Oceanic Eddy Identification Using Pyramid Split Attention U-Net with Remote Sensing Imagery

Nan Zhao, Baoxiang Huang*, *Member, IEEE*, Jie Yang, Milena Radenkovic, *Member, IEEE*, and Ge Chen

***Abstract*—Oceanic eddy is the ubiquitous ocean flow phenomenon, which has been the key factor in the transportation of ocean energy and materials. Consequently, oceanographic understanding can be enhanced by the intelligent identification of eddy. The state-of-the-art deep learning technologies are gradually improving the identification methods. This letter proposes the pyramid split attention(PSA) eddy detection U-Net architecture (PSA-EDUNet) that targets oceanic eddy identification from ocean remote sensing imagery. As for the PSA-EDUNet, its inspiration comes from U-Net, which contains encoder and decoder parts, making the integration of inferior and senior features efficiently and ensuring the feature information will not be lost in large quantities through non-linear connection mode. Meanwhile, PAS module is introduced to enhance feature extraction. In terms of the fusion data, the sea surface feature is the main criterion of eddy identification, including sea surface temperature (SST) and sea level anomaly (SLA). The experiments are implemented on the Kuroshio Extension and the South Atlantic regions, the results demonstrate that the proposed method can outperform other methods, especially for eddy edges and small-scale eddies.**

***Index Terms*—Oceanic eddy identification, Deep learning, Pyramid split attention, U-Net network**

I. Introduction

MESOSCALE eddies are the rotating structures that exist in the global ocean and are a common and complex ocean flow phenomenon, and provide most of the kinetic energy for global ocean circulation [1]. Moreover, mesoscale eddies also affect the mixing and transport of salt [2], heat [3], carbon [3], and other nutrients in the ocean [4]. The effects of mesoscale eddies on ocean dynamics, energy, and large-scale circulation influence global climate change [5]. Especially, eddies have been shown to affect the winds, clouds, and precipitation on the local sea surface to a certain degree through their effects on atmospheric boundary layer turbulence [6]. Accordingly, the identification of mesoscale eddies has a unique significance and research value in the field of ocean science.

According to the direction of rotation, eddies are traditionally divided into cyclonic eddies (CEs) and anticyclonic eddies (AEs), corresponding to negative (positive) SLA. Furthermore, due to the rotation of the earth, the sea water on the surface of the CEs (AEs) is divergent (convergent), which makes the lower (upper) sea water rise (fall) as a supplement, thus making the sea surface present a low (high) SST [7]. Consequently, the distinct characteristics of oceanic eddies can be identified from SLA and SST. With the gradual maturity of satellite observation technology, the spatial resolution of sea surface feature data is competent for eddy identification [8]. The detection methods of mesoscale eddies can be mainly divided into physical parameter based methods [9], geometric contour based methods [10] and artificial intelligence (AI) based methods [11]. Among the AI methods, deep learning technology is the most popular.

In terms of AI methods, Lguensat et al. [12] put forward an EddyNet model rooted in the U-Net architecture [13] for automatic eddy detection and classification. Later, Lguensat et al. proposed EddyResNet and EddyVNet models that outperformed EddyNet respectively [14]. Xu et al. [15] applied the pyramid scene parsing network to eddy detection and identification, which can fully utilize global information to capture details, and identify AEs and CEs from sea surface height anomaly (SSHA) data. The eddy detection models with an encoder-decoder structure are summarized in [16]. These exploratory researches [17] can provide insights and feasibility for eddy identification using deep learning [18], however, the acquisition of edge information and small-scale eddies remains a challenge [19].

This letter aims to achieve a high-precision eddy identification method based on U-Net architecture, adopting the pyramid split attention module and skip connection method. The remaining parts are organized as follows. Section II describes the study area and materials. The architecture of the proposed PSA-EDUNet is depicted in Section III. Section IV presents a detailed analysis of the segmentation results while presenting the eddy identification results of different methods. Finally, Section V summarizes this study and points out future research directions.

II. Study Area and Materials

SST data with temporal and spatial resolutions of days and $0.25^\circ \times 0.25^\circ$ was acquired from National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation(OI) SST product. As for the SLA data, it is a grid L4 product published by the Copernicus Marine Environment Monitoring Service (CMEMS), which has the same spatio-temporal resolutions as SST data. The training of the model requires not

N. Zhao is with the Department of Computer Science and Technology, Qingdao University, 266071 China

Corresponding author: B. Huang is with the Department of Computer Science and Technology, Qingdao University, 266071 China, and Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, 266228, China (email: hbx3726@163.com).

