Automatic Detection of Ocean Eddy based on Deep Learning Technique with Attention Mechanism

Shaik John Saida, Samit Ari, *Member*, *IEEE*

***Abstract*—Ocean eddies are a common occurrence in ocean water circulation. They have an enormous impact on the marine ecosystem. One of the most active study topics in physical oceanography is ocean eddy detection. Although using deep learning algorithms to detect eddies is a recent trend, it is still in its infancy. In this paper, an attention mechanism-based ocean eddy detection approach using deep learning is proposed. Attention mechanism has spatial and channel attention modules that are cascaded to convolution blocks-based encoder model to simulate spatial and channel semantic interdependencies. In the spatial attention module, the feature at each point is aggregated selectively by the sum of the features at all positions. The channel attention module aggregates related data from all channel maps to selectively highlight interdependent channel maps. The original feature map and the feature map obtained through the attention mechanism are appended to enhance the feature representation further, resulting in more accurate segmentation results. The findings of the experiments show that adopting an attention-based deep framework improves eddy recognition accuracy significantly.**

***Index Terms*—Deep learning, semantic segmentation, attention mechanism, eddy detection.**

I. Introduction

The ocean eddies are circular water currents found throughout the world's oceans. They are essential for transporting several ocean traces across the ocean. Eddies influence ocean circulation and biological activity. The cyclonic (anti-cyclonic) eddies correspond to negative (positive) sea surface height (SSH) anomalies. As a result, satellite-measured sea surface height (SSH) images provide information on the eddy characteristics [1], [2], [3]. Automatic eddy detection algorithms derived from sea surface height (SSH) data fall into four categories, physical methods, geometrical methods, sea surface height (SSH) based methods, and hybrid methods. The okubo-weiss (OW) algorithm uses physical parameters to detect eddies automatically [4], [5]. The winding angle (WA) method, based on geometrical attributes, was modified and applied for ocean eddies detection [6], [7]. Another method based on geometrical features is vector geometry (VG) [8], which identifies eddies solely based on the geometrical attributes of ocean current velocity and is unaffected by the parameters calculated using its derivatives. To extract eddies, the vorticity derived from SSH data was processed using an eddy detection approach based on wavelet analysis and projected onto a vorticity space to detect eddies [9]. Because of the dataset's limited dimensions, the eddies detection accuracy of this method is poor. The automatic eddy detection approach based on sea surface height (SSH) is an example of threshold independent eddy detection method [10]. A hybrid detection algorithm combines OW and SSH approaches [11]. When both the OW algorithm and the SSH-based technique are used to initialize cores, the noise impact on the OW algorithm is reduced, and threshold sensitivity is eliminated during initialization. To overcome the shortcomings of aforementioned methods, a machine learning-based strategy for detecting cyclonic and anti-cyclonic eddies was proposed, which is independent of region-specific factors [12]. This machine learning-based approach requires pre-manual eddy labeling, which necessitates specialist expertise. Manual labeling of eddies is a tedious and time-consuming operation.

Artificial intelligence, in particular deep neural network-based models, has demonstrated its ability to solve a wide range of practical issues with tremendous efficiency, including pattern recognition and computer vision, [13], [14], [15], [16], [17], [18], [19]. To find and track maritime eddies in the southwest Atlantic, EddyNet, a network based on the U-Net network, was suggested [20]. A high-resolution network was proposed for detecting eddies from sea surface height data [21]. The aforementioned approaches are unable to fully utilize semantic information to detect eddy borders. Semantic segmentation is a key problem in remote sensing. Mapping of land cover, detection of changes, urban planning, and monitoring of environment are some of the applications of semantic segmentation. The contributions of the proposed framework are as follows:

- An efficient convolution blocks based encoder is used to extract and discern the boundaries of eddies.
- Attention mechanism is proposed to enhance the feature representation.

The rest of the paper is organised as follows: The details of SSH datasets are covered in Section II. The architecture of the proposed framework is illustrated in Section III. Section IV discusses the experimental results of the proposed framework. Conclusion is stated in section V.

