Deep Learning for Arbitrary-Shaped Water Pooling Region Detection on Aerial Images

Pravina Mylvaganam
Department of Electrical and Electronic Engineering
University of Peradeniya
Peradeniya, Sri Lanka
pravina.m@eng.pdn.ac.lk

Maheshi B. Dissanayake
Department of Electrical and Electronic Engineering
University of Peradeniya
Peradeniya, Sri Lanka
maheshid@ee.pdn.ac.lk

***Abstract*— Recent rapid development in Unmanned Aerial Vehicles (UAVs) have extensively promoted several types of civilian tasks. In this paper, we propose and compare two different deep learning and convolutional neural network methods to detect and extract the region of water pooling areas, such as gutters, abandoned ponds, tires, and other water retaining areas on rooftops, using UAVs based aerial images. The performance comparison between the YOLOv4 algorithm and the Mask-RCNN algorithm was explored in the case study to identify the best deep learning method for detecting these uneven regions of water pooling. Experimental results show that the Mask-RCNN approach efficiently detects these uneven areas in an aerial image while simultaneously generating a high-quality segmentation mask for each instance. On the other hand, YOLOv4 detects the best bounding box for the area of interest. The mean average precision (mAP) scores for Mask-RCNN and YOLOv4 are 71.67% and 57.9% respectively. The Mask-RCNN system has shown promising results on test images and video clips. Such real-time detection systems would eventually help to identify mosquito breeding sites to assist the dengue eradication as well as to identify suitable water resources for daily uses, thereby facilitating a better community health system.**

Keywords— water pooling regions, Mask-RCNN, YOLOv4, aerial images, region detection, UAV

I. Introduction

In the recent past, research and development in computer vision have already impacted a wide range of applications in the real world owing to the advancements in deep learning. Deep learning implements a neural network approach with multiple layers of processing units, mainly for object detection, segmentation, and classification [1]. Also, Unnamed Aerial Vehicles (UAVs), mostly drones, coupled with image analysis have experienced a drastic development in diverse fields ranging in civilian and military applications due to their small size, fast deployment, automation capabilities, agility, and low cost during the past few years. Inspection of power lines [2], buildings [4], wildlife conservation [3], traffic and vehicles [8], and effective agriculture [5] are some examples of such applications. Moreover, in [6], a UAV cloud surveillance system is explored to eradicate the damages caused by both natural and man-made disasters. Li et al. [7] propose an unsupervised classification model for the detection of aftereffects of earthquakes, specifically earthquake-triggered roof-holes, using UAV-based aerial images. However, the inherent limitations of drones, such as their weight, power consumption, and limited battery lifetime, cannot be overlooked and should pay careful consideration when running the deep learning algorithms onboard a UAV in itself.

Furthermore, it can be seen that the development in object detection technology in many sectors, including multi-object detection, edge detection, salient object detection, face detection, scene text detection, etc has come a long way owing to the rapid development in the Deep Learning (DL) architectures. The mainstream object detection algorithm can be divided into two categories: (1) two-stage detection algorithms [9]; as the most representative one, it works in two stages, separately yet in series to generate the region proposals and to classify the features extracted from region proposals, in order to refine the location of the object. Typical examples of two-stage detectors are R-CNN (Region-Based Convolutional Neural Networks) [11], Fast R-CNN [12], Faster R-CNN [13], Mask-RCNN [14], etc.; (2) one-stage detection algorithms [9], such as YOLO (You Only Look Once) [15] and SSD [16], does not generate the region proposals separately as it does the classification and bounding box regression concurrently. The main difference between these two categories is that the two-stage detector gives better results with high localization and overall accuracy, whereas the one-stage algorithm predicts the results at high speed [17].

In order to compare the performance of the above mentioned two categories of object detection algorithms and identify the best model for region detection on UAV-based aerial image analysis, this paper evaluates the two state-of-the-art deep learning algorithms from each category; Mask-RCNN and YOLOv4 algorithm, as these two approaches have shown significant performance in their respective categories [17]. We specifically focus on detecting water pooling areas using the aerial image. In line with the practical application scenario, the aerial image dataset generated has water-retaining surfaces which do not have a fixed geometrical shape. Hence the detection algorithm should possess the ability to identify objects and areas with different arbitrary shapes. In countries where mosquito-borne diseases are present, the proposed system can be utilized to identify potential mosquito breeding sites at unreachable locations such as rooftops and gutters of high-rise buildings.

