EDNet: A Mesoscale Eddy Detection Network with Multi-Modal Data

Zhenlin Fan, Guoqiang Zhong*, Hongxu Wei, Haitao Li
Department of Computer Science and Technology
Ocean University of China, Qingdao, China
916056589@qq.com, gqzhong@ouc.edu.cn, weihx1992@163.com, lihaitao@ouc.edu.cn

***Abstract*—Mesoscale eddies play an important role in the transportation and distribution of energy, material and heat in the global ocean. Therefore, mesoscale eddy detection has been researched for a long time. At present, several deep learning models have been proposed for mesoscale eddy detection. However, most of these methods only use single-modal data, while ignoring data of other modals closely related to mesoscale eddy detection. In this paper, we introduce a multi-modal mesoscale eddy dataset, consisting of the satellite data in three modals, i.e., sea surface height (SSH), sea surface temperature (SST) and velocity of flow. Furthermore, we propose an EDNet (Eddy Detection Network), which contains four modules, i.e., multi-modal data fusion module, deep fusion module, region proposal module and head module. We use multi-modal data fusion module to fuse multi-modal data, use deep fusion module to learn the feature representations of the fused multi-modal data and use the region proposal module to generate region proposals containing the mesoscale eddies. There are two branches in the head module, one for classifying and locating the mesoscale eddies, while the other for providing pixel-level instance segmentation of the mesoscale eddies. The experimental results show that EDNet based on multi-modal data fusion significantly improves the accuracy of mesoscale eddy detection over previous approaches.**

***Index Terms*—Deep leaning, Mesoscale eddy detection, Multi-modal data fusion**

I. Introduction

In recent years, deep neural networks [1] have been widely applied in many fields, such as image classification [2], object recognition [3], semantic segmentation [4] and instance segmentation [5]. Techniques such as dropout [6], Bayessian regularization [7] and batch normalization [8] have significantly improved the performance of deep models (e.g., AlexNet [9], VGGNet [10], GoogLeNet [11], ResNet [12] and DenseNet [13]). Particularly, in the field of object detection and instance segmentation, lots of deep models (e.g., Fast R-CNN [14], Faster R-CNN [15], Mask R-CNN [16]) have achieved excellent performances.

In the ocean, mesoscale eddies refer to eddies with a radius of 100 to 300 kilometers and a life span of 2 to 10 months. The mesoscale eddies are usually divided into two types: cyclonic eddies (counterclockwise rotation in the northern hemisphere) and anti-cyclonic eddies (counterclockwise rotation in the southern hemisphere). Because the flow velocity of seawater in mesoscale eddies is several times faster than the average flow velocity of the ocean, mesoscale eddies carry great kinetic energy. Besides, mesoscale eddies can not only bring the nutrients and cold water from the deep ocean to the surface of the ocean, but also press warm water from the sea surface to the deep ocean, playing an important role in the transportation and distribution of energy, material and heat in the global ocean. Therefore, the study of mesoscale eddy detection has very important scientific significance. At present, many mesoscale eddy detection approaches have been proposed [17]–[19].

The early methods for mesoscale eddy detection mainly use satellite remote sensing data [20], [21]. However, they heavily rely on expert analysis, which is laborious and difficult to meet the need of fast and accurate detection. In addition, many traditional mesoscale eddy detection algorithms based on edge detection and flow field geometry are relatively rough, resulting in low detection accuracy for mesoscale eddy detection. In recent years, some methods based on deep neural networks have been proposed. Nevertheless, the architecture of these deep learning models are simple and only single-modal data are used for mesoscale eddy detection. Hence, the detection results are generally not precise.