J. Yang and G. Chen are with the School of Marine Technology, Institute for Advanced Ocean Study, Ocean University of China, 266075 China, and Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, 266228, China.

M. Radenkovic is with the School of Computer Science and Information Technology, The University of Nottingham, Nottingham, NG8 1BB UK.

only the sea surface feature data but also the eddy labeling data at the same resolution as its feature data, which needed to be constructed from the ocean eddy identification and tracking dataset based on the satellite altimeter provided by the Big Earth Data Science Engineering Project (CASEarth).

The eddy identification and tracking dataset contain eddy trajectory information from 2004 to 2018. This letter selects the data from 2004 to 2017 to form a training set, and data from 2018 to form a testing set for training and testing models. To construct the eddy labels, the following processing was performed. (1) First, extract the eddies whose contour boundary value is not empty and classify them into CEs and AEs according to the eddy label. (2) Second, determine the inner region of each eddy. According to the eddy contour, the thresholds of the eddy latitude and longitude are obtained. Meanwhile, the edge and interior points of the rectangle are determined. Considering each eddy is an irregular polygon, the mathematical ray method is used to judge whether each test point is within the eddy contour. (3) Finally, pixels within each polygon are labeled to represent the category of the eddy to which they belong, 0, 1, and 2 represent Non-eddy, anticyclonic eddy, and cyclonic eddy, respectively.

In this letter, the Kuroshio Extension (KE) and the South Atlantic Ocean (SAO) regions are chosen as study areas. The geographic locations of the two study areas in the global ocean, as well as the SST, SLA, and eddy distributions are shown in Fig.1.

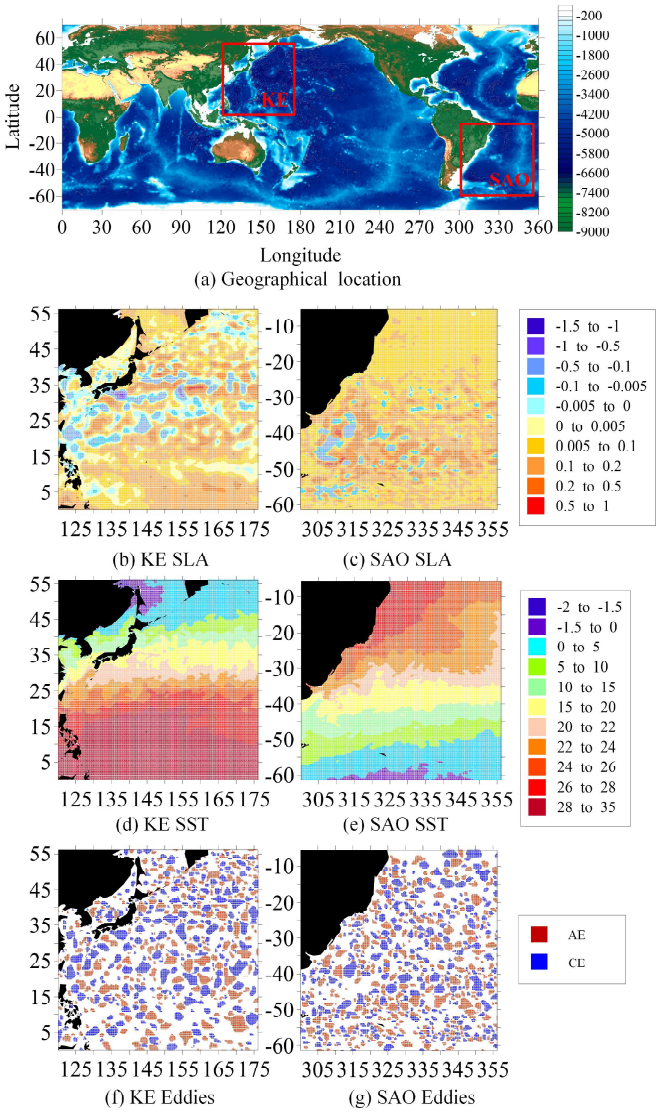

Fig. 1. The geographical locations, SLA data, SST data, and ground truth of the study areas. (a) The location of study areas is highlighted by the red square. (b-g) The SLA, SST, and ground truth of the two study areas, respectively.

