Semantic Segmentation Model for Marine Pollution Detection

Wei-Bin Heng, Yi-Kai Chiu, Xiu-Zhi Chen, Yen-Lin Chen, *Senior Member*, *IEEE*

***Abstract*— Unmanned aerial vehicles (UAV) are nowadays an important piece of equipment in search and rescue missions depending to their flexibility and convenience. Computer vision algorithms such as machine or deep learning are often used in such tasks. However, such approaches rely on many aerial view data to train the machines, but most of the existing publicly available aerial datasets belong to land or traffic scenes, and a few maritime aerial datasets focus on object detection, tiny object detection, and so on. Therefore, this paper introduces a model for marine pollution detection. We use a video of a stranded cargo ship to extract 142 images as a dataset and use the PIDNet[1] as semantic segmentation model for training and validation to demonstrate the applicability to maritime oil pollution detection. Source code is released in https://github.com/HengWeiBin/Oil-Polution-Dataset-with-PIDNet.**

I. Introduction

Camera-equipped UAVs are used in a wide range of applications, mainly for aerial photography and entertainment, they are equipped with visual obstacle avoidance systems to avoid collisions. In the industrial and agricultural fields, UAVs are also used for 3D modeling[2] pesticide spraying, and shipping[3]. In addition, UAVs can also be used to assist in traffic management and are important for search and rescue (SAR) missions[4], as they can quickly reach dangerous areas without risk to humans and collect information to assist in rescue efforts to avoid risks to rescuers.

Currently, most of the above systems and applications may use data-driven methods such as machine and deep learning, and so is the case for maritime oil spill detection. However, there are few existing aerial vision datasets, for example, VisDrone[5] provides a variety of aerial camera vision scenes, including daytime and nighttime urban roads, moving crowds, and so on with object detection and tracking labels; UAVDT[6] covers a large range of urban traffic scenes, but mainly single-object tracking, multi-object tracking, and object detection; SeaDronesSee[7] provides detection datasets and scoring benchmarks for SAR missions. It is clear that the existing maritime environmental datasets are rare and usually based on satellite photography, which has a very different field of view and detection distance from our target, and the detection targets of such datasets are only valuable for large object detectio n from the sea.

In this paper, we introduces a semantic segmentation model for marine pollution detection. We use a video of a stranded cargo ship to extract 142 images as a dataset and use the PIDNet[1] as semantic segmentation model for training and validation to demonstrate the applicability to maritime oil pollution detection.

This research was financially sponsored by the National Science and Technology Council of the Republic of China under the contract No. NSTC-109-2628-E-027-004-MY3, No. NSTC-111-2622-8-027-009 and No. NSTC-111-2218-E-027-003.

II. Method

The network model we use in this research is Proportional-Integral-Derivative Network(PIDNet), they proposed a 3-branch network architecture to solve the similar overshoot problem by bridging a connection between Convolutional Neural Network(CNN) and Proportional Integral Differentiation(PID) controllers. PIDNet has three branches to resolve detail, context and boundary information, and uses boundary attention to guide the final stage of detail and context fusion. **Error! Reference source not found.** shows an overview of the basic architecture of PIDNet.

The PIDNet series achieves an optimal balance between inference speed and accuracy, with test accuracy exceeding all existing models with similar inference speed on the Cityscapes[8], CamVid[9][10], and COCO-Stuff datasets[11].

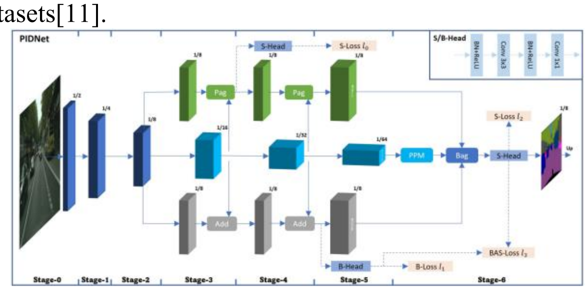

Figure 1. The architecture of PIDNet.[1]

III. Experimental Setup and Result

In this experiment, we provide a marine oil spill dataset for training. To solve the data imbalance problem, we assign corresponding weights to each class. In addition, we add two data augmentation methods to the dataset during the training. Finally, we present the training results of PIDNet and the impact of adding the data augmentation methods separately.

A. Dataset

We provide marine oil spill dataset applied in this research. We use a video of a stranded cargo ship captured by three unmanned aerial vehicles, two of which shot the front and side of the ship, and the other one started from the shore to the stranded area. The total area of the ship was 42 hectares. The UAVs flew at an altitude of 150 meters to collect the data, and we extract 142 RGB source with a resolution of 2048x1024, and label it into 6 categories: sky, ship, sea, wave, shore, and oil. Finally, we split the dataset in the ratio of 8:2, 29 images are used as the validation dataset, and the remaining images will be augmented and used as the training dataset.

B. Experiment Setup

To solve the problem of data imbalance, different weights had assigned for different classes. The classes with the less amount of ground truth in the dataset and our primary target have higher weights, otherwise will have a lower weighting. Table I shows that the class weighting in our experiments.

