Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: 10.61784/jcsee3022

METHOD FOR DETECTING WADING IN SMALL PROBABILITY EVENT SCENARIOS

GuoLin Tan[1,2*], HengHui Xiao[1], FangJiong Chen[2], YanHua Li[1], Ning Lin[1], QunLi Xiao[1]
[1]*The Central Research Institute of Guangdong Communications Services Company Limited, Guangzhou 510630, Guangdong, China.*
[2]*School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, Guangdong, China.*
Corresponding Author: Guolin Tan, Email: tglin2019@gmail.com

Abstract: In the field of computer vision, the most of applications rely on large training datasets, especially tasks such as object detection, which require a large amount of high-quality labeled datasets. However, in real-world tasks, many object detection tasks are small probability events, such as fighting, climbing over railings, wading, etc. These small probability events face difficulties in obtaining labeled datasets with high cost. In order to solve the problem of detecting people wading, this paper proposes a new low-cost and efficient method for detecting personnel wading in small probability event scenarios. This method uses water semantic segmentation and personnel object detection algorithms to segment the water area and detect people positions respectively, and then combines the water area and the target position information to accurately determine whether the person is wading. In addition to personnel wading detection, this method can also be applied to other similar scenarios, such as parks, scenic spots, factories, and other places where people's activities need to be monitored. By adjusting the algorithm and rules, this method can achieve fast and efficient recognition of human behavior in different scenarios.
Keywords: Computer vision; Small probability events; Image; Object detection; Semantic segmentation

1 INTRODUCTION

Collecting large-scale and high-quality datasets is the foundation of deep learning based detection systems. However, obtaining annotated data in many practical scenarios is very costly and difficult, especially in scenarios with imbalanced categories where positive samples are small probability events. In personnel wading detection, image semantic segmentation technology can be used to finely identify different areas in the image, such as water and non-water areas. By training neural network models, the model can learn features from different areas and automatically perform semantic segmentation and annotation in new images. This technology can help us accurately identify whether individuals have entered areas where water entry is prohibited. Unlike image semantic segmentation techniques, in personnel wading detection, object detection technology is used to recognize potential objects and their positions in surveillance images or videos, typically including identifying pedestrians, swimmers, or other objects.

This paper uses image semantic segmentation technology to segment and annotate different areas in the original collected image, identifying water and non-water areas for further detecting wading. Based on object detection technology, it identifies human targets in the original collected image and determines their locations. Finally, by combining the information of people locations and water areas, it is comprehensively judged whether personnel have entered the area where water is prohibited from entering. This method can achieve real-time monitoring and early warning of personnel wading safety, and improve the efficiency and accuracy of wading safety management.

2 RELATED WORK

Although there is not much research on personnel wading detection technology, there is a large amount of research on swimming drowning detection, mainly including the following methods, which are of great reference value.

1) Ultrasonic recognition technology[1]: Underwater ultrasonic sensors and hydraulic pressure detectors are used to detect underwater sound and pressure signals. Ultrasonic features are used to identify the swimmer's posture and hydraulic signals are used to detect the water depth. This technology can comprehensively determine whether the swimmer is drowning and issue a timely alarm.

2) Reference proposed a method for extracting human STCs feature vectors and calculating the Mahalanobis distance from the feature vectors to the standard multivariate Gaussian model[2]. It can achieve unsupervised drowning detection on high-level semantic features, solving the problem of lack of drowning videos.

3) In order to construct a self-made dataset, references carefully selected actors to simulate drowning and used drones to monitor over the swimming pool[3-5]. The captured drone images were carefully screened manually, and finally the YOLO series model was used to conduct experiments on the image dataset. Similarly, reference collected a variety of videos of babies swimming and drowning from many swimming pools and annotated them as datasets[6]. There is also research work collecting images from the Google search engine for experimentation[7]. More related work can refer to research[8-10].