II. Datasets used in the proposed framework

Two datasets, such as the Southern Atlantic Ocean region [20] and the South China Sea [23] are used to evaluate, the EddyNet, DeepLabV3+ with xception backbone (Deep framework), and the proposed framework. Python based eddy

978-1-6654-5136-9/22/$31.00

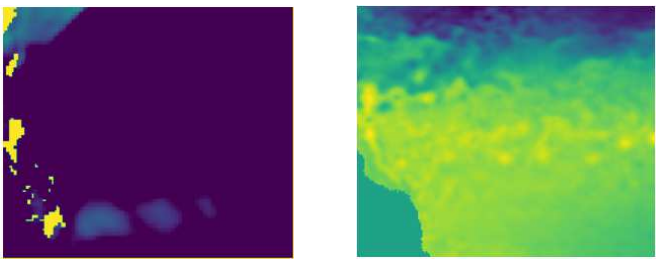
TABLE I: The evolution of existing ocean eddy detection systems and their hierarchy

Literature	Proposition	Remarks
Okubo, *et al*. [4] Weiss, *et al*. [5]	Physical parameters based approach	The physical parameter derivation procedure creates noise and increases the eddy false detection rate making it harder to identify the ideal threshold [6], [7]
Chaigneau, *et al*. [7] Nencioli, *et al*. [8]	Vortex curve and geometric velocity based approach Streampath based approach	These approaches necessitate artificial parameter tweaking and their application range is relatively limited. [23]
Chelton, *et al*. [10]	Threshold independent SSH based approach	It detects eddies with more than one local extreme relying on extra thresholds such as area and horizontal scale [11]
Yi, *et al*. [11]	Hybrid eddy detection approach	It requires a sensitivity test to determine the parameters in the eddy selection phase
Ashkezari, *et al*. [12]	Machine learning (SVM) based approach	It requires pre-manual eddy labeling which necessitates specialist expertise which is a tiresome and time-consuming operation
Lguensat, *et al*. [20] Sun, *et al*. [23]	Deep learning based approach	Eddy detection accuracy and intersection of union score is low.
Proposed work	Deep learning based approach with attention mechanism	Eddy detection accuracy and intersection of union score is better than existing deep learning models.

tracker (PET), a popular method, is used to detect eddies and create ground-truth labels [24]. A pixel-level annotation is performed on the detected eddies to prepare the database required for training the deep learning models. Pixels are labeled as 0 for non-eddy, 1 for anti-cyclonic eddy, and 2 for cyclonic eddy.

1) South china sea dataset: Training and validation data are separated in the china sea data set. The training set contains 4750 SSH images from the first 13 years (2003-2016), and the last two years (2017-2018) data is used to prepare validation set which contains 730 validating SSH images. Each image is a single-channel image with a resolution of 302×184 pixels.

2) Southern Atlantic Ocean region dataset: Fourteen years of data from the Atlantic Ocean (1998-2011) has been divided into 5100 training images, with the latest year's data (365 images) serving as testing. Each image is a single-channel image with a resolution of 240×280 pixels.



(a) (b)

Fig. 1: Representation of sea surface height map of (a) South China Sea data. (b) Southern Atlantic Ocean data.