III. Methodology

A. Overall architecture of Pyramid Split Attention U-Net

Following encoder-decoder architecture, the PSA-EDUNet based on the U-Net [13] is proposed to implement eddy identification from SLA and SST data. Meanwhile, the PSA [20] module is adopted and the way of skip connections is optimized to achieve high efficiency. The overall structure of PSA-EDUNet is depicted in Fig.2. In addition, all the sub-architectures are described detailedly in the following sections.

1) Encoder: SLA and SST data are input into feature map with channel dimension of 2, and then the hierarchical feature representation is generated through three consecutive encoder hierarchical feature representation (Encoder-HFR) modules. Specifically, each Encoder-HFR consists of an HFR block responsible for feature learning representation and a max pooling layer with kernel size of 2 responsible for down-sampling. The feature map first enters two consecutive convolutional layers with a kernel size of 3. Particularly, each convolution layer is followed by a normalization layer and a Relu activation layer. Then the feature map will be input into the PSA module. In the first stage of Encoder-HFR, the width and height of the feature map become half of the initial size, and the number of channels becomes C. In the remaining two stages of Encoder-HFR, only the resolution changed, while the number of channels remains the same. So the hierarchical features with size and channel of $H/2 \times W/2 \times C$, $H/4 \times W/4 \times C$, and $H/8 \times W/8 \times C$ are obtained, respectively.

2) Bottleneck: HFR blocks are employed for deep feature representation learning at the bottleneck, thus the resolution and dimension of the feature remain unchanged.

3) Decoder: Corresponding to the encoder, the decoder part consists of three hierarchical feature representation (Decoder-HFR) modules. In each Decoder-HFR, the feature map is first input into a transposed convolutional layer with a kernel size of 2 for up-sampling, which doubles the size of the feature map. Next, the feature maps from the up-sampling and the Encoder-HFR are spliced together to obtain the feature map with $4C$ channels. Finally, the spliced feature map is input into an HFR module for feature learning and representation, which keeps the size unchanged, and the number of channels is restored to C. In detail, the resolution and channel number of the feature maps output by the three-stage Decoder-HFR are $H/8 \times W/8 \times C$, $H/4 \times W/4 \times C$, and $H/2 \times W/2 \times C$ respectively. A convolutional layer with a kernel size of 1 followed decoder is applied to map the number of feature channels to the number

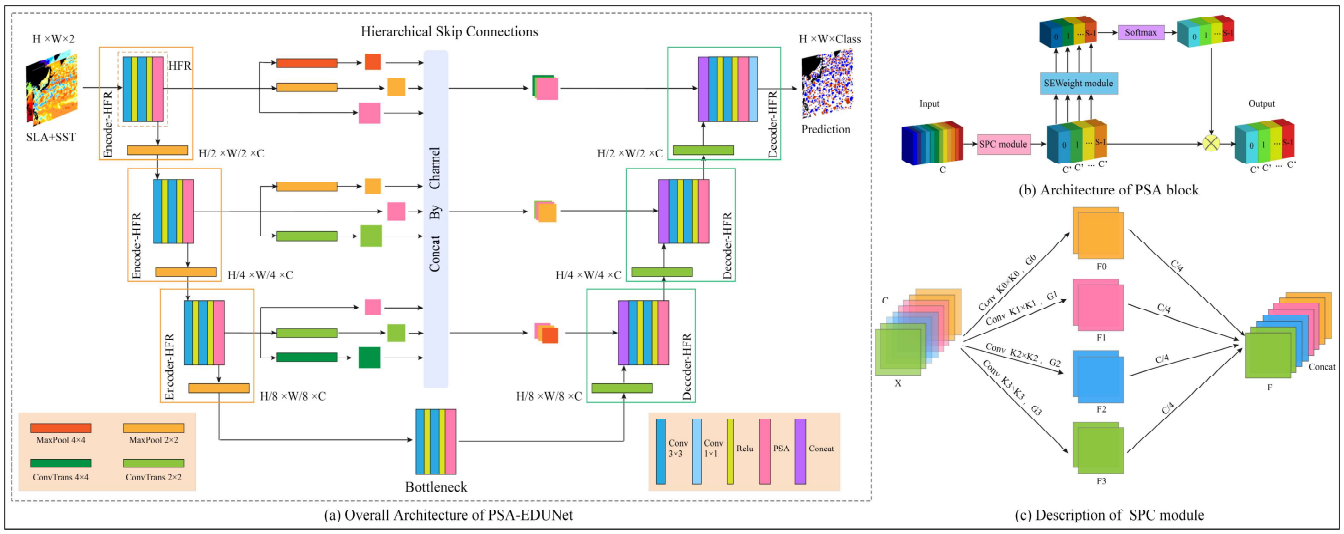

Fig. 2. The architecture of PSA-EDUNet.

of classes predicted by the experimental segmentation, which results in a pixel-level segmentation prediction of the original feature map.