In order to solve the above problems, we propose a novel deep learning model based on multi-modal data fusion for mesoscale eddy detection. Firstly, we build a dataset consisting of the sea surface height (SSH), the sea surface temperature (SST) and the velocity of flow, which are closely related to mesoscale eddy detection. Considering that the current deep learning methods mainly use SSH to detect mesoscale eddies, SSH images are labeled by experts to ease comparison. In the ground truth image, a mesoscale eddy is annotated with a bounding box and a mask for pixel-level segmentation, where the class of the corresponding eddy is also shown. There are two classes, cyclonic eddies and anti-cyclonic eddies. More importantly, the pixel-level segmentation via masks is instance segmentation, which detects each mesoscale eddy as different instances. An example of the ground truth is shown in Fig. 1. Furthermore, we propose an end-to-end mesoscale eddy detection network named EDNet, which contains four modules, i.e., multi-modal data fusion module, deep fusion module, region proposal module and head module. For concreteness, the multi-modal data fusion module fuses multi-modal mesoscale eddy data, the deep fusion module learns the

* Corresponding author

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

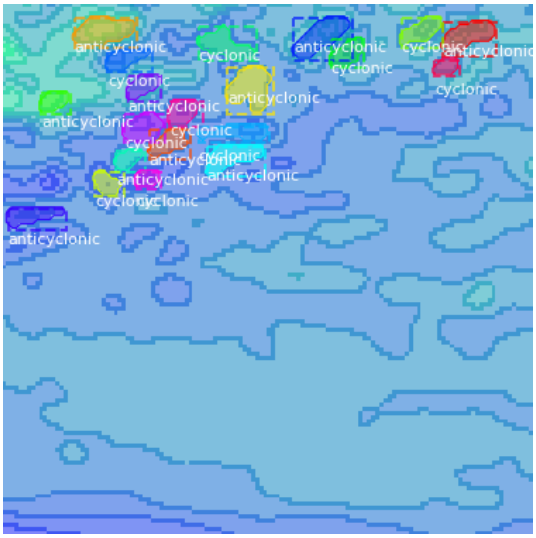


Fig. 1. An example of the ground truth for mesoscale eddy detection.

effective representations of the mesoscale eddies, the region proposal module generates the region proposals containing the mesoscale eddies, and the head module is responsible for classifying, locating and segmenting mesoscale eddies. To our best, our method is the first to use multi-modal dataset for mesoscale eddy detection, and is also the first to detect mesoscale eddies with bounding boxes and masks.

To sum up, the main contributions of our work are:

- We build the first multi-modal mesoscale eddy detection dataset, which is labeled with bounding boxes and masks by experts.
- We propose an end-to-end EDNet, which contains multi-modal data fusion module, deep fusion module, region proposal module and head module, achieving mesoscale eddy detection with both bounding boxes and masks.
- Our work is the first to apply object detection and instance segmentation to mesoscale eddy detection, attaining state-of-the-art performance of mesoscale eddy detection on the multi-modal dataset collected by us.

II. Related Work

In this section, we review some related mesoscale eddy detection algorithms, including the traditional approaches and deep learning approaches.

A. Traditional Approaches

In the early days, mesoscale eddy detection methods mainly relied on experts to manually annotate the remote sensing data. However, this method is time consuming and laborious. Later, mesoscale eddy detection methods using the powerful computing ability of computer began to emerge. Nichol proposed a method which takes advantage of gray scale values of images to extract an eddy-like structure [17]. Ji *et al.* used ellipse edge detection to detect mesoscale eddies based on oceanic remote sensing images [18].

Additionally, there are some mesoscale eddy detection methods based on satellite remote sensing data of different types. These methods are mainly divided into two categories, methods using Euler data and that using Lagrangian data. The Euler data are used to monitor the instantaneous characteristics of the ocean at a certain moment, while the Lagrangian data can record the movement track information of material particles or water masses in a certain period of time. The main methods using Euler data are that based on physical parameters [22] and that based on flow field geometry [23]. The main methods using Lagrangian data belong to the Lagrangian stochastic model [24]. However, these methods are relatively rough, in that the mesoscale eddies may be merged together incorrectly.

In order to solve this problem, Faghmous *et al.* proposed a method named EddyScan, which detects mesoscale eddies by determining the largest possible closed contour of the eddies [25]. However, this method only use single-modal data (SSH) to detect mesoscale eddies.

Although traditional approaches have made great progress so far, they need to be improved with respect to the accuracy of mesoscale eddy detection.