3 METHODS

In view of the difficulties in obtaining training datasets, diverse water environment, and high recognition accuracy requirements in the scene of personnel wading detection, this paper proposes an efficient and low-cost method for personnel wading detection. This method splits the problem of personnel wading detection into two sub-problems, namely, training the water semantic segmentation algorithm and the personnel object detection algorithm respectively, and then combining the two algorithms to comprehensively judge whether the person is wading. The advantage of this method is that the sub problems after splitting are common non-small probability event problems. Specifically, we first introduce the semantic segmentation algorithm for water area images, which identifies water and non-water area pixels in the image and annotates them. Then, we introduce the personnel object detection algorithm, which detects the position of the human body in the image and annotates it with a rectangle. Finally, we introduce how to use water area pixels and human body position information for personnel wading detection.

3.1 Semantic Segmentation Model for Water Area Images

In order to perform water segmentation on water area images with various complex environments, we need to use a large number of water area annotation datasets to train water area image semantic segmentation models. After conducting extensive research on related work, we found that the ADE20K dataset is more suitable for our application scenario[11-12]. We selected over 2100 original images and their corresponding annotated images. Figure 1 shows a sample image in the dataset, with the original image on the left and the annotated image on the right. The black pixels represent non-water areas, and the white pixels represent water areas, which are also the positions that our method needs to identify.

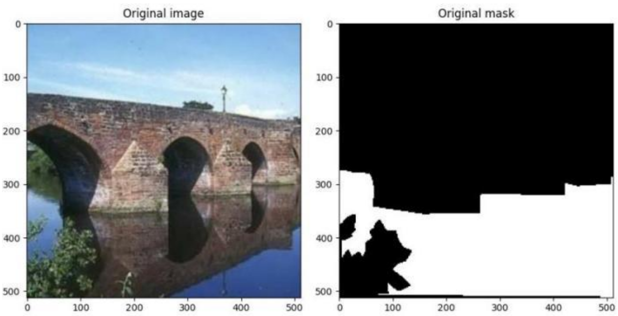

Figure 1 Sample Image from Water Semantic Segmentation Dataset

This paper adopts a neural network model that integrates multiple network architectures as a semantic segmentation model for water images, including using a ResNet pre-trained network to extract image features and using U-Net to segment water pixels in the image. The specific steps are as follows:
1) The basic model uses the ResNet50 network pre-trained on the ImageNet dataset, and then extracts the feature map learned by ResNet50 as the encoder input of the U-Net network. The encoder is responsible for gradually reducing the spatial resolution of the input image and extracting high-level features.
2) Extract the output layer of the ResNet50 network as the skip connections layer of the U-Net network, which is located between the encoder and decoder of U-Net. The skip connections layer can restore lost spatial information in the decoder.
3) Finally, a 1x1 two-dimensional convolutional layer is used to obtain a semantic segmentation result image with the same size as the original input.
By defining the above deep learning network model for semantic segmentation of water area images, using the Adam optimizer and cross entropy loss function for trainning, an average accuracy of 88% can be achieved after simple model parameter tuning. The effect of water pixel segmentation example is shown in the Figure 2. The leftmost is the original image, and the rightmost is the predicted result, which is roughly the same as the annotation result in the middle.

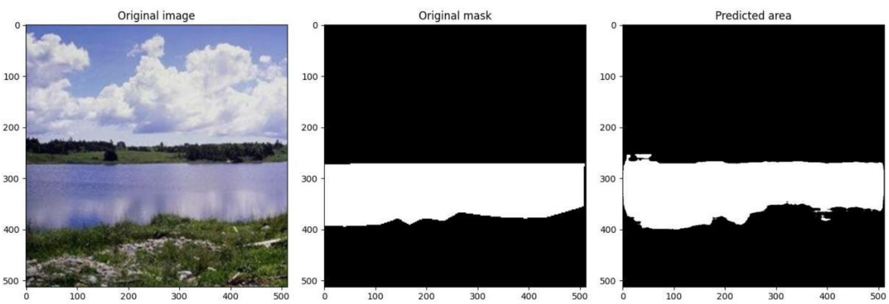

Figure 2 An Example of the Effect of the Semantic Segmentation Model of Water Images

3.2 Personnel Object Detection Model

In order to quickly and accurately detect personnel targets in various complex scene images, this paper adopts the YOLOv10 algorithm as the personnel target detection model which is trained on COCO dataset[13]. This model can recognize various scenes and environmental personnel targets in the real world, including indoor, outdoor, urban, rural, etc. Moreover, objects in the images have different scales, poses, occlusions, and background conditions, so the dataset has high diversity and complexity.