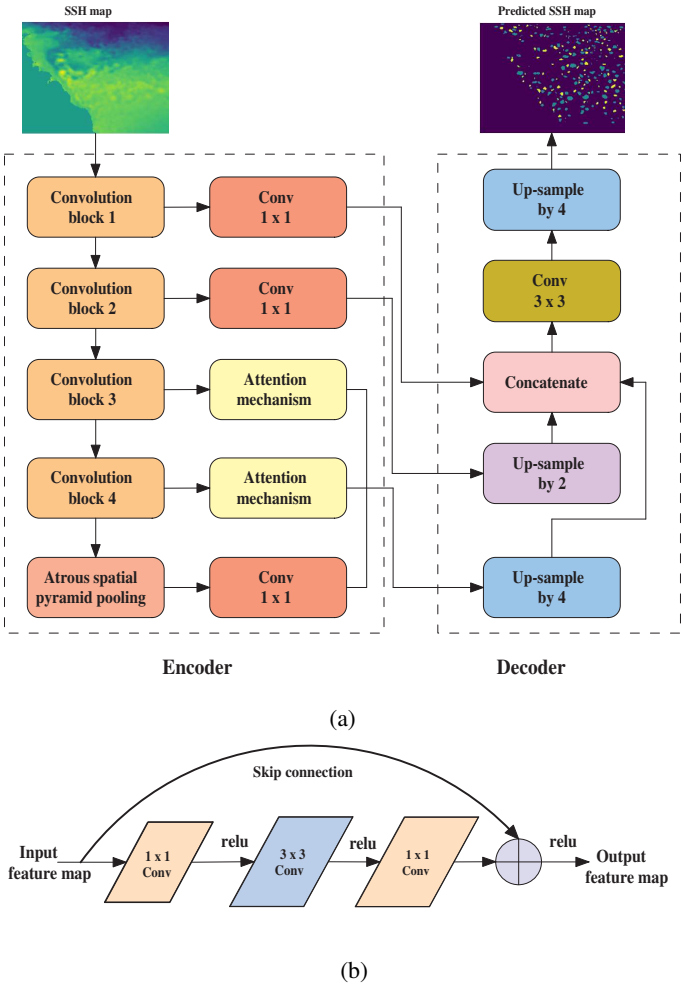


Fig. 2: (a) Block diagram of attention mechanism based ocean eddy detection. (b) Residual structure of convolution blocks.

III. Deep learning technique with attention mechanism

To capture meaningful and accurate border context and multi-scale data, a deep learning model with convolution blocks as a base layer, atrous spatial pyramidal pooling (ASSP), and mechanism based on attention is proposed. As demonstrated in Fig. 2, the two primary components of the proposed framework are an encoder and a decoder.

A. Encoder

The encoder module extracts the features from the input data using multiple convolution blocks. The initial two convolution blocks halve the amount of the input data, while the last two convolution blocks maintain the same input data size. Each convolution block has many residual structures, a 1×1 convolution block of two distinct channels and a 3×3 convolution block. Skip connection is added to each residual structure to deal with the gradient explosion caused by the increase in the depth of the network layers, as shown in Fig. 2(b).

1) Atrous spatial pyramidal pooling (ASSP): Deeplab proposed the ASPP module [25]. It has a parallel atrous convolution structure with multiple sampling rates. ASPP accurately and effectively recognizes features of any scale since the receptive fields of the atrous convolution with different sampling rates are different. A higher sampling rate degenerates 3×3 atrous convolutions in ASPP into 1×1 convolution which unfit it to capture long-range information. Deeplabv3 upgraded the ASPP module, which now comprises a 1×1 convolution and three 3×3 convolutions with sampling rates of 6, 12, and 18 correspondingly, to overcome this problem while maintaining a greater field of vision [26]. The feature extraction module (encoder) executes numerous parallel atrous convolutions as shown in Fig. 3 with variable rates to investigate constitutional features at multiple scales.

2) Attention mechanism: Natural language processing and computer vision are two disciplines where attention is commonly used [27], [28], [29], [30], [31]. The attention mechanism is a matrix multiplication operation that can determine each pixel's reliance relationship in an image and assign a higher weightage to those pixels that have a strong dependence [32]. The attention mechanism constructs a spatial attention matrix from the convolution block's feature mappings to display the pixel relationship in the spatial domain. The final representations reflecting long-range contexts are obtained by multiplying the spatial attention matrix and the original feature map, then conducting an element-wise sum operation on the multiplied resultant matrix. A similar process is applied along the channel and the attention matrix is constructed by multiplying the channel attention matrix and the original feature map, followed by the addition of each element in the multiplied resultant matrix.