B. Deep Learning Approaches

With the wide application of deep learning, mesoscale eddy detection methods based on deep learning emerge. At present, there are not many deep learning algorithms applied to mesoscale eddy detection. Lguensat *et al.* proposed a network named EddyNet [26], which classified each pixel based on a U-Net structure, completing the semantic segmentation of the mesoscale eddies. Recently, Xu *et al.* used pyramid scene parsing network (PSPNet) [27] to detect mesoscale eddies [28], performing pixel-level classification to locate mesoscale eddies same as EddyNet. However, the EddyNet and PSPNet detect mesoscale eddies using the idea of semantic segmentation, which only divide all pixels into three categories, i.e., cyclonic eddies, anticyclonic eddies and background. They cannot detect different mesoscale eddy as different instances, resulting in that overlapping mesoscale eddies belonging to the same class cannot be separated. Besides, only single-modal data (SSH) are used for mesoscale eddy detection in these two methods, while the data of other modals that are helpful for detection are ignored. Additionally, Du *et al.* proposed a method named DeepEddy [19], which used synthetic aperture radar (SAR) images with a high spatial resolution for mesoscale eddy detection. This work mainly applies the principal component analysis network (PCANet) [29]. However, it only classify a mesoscale eddy in a SAR image, not detecting the eddies in a region simultaneously. To summarize, the EddyNet and PSPNet can only achieve semantic segmentation of mesoscale eddies, while the DeepEddy can only classify a single SAR image to be mesoscale eddy or not. In comparison, our proposed EDNet is capable of achieving instance segmentation and location with masks and bounding boxes, respectively, for mesoscale eddies.

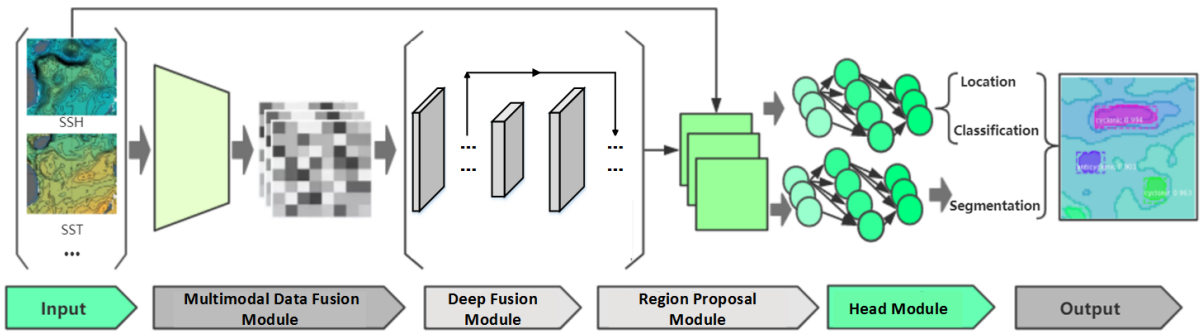


Fig. 2. The structure of EDNet. There are mainly the multi-modal data fusion module, deep fusion module, region proposal module and head module, except the input and output layers in EDNet.

To our best, there is no method to use multi-modal data for mesoscale eddy detection, and also no method to use both bounding boxes and masks to detect mesoscale eddies so far. Therefore, our method is not only the first to use multi-modal data for mesoscale eddy detection, but also the first to apply object detection and instance segmentation to mesoscale eddy detection.

III. Method

In this section, we introduce the structure of the proposed EDNet in detail. As shown in Fig. 2, there are mainly four parts except the input and output layers in EDNet. The first part is a multi-modal data fusion module, which is responsible for fusing the multi-modal mesoscale eddy data. The second part is a deep fusion module, which is used to further learn the effective representations of the mesoscale eddies. The third part is the region proposal module, which finds region proposals containing mesoscale eddies. The last part is the head module that consists of two branches: one can regress bounding boxes to locate and classify mesoscale eddies, and the other can segment the mesoscale eddies in pixel level.

A. The Architecture of EDNet

In the following, we introduce the four parts of EDNet in detail.

1) Multi-Modal Data Fusion Module: The multi-modal data including the SSH, SST and the velocity of flow describe different characteristics of mesoscale eddies. As they are closely related to mesoscale eddy detection, using multi-modal data for mesoscale eddy detection can effectively and accurately detect mesoscale eddies. Hence, for EDNet, we design a multi-modal data fusion module to fuse multi-modal data.