An important parameter in the model is the minimum confidence value of the detected object. Too high confidence value will lead to missed detection of personnel objects. On the contrary, too low confidence value will lead to a large number of false alarms, especially in the case of a large number of dynamic images that need to be detected in video surveillance scenarios. Therefore, it is necessary to set a reasonable minimum confidence value, which can be tested and verified based on specific application scenario images. The personnel object detection effect is shown in the following Figure 3. Although the person in the image is exposed with a head on the water surface, the model can still detect the position of the personnel object.

Figure 3 Example of the Effect of Personnel Object Detection Model

3.3 Personnel Wading Detection Algorithm

As mentioned above, we use the water semantic segmentation model and the personnel object detection model to obtain information such as the pixels of the water area and the location of the person in the image. Next, we will combine this information to quickly detect whether the person is wading, which is mainly divided into the following two steps:

1) Calculate the proportion of the bottom edge of the personnel detection box in the water area according to the pixel positions of the personnel detection box and the water area. Assuming that the four vertex positions of the personnel detection box are (x_1,y_1), (x_m,y_1), (x_1,y_n), (x_m,y_n) from left to right and from top to bottom, then the starting and endding position of the bottom edge of the personnel detection box is (x_1, y_n) and (x_m, y_n), respectively. The x represents the horizontal coordinate position, y represents the vertical coordinate position, and the pixels of the bottom edge of the personnel detection box are (x_1, y_n), (x_{1+1}, y_n), (x_{1+2}, y_n), ..., (x_m, y_n), with a total of M pixels. Assume that the set of water area pixels in the image is S. Traverse all pixels at the bottom edge of the personnel detection box and count the number I of the pixels at the bottom edge where is also in the water area pixels S. The calculation formula for I is as follows:

$$I = \sum_{i=1}^{m} (x_i, y_n) \in S \tag{1}$$

Thus, the proportion of the bottom edge of the personnel detection box in the water area is $p_1=I/M$. When p_1 is greater than a certain threshold α, it can be determined that the personnel are likely preparing or have already entered the water and then the second step below is continued. Otherwise it can be determined that the personnel have not entered the water. The specific threshold α can be set according to the application scenario and the requirements of false positive rate and false negative rate.

2) If the conditions of the first step are met, continue to calculate the proportion of water in the personnel detection box. Assume that the four vertices of the human target detection frame are (x_1, y_n), (x_{1+1}, y_n), (x_{1+2}, y_n), ..., (x_m, y_n) from left to right and from top to bottom, traverse the $M*N$ pixels in the personnel detection box in turn, and count the number of pixels W where is also in the water area S. The calculation formula is as follows:

$$W = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i, y_j) \in S \tag{2}$$

Calculate the proportion of water pixels W to the total number of pixels in the personnel detection box, $p_2=W/(M * N)$. When the proportion of water pixels p_2 in the personnel detection box is greater than a certain threshold β, it is determined that the personnel have waded into the water and corresponding alarm measures need to be taken, such as calling the administrator and using a loudspeaker to call on-site to stop them. The specific threshold β can be set according to the application scenario and the requirements of false positive rate and false negative rate.
The effect of using the method proposed in this paper for personnel wading detection is shown in the following Figure 4. We can see that it can effectively detect all 5 people crossing the river with a red box.