As demonstrated in Fig. 4, F represents feature map, that is, $F \in \mathcal{R}^N$ where $N \equiv C \times H \times W$ of width W, height H and with number of channels C. F is fed into a convolution layer to generate two new feature maps, that is, spatial feature map

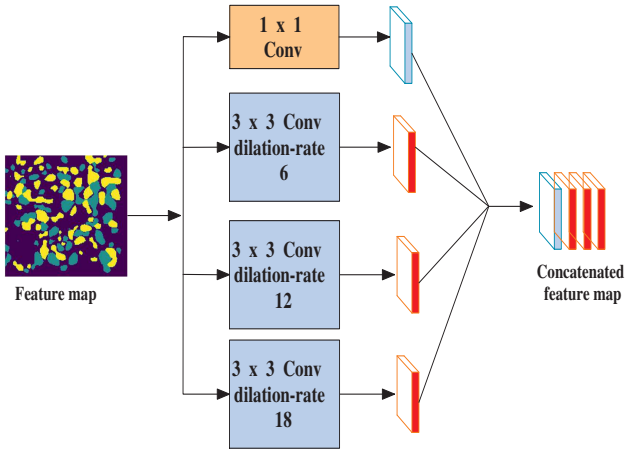


Fig. 3: Block diagram of atrous spatial pyramidal pooling.

$G \in \mathcal{R}^M$, where $M \equiv C \times (H \times W)$ and channel feature map $H \in \mathcal{R}^Q$, where $Q \equiv N \times (H \times W)$, N represents the number of classes. The generated feature map G is reshaped and fed as one of the inputs to multiplication block, where as the feature map H is reshaped and transposed to $\mathcal{R}^K$, with $K \equiv H \times W$ denoting the number of pixels is fed as another input. The multiplication of aforementioned two feature maps yields attention matrix A. The specific operation is as follows:

$$A_{m,n} = \frac{e^{(G_n \times H_m)}}{\sum_{p=1}^{N} e^{(G_n \times H_m)}} \tag{1}$$

where $A_{m,n}$ represents the impact of the n^{th} position on the m^{th}position. Attention matrix A is multiplied with channel matrix to generate intermediate feature map I which is then concatenated with original feature map F to obtain the final feature map F'.

$$I = A \times H \tag{2}$$

$$F' = F + I \tag{3}$$

B. Decoder

The encoded features are up-sampled by four times in the decoder module, and the low-level features from the convolution blocks with the similar spatial resolution are then concatenated. The softmax activation function is used in the final layer of the proposed work to predict the output class probability across the three-channel output layer.

C. Loss Metric

In deep learning, the categorical cross-entropy cost function is widely used to train multi-class classification tasks, while overlap-based metrics are preferred for segmentation problems. In segmentation problems, the dice-coefficient is a well-known and commonly utilized cost function. The expression for dice coefficient and dice-loss is given as follows:

$$Dice - coefficient(P, Q) = \frac{2 \times (P \cap Q)}{|P| + |Q|} \tag{4}$$

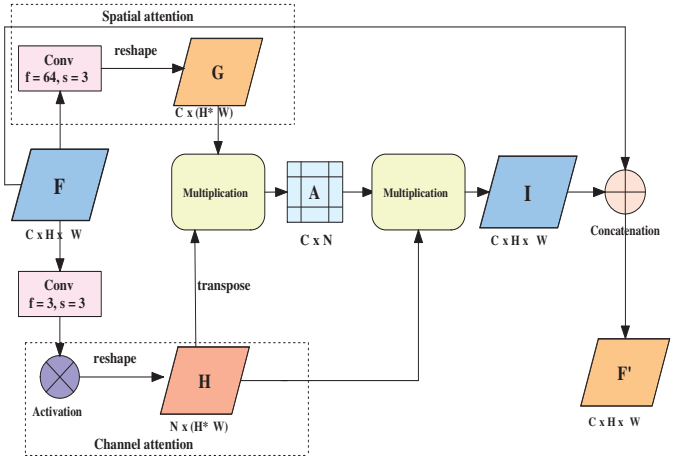


Fig. 4: The process of obtaining an attention matrix using an attention mechanism.

$$Diece - Loss(P, Q) = 1 - \frac{2 \times (P \cap Q)}{|P| + |Q|} \tag{5}$$

where, P: Predicted output image, Q: Ground truth image

IV. Results and Discussions

This section discusses the overall performance and displays the findings gained from the suggested framework.