4) Hierarchical skip connections: To prevent learned features from being gradually forgotten as the model depth increases, the way of skip connections is improved. We apply max-pooling layers and transposed convolutional layers with different-sized convolution kernels to the feature maps output by Encoder-HFR at different stages to generate corresponding feature representations. Then we concatenate the hierarchical features of the same resolution generated by the three Encoder-HFRs by channel and input them into the Decoder-HFR of the corresponding feature map resolution.

5) PSA module: Compared with the Squeeze-and-Excitation (SE) Weight module [21], the PSA module [20] is more efficient, as illustrated by Fig.2(b). The advantage of this algorithm lies in that the feature space is enriched by capturing spatial information at different scales, and the channel attention is considered at the same time. Specifically, the PSA module consists of four steps.

First, multi-scale features are extracted through the Squeeze and Concat (SPC) module. A detailed description of the SPC module when S=4, is displayed in Fig.2(c). For the input X, the number of channels of X is C. First, divide X into S parts, which are represented by $F_i(i = 0, 1, ..., S-1)$, the number of channels of each divided part both are $C^{'} = C/S$. It is worth noting that C must be divisible by S. Since the number of parameters can be reduced by grouping convolution, the multi-scale convolution kernel grouping convolution method is used to extract spatial information of different scales for each channel feature map divided. The relationship between group and kernel size is $G = 2^{\frac{K-1}{2}}$, where K and G represent kernel size and group size, respectively. Therefore, the specific calculation method of the multi-scale feature extraction process is as the following Eq. 1.

$$F_i = Conv(k_i \times k_i, G_i)(X) \quad i = 0, 1, ..., S-1 \tag{1}$$

Where $k_i = 2 \times (i+1) + 1$ denotes the ith kernel size, $G_i = 2^{\frac{K_i-1}{2}}$ denotes the ith group size. Moreover, $F_i \in R^{C \times H \times W}$ denotes the feature maps of different scales.

After convolution of different sizes, the obtained feature maps are spliced on the channel to get $F = Cat([F_0, F_1, ..., F_{S-1}])$, where $F \in R^{C \times H \times W}$.

Second, channel-level attention vectors, which can be expressed as Eq. 2, are obtained by extracting attention from feature maps at different scales using the SE-Weight module.

$$Z_i = SEWeight(F_i) \quad i = 0, 1, ..., S-1 \tag{2}$$

Where $Z_i \in R^{C^{'} \times 1 \times 1}$ is the attention weight. Meanwhile, the obtained attention vectors are spliced to better realize the interaction and fuse cross dimensional attention information, as shown in Eq. 3.

$$Z = Z_0 \oplus Z_1 \oplus ... \oplus Z_S - 1 \tag{3}$$

Where $\oplus$ represents the splicing operation of the vector, Z_i is the attention value corresponding to F_i, and Z is the multi-scale attention weight vector.

Third, re-calibration weights for multi-scale channels are obtained by re-calibrating the channel-wise attention vector using Softmax. To achieve the interactive connection between local and global channel attention, Softmax is used for normalization, as shown in Eq. 4.

$$att_i = Softmax(Z_i) = \frac{exp(Z_i)}{\sum_{i=1}^{S-1} exp(Z_i)} \tag{4}$$

Finally, the element-level multiplication operation is applied between the recalibrated weights and corresponding source feature maps to obtain weighted feature maps. The multiplication process is illustrated in Eq. 5.

$$Y_i = F_i \otimes att_i \quad i = 0, 1, ..., S-1 \tag{5}$$

Here, $\otimes$ means the multiplication of channel direction, and Y_i represents the obtained feature map of the attention weights

in the multi-scale channel direction. The final output is $Out = Cat([Y_0, Y_1, ..., Y_{S-1}])$.