Firstly, we use different hyperparameters (α, β, γ) to scale the data in different modals to explore the importance of different modals in mesoscale eddy detection:

$$H = h * \alpha, \tag{1}$$

$$T = t * \beta, \tag{2}$$

$$V = v * \gamma, \tag{3}$$

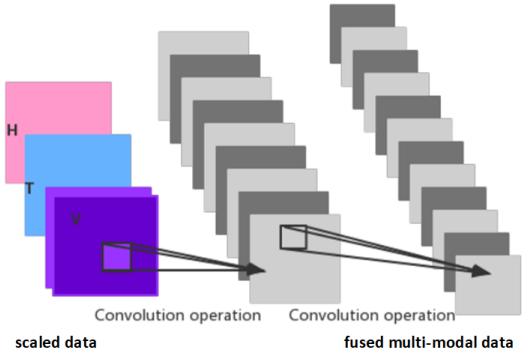


Fig. 3. The structure of the multi-modal data fusion module. We use two convolutional operations to fuse the multi-modal data.

where $\alpha + \beta + \gamma = 1$, h denotes the SSH, t the SST, and v the velocity of flow. H is the scaled SSH, T is the scaled SST, and V is the scaled velocity of flow. Then we use the multi-layer convolutional network to fuse the scaled data. The architecture is shown in Fig. 3. Finally, we can obtain the fused multi-modal data from the multi-modal data fusion module.

2) Deep Fusion Module: After going through the multi-modal data fusion module, the fused multi-modal data are input to the deep fusion module, which is used for feature extraction of mesoscale eddies. We use ResNet-101 [12] as the backbone of the deep fusion module. However, due to lots of convolutional operations in ResNet-101, fine-grained information may be lost as the resolution of feature maps decreases. Thus, we fuse feature maps of high resolution with rich contextual information and feature maps of low resolution with rich semantic information, which is illustrated in Fig. 4. In particular, there are 5 blocks in ResNet-101. We denote outputs of conv2, conv3, conv4, and conv5 as C2, C3, C4 and C5, respectively. Firstly, C5 upsampled by a factor of 2 is fused with C4 undergoing a 1×1 convolutional layer by element-wise addition. Then, we iterate this process until C3 is fused with C2. To the end, feature maps with rich contextual information and semantic information are generated. The

output of conv1 is ignored considering that feature maps of high resolution cause a large amount of computation.

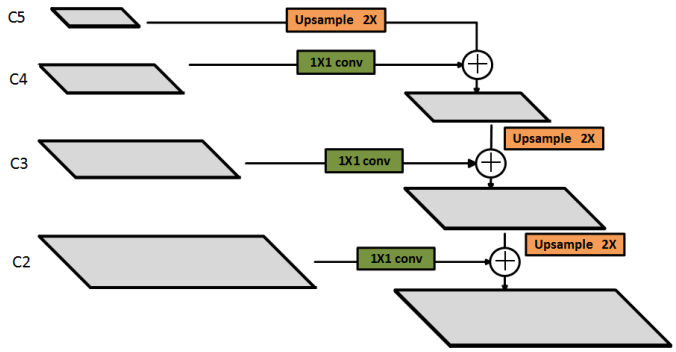


Fig. 4. The structure of the deep fusion module.

In summary, feature maps with rich contextual information and semantic information can be obtained from the deep fusion module, which is beneficial to the subsequent region proposal module and head module.

3) Region Proposal Module: We design a region proposal module, which can obtain multiple proposals that contain mesoscale eddies with high probability. The region proposal module can be regarded as a rough object detection network, which takes advantage of anchor boxes with different scales and aspect ratios to locate and classify mesoscale eddies.

Firstly, we take the feature maps from the deep fusion module as input, and run a 3×3 sliding window on the feature maps. As shown in Fig. 5, we generate 9 anchor boxes with three scales and three aspect ratios (1:1, 1:2 and 2:1) for every center pixel of the sliding window. Then we classify and regress every anchor box with two 1×1 convolutional operations. One 1×1 convolutional layer classify whether the anchor box is the foreground or the background, which is just a two-class classification task. The other 1×1 convolutional layer regresses every anchor box to perform the first coordinate correction, making anchor box close to the ground truth box. Finally, region proposals containing mesoscale eddies with high probability are obtained by this region proposal module.

