C] // Proceedings of the IEEE / CVF conference on computer vision and pattern recognition. 2019: 5693-5703.
- [2] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in neural information processing systems, 2015, 28.
- [3] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size [J]. arXiv preprint arXiv:1602.07360, 2016.
- [4] Zhu Y, Newsam S. Densenet for dense flow[C]//2017 IEEE international conference on image processing (ICIP). IEEE, 2017: 790-794.
- [5] Zuo Jianjun, Wu Youfu. Intelligent monitoring technology of floating objects on the water surface [J]. Software Guide, 2013, 20(4): 150-152.
- [6] Zhu Dian, Liao Min, Liu Xinshun, et al. A novel floating debris collection and monitoring system in waters[J]. Modern Electronics Technology, 2019, 42(5): 27-31.
- [7] ZHU He, LI Chenming, ZHANG Lili, et al. River channel extraction combined with grayscale threshold segmentation and contour morphology recognition[J]. Journal of Electronic Measurement and Instrumentation, 2014, 28(11): 1288-1296.
- [8] LI Ning, WANG Yuxuan, XU Shoukun, et al. Identification of small sample floating objects based on AlexNet[J]. Computer Applications and Software, 2019, 36(2): 245-251.
- [9] Li Changlong. Discussion on intelligent recognition technology of floating objects on water surface based on R-CNN training method[J]. Information Technology Research and Application, 2019(Suppl 1): 18-21.
- [10] DENG Lei, YAN Lifu, ZHANG Shihan, et al. Intelligent Identification and Decision System for Water Surface Floating Objects Based on Machine Vision [J]. electronic testing, 2019(17) : 133-134.
- [11] LEI Liyi, AII Jiaoyan, PENG Jing, et al. Target detection and evaluation of floating debris based on deep learning[J]. Environment and Development, 2019, 31(6): 117-120.v
- [12] Zhang L, Zhang Y, Zhang Z, et al. Real-time water surface object detection based on improved faster R-CNN[J]. Sensors, 2019, 19(16): 3523.