TABLE I. TABLE I. CLASS WEIGHTING IN OUR EXPERIMENTS

Class Weights					
Sky	***Ship***	***Ocean***	***Wave***	***Shore***	***Oil***
1.1000	1.0009	0.9000	1.5000	1.8000	1.9000

To increase the robustness of the trained model, we adopted various hue and mixing background augmentation for training.

We converts the hue value of the oil part by gamma correction so that it can be adjusted to other colors.

- $$c = \theta \cdot \left(\frac{x^{oil}}{\theta}\right)^{\gamma} \quad (1)$$

By setting γ to a random value between 0 and 2.0, we generate new input data c by using (1) and the amount of overall data can be increased up to eleven times. Fig. 2 shows that the example results of Hue Data Augmentation when γ = 0, 0.4, 0.8, 1.0, 1.4 against oil spill area.

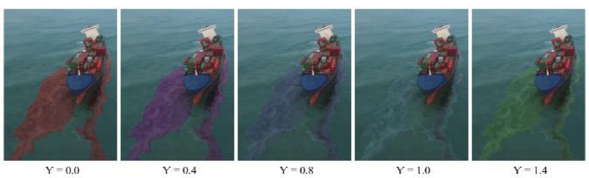

Figure 2. Gamma correction as Hue data augmentation

In MSPNet[12], the mixing background augmentation (MBA) is proposed to include foreground information obtained by background subtraction to generate more training samples. They assume road objects as foreground objects and other pixels as background to allow the detectors to learn the foreground features better. In our case, we want to better segment the area of the oil on the sea, so we generate the mask x_i^{oil} for data augmentation by using the oil pixels as foreground information only.

$$\bar{x}_i^{\alpha} = (1-\alpha)x_i + \alpha x_i^{oil} \quad (2)$$

By setting α to a random weight between 0 and 0.3, we generate new input data $\bar{x}_i^{\alpha}$ by using (2) and the amount of overall data can be increased up to three times. Fig. 3 shows the results of data augmentation using MBA when α = 0, 0.1, 0.2 against oil spill area.

Figure 3. Fig. 1. Mixing background augmentation with difference α.

C. Result

Table II shows the results of comparing different data augmentation methods by training PIDNet-S with the official ImageNet pre-trained model provided by PIDNet. Clearly, the results show that adding data augmentation can significantly improve the model, especially for the Oil and Shore categories, and the overall accuracy also increases due to the increased amount of data, and the model can be adapted to more scenarios. The visualization result of the trained model is shown as **Error! Reference source not found.**.

TABLE II. TABLE II. THE COMPARISON OF SEMANTIC SEGMENTATION F1-SCORES

DA	**F1-Scores**						
	Sky	***Ship***	***Ocean***	***Wave***	***Shore***	***Oil***	***Mean***
N/A	0.973	0.946	0.968	0.729	0.850	0.768	0.873
Hue	0.981	0.971	0.978	0.782	0.945	0.853	0.918
Hue MBA	0.982	0.971	0.979	0.793	0.956	0.862	0.923

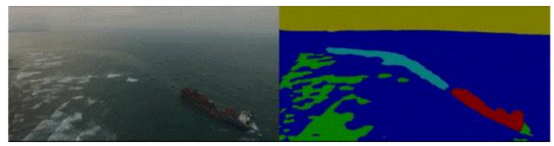

Figure 4. The marine pollution semantic segmentation result.

REFERENCES

[1] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired from pid controller," arXiv preprint arXiv:2206.02066, 2022.

[2] R. a. G. A. a. H. X. Qin, "UAV-Project- Building a reality-based 3D model," Coordinates, vol. IX, pp. 18-26, 01 2013.

[3] K. T. San, S. J. Mun, Y. H. Choe, and Y. S. Chang, "UAV Delivery Monitoring System," MATEC Web of Conferences, vol. 151, p. 04011, 2018-01-01 2018, doi: 10.1051/matecconf/201815104011.

[4] I. A. Azzollini, N. Mimmo, L. Gentilini, and L. Marconi, "Uav-based search and rescue in avalanches using arva: An extremum seeking approach," arXiv preprint arXiv:2106.14514, 2021.

[5] P. Zhu et al., "Visdrone-det2018: The vision meets drone object detection in image challenge results," in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0-0.

[6] D. Du et al., "The unmanned aerial vehicle benchmark: Object detection and tracking," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 370-386.

[7] L. A. Varga, B. Kiefer, M. Messmer, and A. Zell, "Seadronessee: A maritime benchmark for detecting humans in open water," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2260-2270.

[8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[9] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in European conference on computer vision, pp. 44-57, Springer, 2008.

[10] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," Pattern Recognition Letters, vol.30, no.2, pp. 88-97, 2009.

[11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Common objects in context," in Proc. Eur. Conf. Comput. Vis., Cham, Switzerland: Springer, 2014, pp. 740-755.

[12] C. Ping-Yang, J.-W. Hsieh, M. Gochoo, and Y.-S. Chen, "Light-weight mixed stage partial network for surveillance object detection with background data augmentation," in 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 3333-333.