Although several datasets [19-24] are available in public domain for water detection assignments, they were designed to address different tasks, than ours. Strictly speaking, most of

978-1-6654-8786-3/22/$31.00 ©2022 IEEE

these datasets contain satellite images and these images are low in resolution while been captured from high altitudes. They are affected by several noisy artefacts such as clouds and smoke. Moreover, while deploying satellites and collecting high altitude satellite images are costly alternative, it would not suite low altitude image analysis tasks. To address these issues, we have created a dataset using locally collected high-resolution images taken from a UAV operated at low altitudes. These characteristics of the dataset bring more clarity to the task as well as novelty. Furthermore, it helps deep learning models to make more accurate and highly precise decisions regarding water pooling uneven region detection.

In addition, DL models by inheritance are data-hungry. In literature transfer learning approach is proposed to address the lack of data in task-specific applications of DL [10]. Transfer learning is an approach where knowledge is transferred from one domain to another. Hence, with this approach, a DL model pre-trained on a generalized larger dataset such as the COCO dataset or ImageNet dataset, can be adapted to perform a different yet specific task using the knowledge gained at the initial training.

In our implementation, we have utilized a transfer learning approach with the pre-trained weights of each model to address the lack of a significantly large dataset. Later we fine-tune the pre-trained weights to meet the application scenario using our custom dataset with aerial images. Finally, we compare and analyze the results obtained from both of the proposed models in terms of efficiency, effectiveness, and accuracy to understand the model most suited for the application.

The rest of the paper is organized as follows. Section II describes the methodology adopted, models, and hyperparameters we have experimented with our custom dataset in this study. In Section III, we present and discuss the results obtained in the experimental analysis. Finally, we conclude the paper by summarizing our results in Section IV.

II. Methodology

In this research, we propose a model to detect and locate water retaining areas within objects such as gutters, abandoned ponds, tires, and rooftop objects, using UAV-based aerial images with high efficiency. At the initial stage of the research, a dataset of water retaining sources, which have uneven boundaries of water, was created using images captured from a drone camera. The quality of the captured images depends on the sensor capacity of the drone camera and the Ground Sampling Distance (GSD). We have considered a GSD of 7cm or larger, which is sufficient to identify the objects clearly using the classification system. The custom built dataset contains a total of 600 images belonging to two classes; objects with water (a total of 300 images) and objects without water (total of 300 images). Next, all images in the dataset were divided into a training set and a validation set with an 80:20 ratio respectively. Then the training dataset was used to fine-tune the pre-trained weights of both Mask-RCNN and YOLOv4 algorithms for the feature extraction and classification.

A. *Mask-RCNN*

After dividing the dataset into a training set and a validation set, all aerial images were annotated using VGG Image Annotator (VIA) tool, which converts the dataset to the '.json' format to fit the proposed model. Since deep learning is data-hungry and the dataset has only 600 images, we adopted improved Mask-RCNN with transfer learning for our experimental setup. In this research, we have adopted the pre-trained model, trained initially on large COCO datasets. The model with pre-trained weights is further trained and fine-tuned using the locally collected dataset. Also, we have fine-tuned hyper-parameters of the pre-trained CNN model by manual search to achieve better results and the optimized hyper-parameters are shown in Table I.

TABLE I. The optimized hyper-parameters

Parameter	Value
Learning rate	0.001
Weight decay	0.0001
Minimum confidence of detection	0.9
Steps per epoch	10
Number of classes	2
Pool size of Mask	14
Pool size	7
Validation steps	50

Fig. 1 shows the complete system with the Mask RCNN model proposed in this study. The model was trained and fine-tuned, over 20 epochs with the softmax classifier to perform binary classification, using frame regression to get more precise information on the candidate-frame position, and eliminating the part of the region of interest (ROI) by non-maximum suppression. Once the model is trained, it is tested against new generalized as well as unseen data to check the accuracy of the proposed model in detecting water retaining areas.