B. Evaluation metrics

The pixel level alone cannot reflect the structural information of ocean dynamics. Therefore, the pixel accuracy (PA) and mean dice coefficient (MD) at the pixel level as well as the accuracy, precision, and recall at the structure level are used to evaluate the performance of PSA-EDUNet in an overall way.

IV. Experimental Results

The RMSprop optimizer is used to optimize the experiment. The loss function is the cross entropy loss function, and the learning rate is 1e-3. Each small batch is composed of 8 images. When the training set loss stopped improving for 30 consecutive epochs, the strategy of reducing the learning rate will be employed to improve the learning process. The code is written using the Pytorch deep learning library. Training is performed using an Nvidia GeForce RTX 3090 GPU.

To confirm the performance of the PSA-EDUNet, the comparative experiments results are presented in TABLE I and TABLE II. In the KE region, the average pixel accuracy of our model can reach 91.09%, which is 2.77%, 3.45%, 1.82%, 8.76%, and 3.45% higher than the EddyNet [12], Swin-Unet [22], U-Net [13], DeepLabV3+ [23] and a two branch convolutional neural network (TBCNN) [24] models respectively; while in the SAO region, the average pixel accuracy of our model is 91.27%, which is increased by 1.98%, 4.26%, 2.07%, 9.76% and 2.89% respectively based on EddyNet, Swin-Unet, U-Net, DeepLabV3+, and TBCNN models. The proposed PSA-EDUNet has the most brilliant performance. EddyNet and U-Net models have comparative performance in eddy identification. Swin-Unet and TBCNN models are inferior to the U-Net model. In addition, the ablation experiment for the PSA module has also been completed, and the results in TABLE I and TABLE II can demonstrate the effectiveness of the PSA module.

TABLE I. Metric evaluation results of 6 models on the 2018 KE regional testing set.

Method	PSA-EDUNet		EddyNet [12]	Swin-Unet [22]	U-Net [13]	DeepLabV3+ [23]	TBCNN [24]
	all	no psa					
Region*	KE	KE	KE	KE	KE	KE	KE
Test PA	91.09%	89.88%	88.32%	87.64%	89.27%	82.33%	87.64%
MD	88.65%	87.26%	84.57%	84.17%	86.34%	77.16%	83.98%
Accuracy	94.06%	93.25%	92.21%	91.76%	92.84%	88.22%	91.76%
Precision	88.72%	86.64%	87.04%	84.69%	86.41%	77.94%	85.17%
Recall	88.62%	87.99%	82.70%	83.75%	86.34%	76.52%	82.97%

* Lon: $120.125° - 176.125°E$ Lat: $0.125° - 56.125°N$

TABLE II. Metric evaluation results of 6 models on the 2018 SAO regional testing set.

Method	PSA-EDUNet		EddyNet [12]	Swin-Unet [22]	U-Net [13]	DeepLabV3+ [23]	TBCNN [24]
	all	no psa					
Region*	SAO	SAO	SAO	SAO	SAO	SAO	SAO
Test PA	91.27%	90.54%	89.29%	87.01%	89.20%	81.51%	88.38%
MD	88.95%	88.00%	86.18%	83.52%	86.34%	76.09%	85.22%
Accuracy	94.18%	93.67%	92.86%	91.34%	92.80%	87.67%	92.25%
Precision	89.13%	88.26%	87.57%	87.57%	86.40%	77.11%	85.66%
Recall	88.89%	87.74%	84.98%	83.25%	86.35%	75.23%	84.91%

* Lon: $3.875° - 59.875°W$ Lat: $5.375° - 61.375°S$

Fig.3 is the learning curve on the training set. Combined with the evaluation results in TABLE I and TABLE II, the loss value of the U-Net model is lower than that of PSA-EDUNet, but the test results are not as good as PSA-EDUNet, which shows that the model proposed in this letter has better generalization ability than U-Net. In addition, the loss value obtained by PSA-EDUNet on the training set and the results of the testing set are superior to the EddyNet, DeepLabV3+ and TBCNN models, which indicates that the proposed PSA-EDUNet has excellent learning ability and can adapt to different situations of dataset to learn the features of the data.