4) Head Module: There are two branches in the head module, which is shown in Fig. 6. We input region proposals from the region proposal module into these two branches, achieving mesoscale eddies classification, location with bounding boxes and instance segmentation in pixel-level.

Before inputting region proposals into the head module, we use bilinear interpolation to obtain feature maps with fixed size. Then the feature maps are input into two branches in the head module. In one branch, we use fully connected networks to perform feature extraction of feature maps, then obtain specific classes (cyclonic eddies or anti-cyclonic eddies) and regress the bounding boxes of proposals for the second time by two sibling fully connected layers. As shown in Fig. 7, proposal is regressed to predicted bounding box by this branch, which is closer to ground truth box than the original proposal. In the other branch, fully convolutional networks [4] are added to generate masks prediction for mesoscale eddies in region proposals.

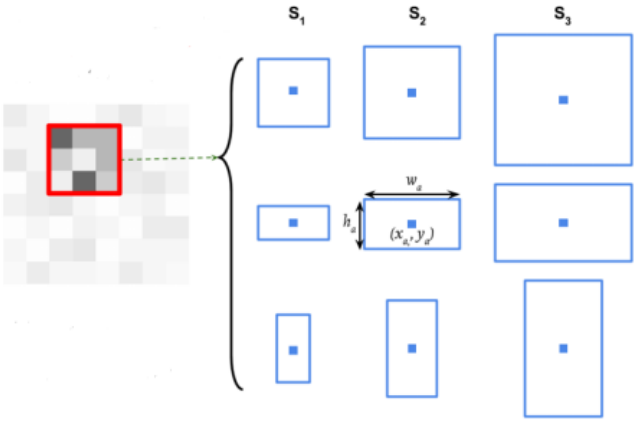


Fig. 5. We generate 9 anchor boxes with three scales and three aspect ratios (1:1, 1:2 and 2:1) for a center pixel of the sliding window. (x_a, y_a) is the center pixel of the sliding window, w_a, h_a are the width and height of the anchor box based on different scales and different aspect ratios.

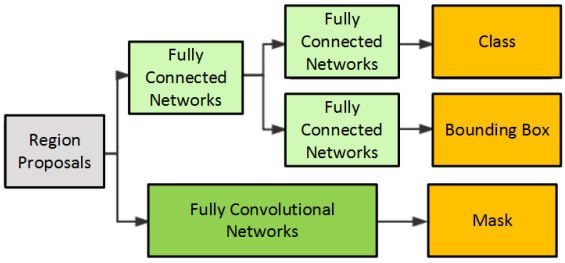


Fig. 6. The structure of the head module. We input region proposals from the region proposal module into the head module, obtaining class, bounding boxes and masks about mesoscale eddies.

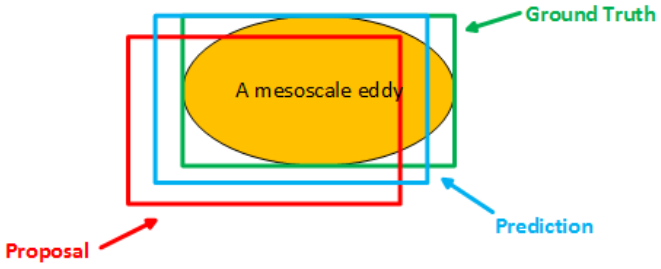


Fig. 7. The proposal, prediction and ground truth about a mesoscale eddy. We use the head module to regress the bounding box of proposal, obtaining bounding box of prediction which is close to the ground truth box.

To the end, we achieve mesoscale eddy detection with bounding boxes and masks by the whole EDNet.

B. The Loss Function

Based on three outputs from the EDNet, the loss function includes three parts, which can be written as follows:

$$L = L_{cls} + L_{box} + L_{mask}. \tag{4}$$

Among them, L_{cls} and L_{box} refer to the loss of classification and bounding boxes prediction, which are determined by the region proposal module and head module, as there are class prediction and bounding box regression in both these two modules. L_{cls} is a cross-entropy loss function between prediction and ground truth, which is defined as:

$$L_{cls} = CrossEntropy(p, q), \tag{5}$$

where p, q represent the labels of prediction and ground truth separately. The classification in the region proposal module is a binary classification problem, while the classification in the head module is a multi-class classification problem. L_{box} is a smooth $l1$ loss function about bounding box regression, which is defined as follows:

$$L_{box} = g(t - t^*), \tag{6}$$

where t denotes the offset of predicted bounding box from the original bounding box, t^* denotes the offset of ground truth from the original bounding box. Besides, the smooth l1 loss function can avoid exceeding the optimal value due to the decrement of gradient, which is defined as follows:

$$g(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{7}$$

L_{mask} is an average binary cross-entropy loss function about the mask branch in the head module. L_{mask} is defined as:

$$L_{mask} = CrossEntropy(p^*_{i,k}, q^*_{i,k}), \tag{8}$$

where $p^*_{i,k}$ refers to the probability of the i-th pixel belonging to the k-th class, and $q^*_{i,k}$ is the probability from the ground truth.

In summary, we minimize the loss function consisting these three parts to attain good performance in mesoscale eddy detection.

IV. Experiments

We report the experimental results in this section. Firstly, the collected and labeled multi-modal mesoscale eddy dataset is introduced in detail. Then we present the parameter settings. Finally, we demonstrate the effectiveness of the proposed EDNet with comparison to the state-of-the-art.

A. The Collected and Labeled Dataset

In our work, we build a multi-modal dataset consisting of the satellite data of three modals, i.e., SSH, SST and velocity of flow, which are closely related to mesoscale eddy detection. Firstly, we download the SSH, SST and velocity of flow for a total of ten years from January 2000 to December 2009 in CMEMS, all of which are from the same sea area. The dimensions of these three-modal data are $681 \times 1440 \times 120$, where 681 is the dimension of the latitude, 1440 is the dimension of the longitude, and 120 indicates that the data come from 120 consecutive months. Among the data of these three modals, the SSH and SST occupy one channel respectively, while the velocity of flow occupies two channels, because velocity of flow contains two directions, which can be understood as the velocity vector of a certain point in the ocean is decomposed into the east/west direction and north/south direction. Therefore, there are four channels in our multi-modal data. Then, we choose the data of 40 months for a three-month interval of 120 months to make the data diverse. Lastly, we randomly select the data of these three modals from multiple regions in each month without repetition, and the regions selected are resized to 128×128. In addition, the corresponding position of the SSH, SST and the velocity of flow from the same multi-modal data is consistent. In order to compare our method with those methods only use single-modal data (SSH) for mesoscale eddy detection, SSH images are labeled by experts as ground truth.

In the end, to reduce the overlapping regions of the selected samples, we only randomly select 512 samples as our multi-modal mesoscale eddy dataset. All the data are divided into training set and test set according to the ratio 7:3. Based on the experimental results in Section IV.C, we can see that the small amount of data randomly selected in our dataset are enough to prove the superiority of our method.

B. Parameter Settings

In our work, we perform experiments on two GPUs, and the computational time is greatly reduced. We pre-train our EDNet on the COCO dataset, initializing it for training on mesoscale eddy detection. The optimization method used in our method is stochastic gradient descent (SGD). We set three anchor ratios as 1:1, 1:2 and 2:1, which are commonly used in object detection. However, we need define three anchor scales and learning rate by experiments.

For simplicity, only SSH is used as the training data for model optimization. During training, we train the head module for 50 epochs firstly and then train the whole network (EDNet) for 100 epochs. According to the experiment on the COCO dataset, we find that the learning rate of the head module has little effect on the prediction results, so that we follow the pre-trained network to use 0.001 as the learning rate of the head module. Therefore, only anchor scales and the learning rate of the whole network are not determined. For the size of our collected dataset, we only compare anchor scales (8, 32, 128) and (4, 16, 64), and learning rates 0.001 and 0.0001 which are most commonly used for object detection tasks. Then we use the optimal hyper-parameters as the baseline model for subsequent training. Training results based on different hyper-parameters are shown in Table I.

As can be seen from the table, we obtain the best mAP of 56.35% when the anchor scale is $(4, 16, 64)$ and the initial learning rate of the whole network is 0.0001.