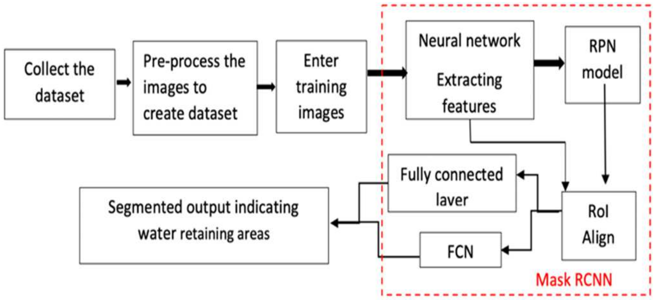


Fig. 1. Flow chart of the proposed system with Mask-RCNN architecture

B. *YOLOv4*

In this setup, the same dataset, which is used for the Mask-RCNN model, was utilized. The input images are annotated using the BBox tool to refine the coordinates of the instance of the object presented in the aerial images to make it suitable for the model training purposes. Google Collab with GPU was used for the simulations. Fig. 2 shows the outlook of the proposed YOLOv4 architecture.

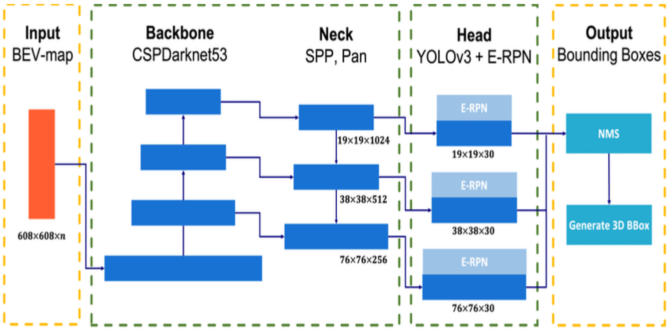


Fig. 2. Proposed system with YOLOv4 architecture

After initializing hyperparameters, learning rate, regularization, dropout as 0.0026, 0.001, 0.5 respectively, we trained the model on the locally collected image dataset using GPU. The two hyperparameters, batch normalization, and anchor model are used to predict the bounding box and the location of the object region in the given dataset. Finally, the test images and video clips captured using a drone camera were tested using the trained model to evaluate the model performance. Training loss variation through the 6000 iterations for the model presented at Fig.2 is shown in Fig. 3.

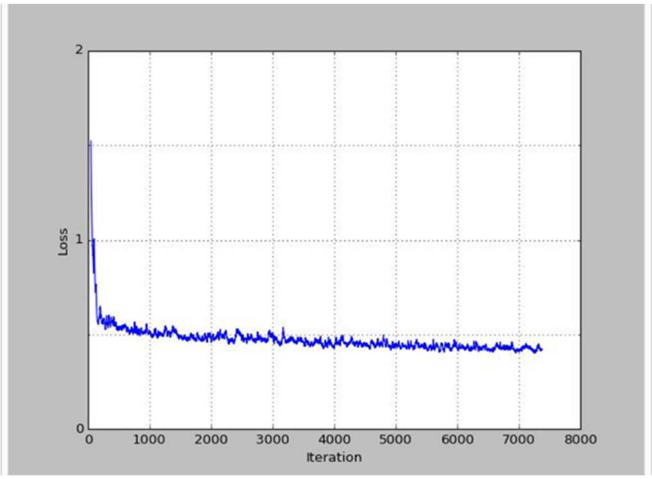


Fig. 3. Training loss graph of YOLOv4 model

III. Results and Discussion

The outputs of both models were evaluated by using mean average precision (mAP). Eq. (1) defines mAP, which is obtained by using Intersection over Union (IoU), defined as the ratio of the area of overlap between the predicted and ground truth box.

$$\text{mAP} = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \;, \qquad (1)$$

where AP_k means average precision (AP) of class k, which is given by (2), and n stands for the number of classes.

$$\text{AP} = \frac{1}{11}\sum_{r\in\{0,0.1,\ldots,1\}} p_{interp}(r)\;, \qquad (2)$$

where the interpolated precision (p_{interp}) (3) at a certain level of recall r is defined as the highest precision found for any recall level $\tilde{r} \geq r$.