A. Experimental Results

EddyNet [20], Deep framework [23] and the proposed methods are implemented on the python platform (version 3.8.8) with TensorFlow as the backend and Keras as the frontend on an Intel Xeon E5-2630 v4 (10 core, 2.2 GHz, 32 GB RAM) processor with NVIDIA Quadro M400 8 GB GPU. The following parameters were used for the comparison of the proposed method with [20], [23] in terms of quantitative evaluation: accuracy, sensitivity, and mean intersection of union score (MIOU). Table II specifies the quantitative analysis of three deep learning based frameworks. Table II shows that the proposed framework outperforms the compared strategies on all of the previously specified parameters.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

where, TP: True Positive, TN: True Negative

$$IOU(P, Q) = \frac{2 \times (P \cap Q)}{P \cup Q} \tag{8}$$

where, P: Predicted image, Q: Ground truth image

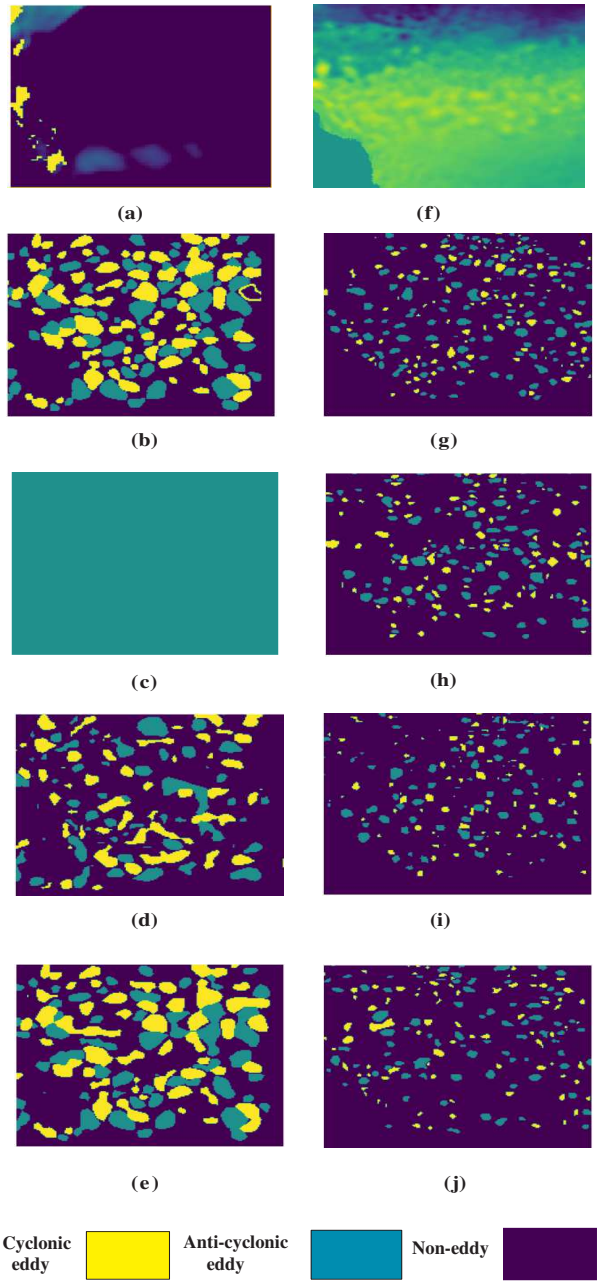


Fig. 5: Representation of sea surface height map of (a) South China Sea. (b) Ground truth. (c) EddyNet model prediction (d) Deep framework prediction (e) Proposed work prediction. (f) Southern Atlantic Ocean region (g) Ground truth. (h) EddyNet model prediction (i) Deep framework prediction (j) Proposed work prediction.