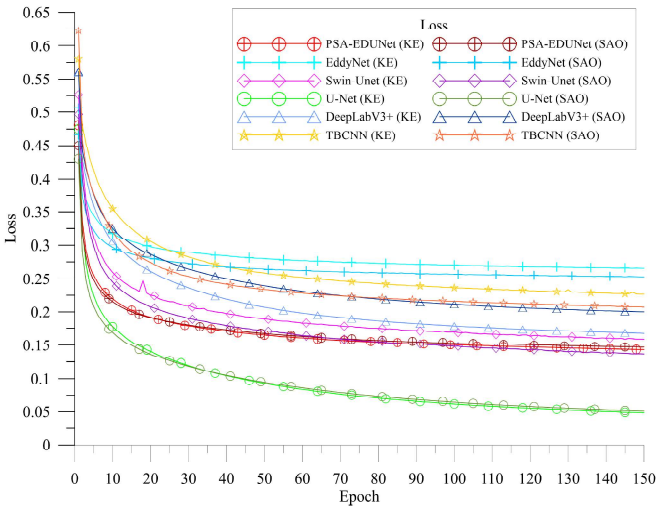

Fig. 3. Learning curves of the different methods. The x-axis represents the quantity of epochs in the training process, and the y-axis represents the loss value.

For the convenience of comparison, we take the last day of 2018 as an example to draw the eddy identification results in the KE region. As shown in Fig.4 and Fig.5, the results demonstrate that deep learning methods can identify most of the eddy regions, and the segmentation results of the PSA-EDUNet are most approximate to the ground truth, especially in the segmentation of small-scale eddies and eddy edges.

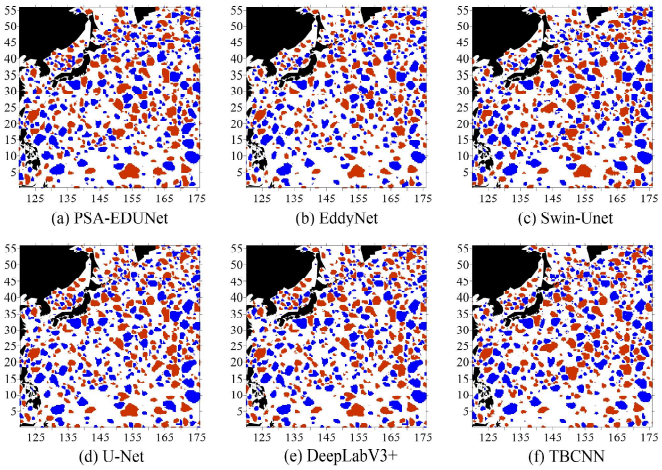

Fig. 4. Oceanic eddy identification results of different methods in the KE region on 31 December 2018: AEs (red), CEs (blue), Non-eddy (white).

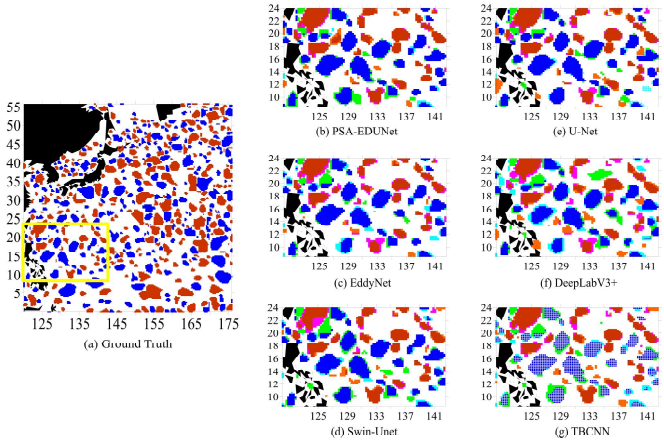

Fig. 5. Oceanic eddy identification results of different methods in the local area of KE region on 31 December 2018: AEs (red), CEs (blue), Non-eddy (white), unrecognized AEs (pink) and CEs (cyan), newly recognized AEs (orange) and CEs (green).

V. Conclusions

This letter builds a multivariate fusion dataset consisting of SLA and SST. With the fused dataset, a high-precision eddy identification model PSA-EDUNet is proposed. The proposed method can capture spatial information of different scales from channel and spatial attention directions, enrich the feature space by the PSA module and improve the way of skip connection, thus ensure the integrity of feature information in the decoder to the greatest extent. In addition, grouped convolution is also employed to avoid the problem of excessive data volume during the learning process. It can be reached from the identification results that the proposed PSA-EDUNet approach is superior to other methods. Meanwhile, the results also indicate the presence of eddies in the region with no eddies on the ground truth.