$$p_{interp}(r) = max_{\tilde{r}\geq r}\, p(\tilde{r}) \qquad (3)$$

and $p(\tilde{r})$ is the measured precision at recall $\tilde{r}$. The Precision (P) and Recall (R) values alone also were used to analyse the target detection performance.

$$\text{Precision} = \text{TP}/((\text{TP+FP})), \quad \text{Recall} = \text{TP}/((\text{TP+FN})), \qquad (4)$$

where true positive (TP) is the number of correctly predicted samples. False positive (FP) is the number of samples that are incorrectly marked. False Negative (FN) illustrates the number of samples that are incorrectly marked as negative samples. The results obtained during the testing process are shown in Table II.

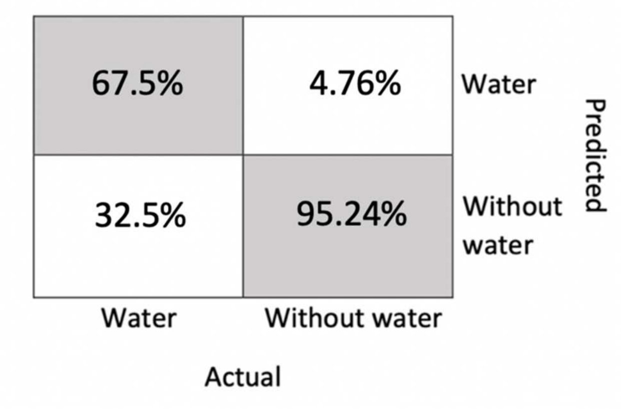
TABLE II. Segmentation performance of Mask-RCNN based system and YOLOv4 based system

Model	mAP	Precision	Recall
Mask-RCNN	71.67%	0.625	0.75
YOLOv4	57.9%	0.57	0.53

According to Table II, the test results show that the *mAP* of Mask-RCNN (71.67%) is much higher than that of YOLOv4, (57.9%). This high *mAP* performance is the major advantage of using Mask-RCNN for this application. Moreover, the boundary outline of the detected area drawn by the Mask-RCNN has high precision compared to the bounding box drawn by YOLO. The flexibility within Mask-RCNN to draw an arbitrary shape contour is an added advantage for the given task.

By contrast, YOLOv4 runs a lot faster than the Mask-RCNN at testing due to its simpler architecture with a direct approach for detecting the ROI. It predicts the output in 1 s, whereas Mask-RCNN predicts the result in 5 s. This could be due to YOLOv4 being trained to do classification and bounding box regression at the same time. Hence, in a time critical aerial image analysis environment, YOLO performs faster than Mask-RCNN.

When we used an aerial video clip for the testing, Mask-RCNN located the water resources very accurately because of its high mAP, compared to YOLOv4. Moreover, Mask-RCNN shows good precision and recall values, which are at 0.625 and 0.75 respectively, and these values are computed using the confusion matrix shown in Fig. 4. Furthermore, Fig. 5 presents the mAP graph for YOLOv4, for 6000 iterations.



		Actual	
		Water	Without water
Predicted	Water	67.5%	4.76%
	Without water	32.5%	95.24%