B. Discussions

The proposed method is compared with the existing techniques such as EddyNet and deep framework to thoroughly validate our model. The findings of the aforementioned methods, as well as the proposed method, are presented in Table II with an average intersection of union score of 82.00 percent

TABLE II: Performance evaluation of existing techniques and proposed work

Dataset	Method	Accuracy (%)	Sensitivity (%)	Mean IOU (%)
Southern Atlantic Ocean [20]	Lguensat, *et al*. [20]	88.60	88.56	81.00
	Sun, *et al*. [23]	89.56	89.47	80.21
	Proposed work	**90.85**	**90.85**	**82.00**
South China Sea [23]	Lguensat, *et al*. [20]	17.35	17.36	9.51
	Sun, *et al*. [23]	60.37	60.05	42.00
	Proposed work	**63.25**	**64**	**46.11**

on the Atlantic Ocean dataset and 46.11 percent on the South China Sea dataset. The suggested method performs better than the EddyNet and Deep framework in detecting eddies. It also outperforms both methods in terms of detection accuracy.

The proposed method is trained with two different three-class datasets that had a batch size of 4. The training constituted of 300 epochs. The model used 5100 and 4750 images as datasets for this complete comparison. Day 10 detection results of EddyNet, Deep framework and the proposed method are visualized as shown in Fig. 5, to better comprehend our findings. Fig. 5(a) to 5(e) represents one day (day 10) sea surface height (SSH) data of South China Sea, ground truth (eddy labeled) SSH image of South China Sea, EddyNet predictions, Deep framework predictions and proposed method predictions on South China Sea data. As it shown in Fig. 5(e) proposed method detected more eddies than the other two methods, where as EddyNet is inaccurate in detecting eddies in the South China Sea dataset. Fig. 5(f) to 5(j) represents the South Atlantic Ocean region one day (day 10) SSH data, ground truth (eddy labeled) SSH map, EddyNet predictions, Deep framework predictions and the proposed method predictions. As it shown in Fig. 5(j) the proposed method detected eddies more accurately than the other two methods. The suggested technique detects eddies with better accuracy in the South Atlantic Ocean and the South China Sea dataset due to the attention mechanism incorporated in the proposed work. The proposed method is a substantial improvement over current methods. The performance of the proposed work on the South Atlantic Ocean region (dataset 1) is better than the south china sea (dataset 2) due to the over-fitting nature of dataset 2.

V. Conclusion

The proposed work's purpose is to solve the challenge of autonomous ocean eddies recognition by employing semantic segmentation approaches based on the most efficient feature representation backbone. The attention module is appended to the convolution blocks-based base layer to improve the boundary accuracy of eddies detection. The quantitative and qualitative result analysis shows that the proposed framework performs better than the previous models.

References

[1] C. Dong, J. C. Mc. Williams, Y. Liu, "Global heat and salt transports by eddy movement," *Nature Communications*, vol. 5, 2014.

[2] Z. Zhang, W. Wang, and B. Qiu, "Oceanic mass transport by mesoscale eddies," *Science*, vol. 345, no. 6194, pp. 322-324, 2014.

[3] Y. Liu, L. Yu, and G. Chen, "Characterization of Sea Surface Temperature and Air Sea Heat Flux Anomalies Associated With Mesoscale Eddies in the South China Sea," *Journal of Geophysical Research: Oceans*, vol. 125, no. 4, 2020.

[4] A. Okubo, "Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences," *Deep Sea Research and Oceanographic Abstracts*, pp. 445–454, doi:10.1016/0011-7471(70)90059-8, 1970.