Acknowledgments

The dataset and Python code of application examples are available at GitHub DBAI to encourage competing methods. This research was financially supported by the International Research Center of Big Data for Sustainable Development Goals (CBAS2022GSP01), the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No.2022QNLM050301), National Natural Science Foundation of China (No.42276203), and the Natural Science Foundation of Shandong Province (No.ZR2021MD001).

References

[1] M. Ding, P. Lin, H. Liu, A. Hu, and C. Liu, "Lagrangian eddy kinetic energy of ocean mesoscale eddies and its application to the northwestern pacific," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[2] R. S. Patel, J. Llort, P. G. Strutton, H. E. Phillips, S. Moreau, P. Conde Pardo, and A. Lenton, "The biogeochemical structure of southern ocean mesoscale eddies," *Journal of Geophysical Research: Oceans*, vol. 125, no. 8, p. e2020JC016115, 2020.

[3] J. Martínez-Moreno, A. M. Hogg, M. H. England, N. C. Constantinou, A. E. Kiss, and A. K. Morrison, "Global changes in oceanic mesoscale currents over the satellite altimetry record," *Nature Climate Change*, vol. 11, no. 5, pp. 397–403, 2021.

[4] A. Mikaelyan, A. Zatsepin, and A. Kubryakov, "Effect of mesoscale eddy dynamics on bioproductivity of the marine ecosystems," *Physical Oceanography*, vol. 27, no. 6, pp. 590–618, 2020.

[5] R. M. van Westen and H. A. Dijkstra, "Ocean eddies strongly affect global mean sea-level projections," *Science advances*, vol. 7, no. 15, p. eabf1674, 2021.

[6] I. Frenger, N. Gruber, R. Knutti, and M. Münnich, "Imprint of southern ocean eddies on winds, clouds and rainfall," *Nature geoscience*, vol. 6, no. 8, pp. 608–612, 2013.

[7] H. Zhang, B. Huang, G. Chen, L. Ge, and M. Radenkovic, "An efficient oceanic eddy identification method with xbt data using transformer," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9860–9872, 2022.

[8] Z. You, L. Liu, B. J. Bethel, and C. Dong, "Feature comparison of two mesoscale eddy datasets based on satellite altimeter data," *Remote Sensing*, vol. 14, no. 1, p. 116, 2021.

[9] 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.

[10] E. Mason, A. Pascual, and J. C. McWilliams, "A new sea surface height–based code for oceanic mesoscale eddy tracking," *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 5, pp. 1181–1188, 2014.

[11] M. D. Ashkezari, C. N. Hill, C. N. Follett, G. Forget, and M. J. Follows, "Oceanic eddy detection and lifetime forecast using machine learning methods," *Geophysical Research Letters*, vol. 43, no. 23, pp. 12–234, 2016.

[12] 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 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 1764–1767.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[14] R. Lguensat, S. Rjiba, E. Mason, R. Fablet, and J. Sommer, "Convolutional neural networks for the segmentation of oceanic eddies from altimetric maps," *Remote Sensing*, vol. 1, no. 1, pp. 1–16, 2018.

[15] 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.

[16] X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang, and F. Wang, "Deep-learning-based information mining from ocean remote-sensing imagery," *National Science Review*, vol. 7, no. 10, pp. 1584–1605, 2020.

[17] G. Xu, W. Xie, C. Dong, and X. Gao, "Application of three deep learning schemes into oceanic eddy detection," *Frontiers in Marine Science*, vol. 8, p. 672334, 2021.

[18] G. Chen, B. Huang, X. Chen, L. Ge, M. Radenkovic, and Y. Ma, "Deep blue ai: A new bridge from data to knowledge for the ocean science," *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 190, p. 103886, 2022.

[19] B. Huang, L. Ge, X. Chen, and G. Chen, "Vertical structure-based classification of oceanic eddy using 3-d convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[20] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "Epsanet: An efficient pyramid squeeze attention block on convolutional neural network," *arXiv preprint arXiv:2105.14447*, 2021.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[22] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537*, 2021.

[23] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[24] R. Hang, G. Li, M. Xue, C. Dong, and J. Wei, "Identifying oceanic eddy with an edge-enhanced multiscale convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9198–9207, 2022.