Fig. 4. Confusion matrix of Mask-RCNN model

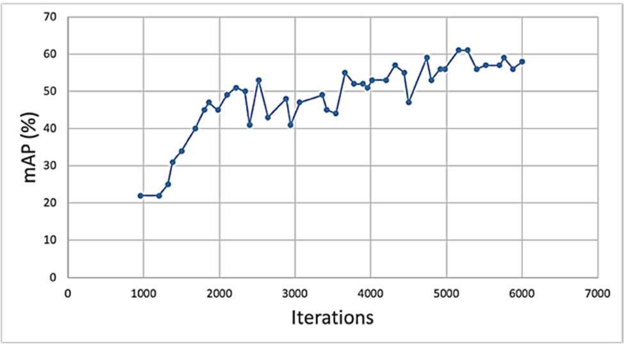
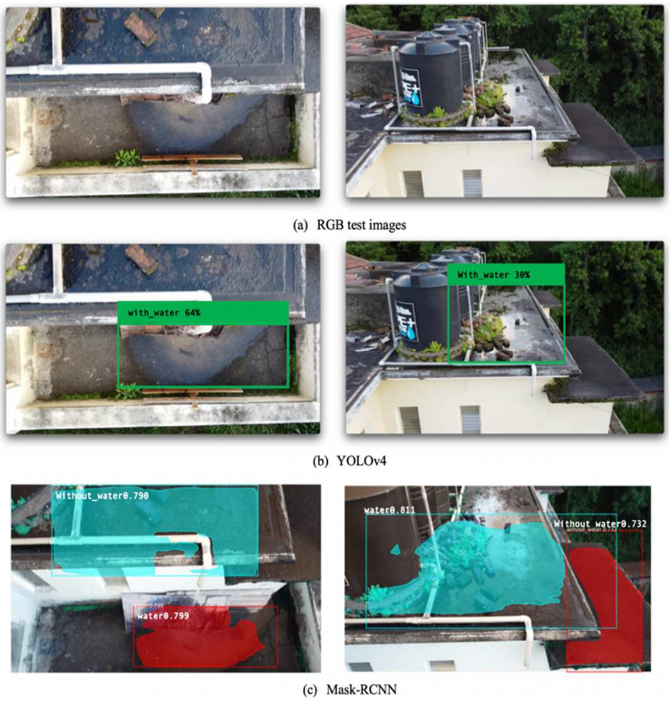


Fig. 5. mAP variation of YOLOv4 model

In addition, Fig. 6 shows samples of RGB test images and detected outputs of the YOLOv4 and Mask-RCNN models respectively.

IV. Conclusion

The main task of this study is the detection of water pooling areas using aerial images and deep learning algorithms. Such a system can be used to eradicate the ecological and hydrological-based problems, especially identifying potential mosquito breeding sites. For this purpose, we propose two deep learning algorithms to automatically detect the water retaining areas with arbitrary shapes in aerial images. We built two complete detection systems using Mask-RCNN and YOLOv4 architecture with the newly acquired dataset, which contains a total of 600 images belonging to two categories; with water, and without water. The performance evaluations show that the Mask-RCNN algorithm effectively detects water retention areas with an accuracy of 71.67% compared to the YOLOv4 algorithm which only achieves a mAP score of 57.9%. Similar superior performance in Mask-RCNN was observed in precision and recall as well. By contrast, the proposed YOLO model predicts the outputs faster than the Mask-RCNN. Time delay is the only disadvantage present in Mask-RCNN when compared to YOLOv4 for the given task.



(a) RGB test images

(b) YOLOv4

(c) Mask-RCNN

Fig. 6. Sample outputs; (a) RGB test images; (b) Predicted outputs from the trained YOLOv4 model with confidence intervals; (c) Segmentation outputs from the trained Mask-RCNN model with confidence intervals

Acknowledgment

The authors wish to acknowledge the University Research Grant, URG/2021/14/E, of University of Peradeniya for providing the support to purchase the camera drone.

References

[1] A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet classification with deep convolutional neural networks", In Advances in neural information processing systems, pp. 1097–1105, 2012

[2] C. Martinez, C. Sampedro, A. Chauhan, and P. Campoy, "Towards autonomous detection and tracking of electric towers for aerial power line inspection," in Proceedings of the 2014 International Conference on Unmanned Aircraft Systems, ICUAS 2014, pp. 284–295, May 2014.

[3] M. A. Olivares-Mendez, C. Fu, P. Ludivig et al., "Towards an autonomous vision-based unmanned aerial system against wildlife poachers," Sensors, vol. 15, no. 12, pp. 31362–31391, 2015.

[4] A. Carrio, J. Pestana, J.-L. Sanchez-Lopez et al. et al., "Ubristes: uav-based building rehabilitation with visible and thermal infrared remote sensing," in Proceedings of the Robot 2015: Second Iberian Robotics Conference, pp. 245–256, Springer International Publishing, 2016.