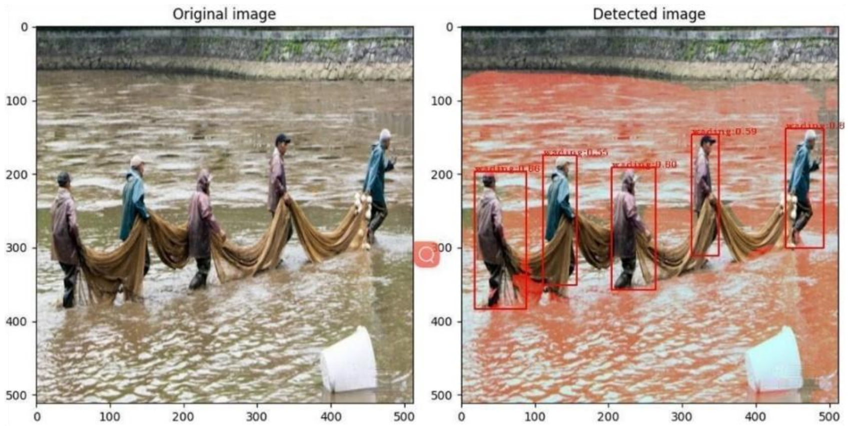

Figure 4 Example of the Effect of Personnel Wading Detection

4 CONCLUSION

The biggest advantage of this paper is that it solves the problems of high cost and difficulty in obtaining labeled datasets in traditional methods, and proposes a new method for detecting people wading in water that combines image semantic segmentation and object detection. This method splits the problem of detecting small-probability events in actual scenes into two common and easy to implement sub-problems. Therefore we can train image semantic segmentation and object detection models separately, and then combine the detection results of the two models to accurately determine whether personnel are wading in water, thereby achieving the effect of personnel wading security monitoring. This method has the characteristics of low implementation difficulty, low deployment cost, and high recognition accuracy in complex environments, and plays an important role in the field of image and video security monitoring.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The project was supported by Guangzhou Science and Technology Plan Project: High-Level Enterprise Research Institute Independently Selected Project Topic (202205110004).

REFERENCES

[1] Liu H, Frej M B H, Wen B. A novel method for recognition, localization, and alarming to prevent swimmers from drowning//2019 IEEE Cloud Summit. IEEE, 2019: 65-71.
[2] Liu T, He X, He L, et al. A video drowning detection device based on underwater computer vision. IET image processing, 2023, 17(6): 1905-1918.
[3] Yang R, Wang K, Yang L. An Improved YOLOv5 Algorithm for Drowning Detection in the Indoor Swimming Pool. Applied Sciences, 2023, 14(1): 200.
[4] Li D, Yu L, Jin W, et al. An improved detection method of human target at sea based on Yolov3//2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2021: 100-103.
[5] Handalage U, Nikapotha N, Subasinghe C, et al. Computer vision enabled drowning detection system//2021 3rd International Conference on Advancements in Computing (ICAC). IEEE, 2021: 240-245.
[6] He Q, Mei Z, Zhang H, et al. Automatic real-time detection of infant drowning using YOLOv5 and faster R-CNN models based on video surveillance. Journal of Social Computing, 2023, 4(1): 62-73.
[7] Shatnawi M, Albreiki F, Alkhoori A, et al. Deep learning and vision-based early drowning detection. Information, 2023, 14(1): 52.

[8] Alqahtani A, Alsubai S, Sha M, et al. Falling and drowning detection framework using smartphone sensors. Computational intelligence and neuroscience, 2022, 2022(1): 6468870.
[9] Kozlov E, Gibadullin R. Prerequisites for developing the computer vision system for drowning detection//E3S Web of Conferences. EDP Sciences, 2024, 474: 02031.
[10] Jalalifar S, Belford A, Erfani E, et al. Enhancing water safety: exploring recent technological approaches for drowning detection. Sensors, 2024, 24(2): 331.
[11] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 633-641.
[12] Zhou B, Zhao H, Puig X, et al. Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision, 2019, 127: 302-321.
[13] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.