C. Experimental Results

1) Multi-Modal Data Fusion Experiments: The experiment in this section is designed to verify the effectiveness of multi-modal data fusion. α, β and γ are hyper-parameters of the SSH, SST and velocity of flow, respectively, where α + β + γ

TABLE I
THE RESULTS OBTAINED BY USING DIFFERENT ANCHOR SCALES AND LEARNING RATES.

anchor scale	learning rate of the whole network	mAP %
$(8, 32, 128)$	0.001	45.50
$(4, 16, 64)$	0.001	53.89
$(8, 32, 128)$	0.0001	52.62
(4, 16, 64)	**0.0001**	**56.35**

$= 1$. Hyper-parameters not only reflect the importance of the data of these three modals on mesoscale eddy detection, but also demonstrate whether multi-modal data fusion is helpful for mesoscale eddies detection or not. We try a lot of different combinations of α, β and γ. The experimental results are shown in Table II.

TABLE II
THE RESULTS OBTAINED BY USING DIFFERENT COMBINATIONS OF α, β AND γ.

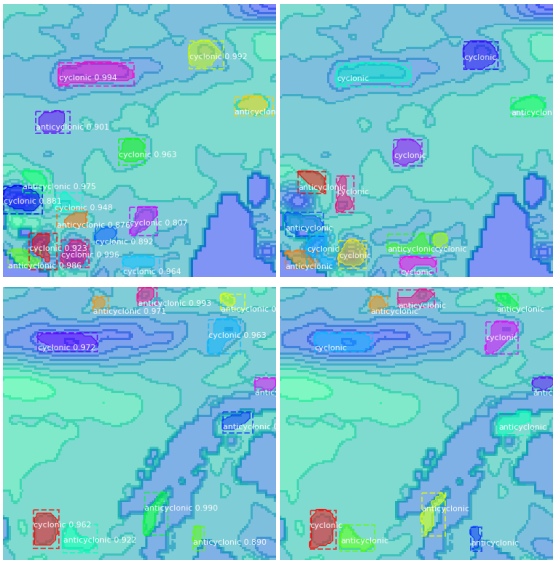
$\alpha : \beta : \gamma$	mAP %
0 : 1 : 0	44.91
0 : 0 : 1	49.26
1 : 0 : 0	56.35
0 : 1 : 1	52.91
1 : 0 : 1	57.43
1 : 1 : 0	56.48
1 : 1 : 1	60.48
1 : 1 : 2	59.62
1 : 2 : 1	58.49
2 : 1 : 1	66.85
2 : 2 : 1	58.69
2 : 1 : 2	58.70
1 : 2 : 2	57.43
2 : 3 : 5	59.73
2 : 5 : 3	57.75
3 : 2 : 5	60.76
3 : 5 : 2	60.15
5 : 3 : 2	67.47
5 : 2 : 3	**70.27**

According to the experimental results, we can see that the best mAP of 70.27% is obtained when the ratio of α, β and γ is 5:3:2. It is also observed in the Table II that multi-modal data fusion is better than using single-modal data for detecting mesoscale eddies.

2) Comparison between the Mesoscale Eddies Detection Algorithms: Based on the selected optimal hyper-parameters, we compare different mesoscale eddy detection algorithms with the proposed network to demonstrate the effectiveness of EDNet. Since there are not many deep learning methods for mesoscale eddy detection, we choose EddyNet and PSPNet for comparison, which can detect mesoscale eddies in pixel-level. However, the EddyNet and PSPNet achieve semantic segmentation, while the proposed EDNet can achieve instance segmentation. Hence, we evaluate our method at the level of semantic segmentation for fair comparison. The final experimental results obtained by using different networks on our multi-modal mesoscale eddy dataset are shown in Table III. As can be seen from the table, our proposed network yields 86.18% in terms of mAP at the level of semantic segmentation, which is 5.4% higher than EddyNet and 4.12% higher than PSPNet. Therefore, our method achieves better performance than the other methods. Some results of mesoscale eddy detection using our method can be seen in Fig. 8. Through comparison with the ground truth, it is clear that results of mesoscale eddy detection using our method is very close to the ground truth.

TABLE III
THE RESULTS OBTAINED BY USING DIFFERENT NETWORKS ON OUR MULTI-MODAL MESOSCALE EDDY DATASET.

Methods	mAP %
EddyNet	80.78
PSPNet	82.06
EDNet	**86.18**



(a) Our method (b) Ground truth

Fig. 8. Results of mesoscale eddy detection using our method on our multi-modal dataset. (a) Examples of mesoscale eddy detection results using our method; and (b) The ground truth labeled by experts.