[5] L. Li, Y. Fan, X. Huang, and L. Tian, "Real-time uav weed scout for selective weed control by adaptive robust control and machine learning algorithm," in Proceedings of the 2016 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers, p. 1, 2016.

[6] C. Luo, J. Nightingale, E. Asemota, and C. Grecos, "A UAV-cloud system for disaster sensing applications," in Proceedings of IEEE 81st Veh. Technol. Conf. (VTC Spring), Glasgow, Scotland, pp. 1–5 May 2015,.

[7] S. Li et al., "Unsupervised detection of earthquake-triggered roof-holes from UAV images using joint color and shape features," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 9, pp. 1823–1827, Sep. 2015.

[8] T. Moranduzzo and F. Melgani, "Detecting cars in UAV images with a catalog-based approach," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[9] Lohia, Aditya; Kadam, Kalyani Dhananjay; Joshi, Rahul Raghvendra; and Bongale, Dr. Anupkumar M., "Bibliometric Analysis of One-stage and Two-stage Object Detection" (2021). Library Philosophy and Practice (e-journal). 4910.

[10] Pan, Sinno Jialin; Yang, Qiang (2010). “A Survey on Transfer Learning”. IEEE Transactions on Knowledge and Data Engineering, vol.22 no.10, pp. 1345–1359. doi:10.1109/TKDE.2009.191

[11] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

[12] Girshick, Ross. "Fast r-cnn." in Proceedings of the IEEE international conference on computer vision, pp. 1440-1448, 2015.

[13] Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." in International Conference on Neural Information Processing Systems , pp. 91-99, 2015.

[14] He, Kaiming, et al. "Mask R-CNN." IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 99, pp. 1-1, 2017.

[15] Redmon, Joseph, et al. "You only look once: Unified, real- time object detection." in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.

[16] Liu, Wei, et al. "Ssd: Single shot multibox detector." in European conference on computer vision. Springer, Cham, pp. 21- 37, 2016.

[17] Manuel Carranza-García; Jesús Torres-Mateo; Pedro, Lara-Benítez; Jorge García-Gutiérrez; (2020). “On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data . Remote Sensing”, (), –. doi:10.3390/rs13010089

[18] Nhat-Duy Nguyen, Tien Do, Thanh Duc Ngo, Duy-dinh Le, “an Evaluation of Deep learning methods for small object detection”,Journal of Electrical and Computer Engineering, vol. 2020, Article ID 3189691, 18pages, 2020. https://doi.org/10.1155/2020/ 3189691

[19] Munawar,H.S.;Ullah,F.; Qayyum, S.; Heravi, A. Application of Deep Learning on UAV-Based Aerial Images for Flood Detection. Smart Cities 2021, *4*, 1220–1242. https://doi.org/10.3390/ smartcities4030065

[20] Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., & Murphy, R. R. (2021). *FloodNet: A* High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. IEEE Access*, 9, 89644–89654.*

[21] V. V. Khryashchev, V. A. Pavlov, A. Priorov and A. A. Ostrovskaya, "Deep Learning for Region Detection in High-Resolution Aerial Images," in Proceedings of *IEEE East-West Design & Test Symposium (EWDTS)*, 2018, pp. 1-5, 2018.

[22] Song, Shiran & Jianhua, Liu & Yuan, Liu & Feng, Guoqiang & Han, Hui & Yao, Yuan & Du, Mingyi. (2020). Intelligent Object Recognition of Urban Water Bodies Based on Deep Learning for Multi-Source and Multi-Temporal High Spatial Resolution Remote Sensing Imagery. Sensors. 20. 397. 10.3390/s20020397.

[23] Akiyama, T. S., Marcato Junior, J., Gonçalves, W. N., Bressan, P. O., Eltner, A., Binder, F., and Singer, T.: DEEP LEARNING APPLIED TO WATER SEGMENTATION, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2020, 1189–1193, https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1189-2020, 2020.

[24] J. Taipalmaa, N. Passalis, H. Zhang, M. Gabbouj and J. Raitoharju, "High-Resolution Water Segmentation for Autonomous Unmanned Surface Vehicles: a Novel Dataset and Evaluation," *in Proceedings of IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1-6, 2019