V. CONCLUSION

In this paper, we introduce a novel multi-modal mesoscale eddy detection dataset, which contains the SSH, SST and velocity of flow. In addition, a mesoscale eddies detection model named EDNet is proposed. EDNet contains four modules, i.e., multi-modal data fusion module, deep fusion module, region proposal module and head module. The proposed EDNet is the first using bounding boxes and masks to detect mesoscale eddies, achieving a combination of object detection and instance segmentation. In the consequence, the experimental results by using EDNet on the multi-modal mesoscale eddy dataset are much better than that obtained by previous approaches for mesoscale eddy detection.

ACKNOWLEDGMENTS

This work was supported by the Major Project for New Generation of AI under Grant No. 2018AAA0100400, the National Key R&D Program of China under Grant No.2016YFC1401004, the National Natural Science Foundation of China (NSFC) under Grant No. 41706010, the Joint Fund of the Equipments Pre-Research and Ministry of Education of China under Grant No. 6141A020337, the Open Project Program of Key Laboratory of Marine Hazards Forecasting, Ministry of Natural Resources, under Grant No. LOMF1802, and the Fundamental Research Funds for the Central Universities of China.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *CVPR*, 2012, pp. 3642–3649.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014, pp. 580–587.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[5] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *CVPR*, 2017, pp. 2359–2367.

[6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[7] F. Burden and D. Winkler, "Bayesian regularization of neural networks," *Neural Networks*, vol. 83, pp. 75–85, 2016.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015, pp. 448–456.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017, pp. 4700–4708.

[14] R. Girshick, "Fast r-cnn," in *ICCV*, 2015, pp. 1440–1448.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *NIPS*, 2015, pp. 91–99.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017, pp. 2961–2969.

[17] D. G. Nichol, "Autonomous extraction of an eddy-like structure from infrared images of the ocean," *IEEE Transactions on Geoscience and Remote Sensing*, no. 1, pp. 28–34, 1987.

[18] G. Ji, X. Chen, Y. Huo, and T. Jia, "A automatic detection method for mesoscale eddies in ocean remote sensing image," *Ocean and Lake*, vol. 33, no. 2, pp. 139–144, 2002.

[19] Y. Du, W. Song, Q. He, D. Huang, A. Liotta, and C. Su, "Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection," *Information Fusion*, vol. 49, pp. 89–99, 2019.

[20] S. H. Peckinpaugh and R. J. Holyer, "Circle detection for extracting eddy size and position from satellite imagery of the ocean," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 267–273, 1994.

[21] A. Fernandes and S. Nascimento, "Automatic water eddy detection in sst maps using random ellipse fitting and vectorial fields for image segmentation," in *DS*, 2006, pp. 77–88.

[22] J. Isern-Fontanet, E. García-Ladona, and J. Font, "Identification of marine eddies from altimetric maps," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 772–778, 2003.

[23] D. B. Chelton, M. G. Schlax, and R. M. Samelson, "Global observations of nonlinear mesoscale eddies," *Progress in Oceanography*, vol. 91, no. 2, pp. 167–216, 2011.

[24] M. Lankhorst, "A self-contained identification scheme for eddies in drifter and float trajectories," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 11, pp. 1583–1592, 2006.

[25] J. H. Faghmous, L. Styles, V. Mithal, S. Boriah, S. Liess, V. Kumar, F. Vikebø, and M. dos Santos Mesquita, "Eddyscan: A physically consistent ocean eddy monitoring application," in *CIDU*, 2012, pp. 96–103.

[26] R. Lguensat, M. Sun, R. Fablet, P. Tandeo, E. Mason, and G. Chen, "Eddynet: A deep neural network for pixel-wise classification of oceanic eddies," in *IGARSS*, 2018, pp. 1764–1767.

[27] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *CVPR*, 2017, pp. 2881–2890.

[28] G. Xu, C. Cheng, W. Yang, W. Xie, L. Kong, R. Hang, F. Ma, C. Dong, and J. Yang, "Oceanic eddy identification using an ai scheme," *Remote Sensing*, vol. 11, no. 11, p. 1349, 2019.

[29] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.