[5] J. Weiss, "The dynamics of enstrophy transfer in two-dimensional hydrodynamics," *Phys. D: Nonlinear Phenom*, vol. 48, pp. 273–294, 1991.

[6] I. Ari. Sadarjoen, F. H. Post, "Detection,quantification, and tracking of vortices using streamline geometry," *Computers and Graphics*, pp. 333–341, doi: 10.1016/S0097-8493(00)00029-7, 2000.

[7] A. Chaigneau, A. Gizolme, and C. Grados, "Mesoscale eddies off Peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns," *Progress in Oceanography*, vol. 79, pp. 106–119, 2008.

[8] F. Nencioli, Dong. Changming, T. Dickey, "A vector geometry-based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in the southern California bight," *Journal of Atmospheric and Oceanic Technology*, pp. 564–579, doi: 10.1175/2009JTECHO725.1., 2010.

[9] A. M. Doglioli, B. Blanke, S. Speich, "Tracking coherent structures in a regional ocean model with wavelet analysis: Application to Cape Basin eddies," *Journal of Geophysical Research: Oceans*, vol. 112, 2007.

[10] D. B. Chelton, M. G. Schlax, R. M. Samelson, "Global observations of nonlinear mesoscale eddies," *Progress in Oceanography*, pp. 167–216, doi: 10.1016/j.pocean.2011.01.002, 2011.

[11] J. Yi, Y. Du, Z. He, and C. Zhou, "Enhancing the accuracy of automatic eddy detection and the capability of recognizing the multi-core structures from maps of sea level anomaly," *Ocean Sci.*, pp. 39–48, https://doi.org/10.5194/os-10-39-2014, 2014.

[12] Mohammad. D. Ashkezari, N. Christopher Hill, N. Christopher Follett, Gael. Forget, J. Michael Follows, "Oceanic eddy detection and lifetime forecast using machine learning methods," *Geophysical Research Letters*, doi:10.1002/2016GL071269, 2016.

[13] 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.

[14] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European Conference on Computer Vision*, Springer, pp. 746–760, 2012.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 3354–3361, 2012.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, pp. 91–99, 2015.

[17] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, pp. 21–37, 2016.

[20] 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," IGARSS *IEEE International Geoscience and Remote Sensing Symposium*, pp. 1764-1767, doi: 10.1109/IGARSS.2018.8518411, 2018.

[21] X. Lu, S. Guo, M. Zhang, J. Dong, X. Chen and X. Sun, "Mesoscale Ocean Eddy Detection Using High-Resolution Network," *International Conference on Awareness Science and Technology (iCAST)*, pp. 1-6, doi: 10.1109/iCAST51195.2020.9319490, 2020.

[22] X. Bai, C. Wang and C. Li, "A Streampath-Based RCNN Approach to Ocean Eddy Detection," in *IEEE Access*, vol. 7, pp. 106336-106345, doi: 10.1109/ACCESS.2019.2931781, 2019.

[23] X. Sun, M. Zhang, J. Dong, R. Lguensat, Y. Yang and X. Lu, "A Deep Framework for Eddy Detection and Tracking From Satellite Sea Surface Height Data," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, pp. 7224-7234, doi: 10.1109/TGRS.2020.3032523, 2021.

[24] E. Mason, A. Pascual, and J. C. McWilliams, "A new sea surface height–based code for oceanic mesoscale eddy tracking," in *Ocean.Technol*, vol. 31, pp. 1181–1188, May 2014

[25] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs,", *arXiv preprint arXiv:1412.7062*, 2016.

[26] L. C. Chen, G. Papandreou, F. Schroff, H. Adam, "Rethinking atrous convolution for semantic image segmentatio,", *arXiv preprint arXiv:1706.05587*, 2017.

[27] Ashish. Vaswani, Noam. Shazeer, Niki. Parmar, Jakob. Uszkoreit, Lion. Jones, Aidan. N. Gomez, Lukasz. Kaiser, and Illia. Polosukhin, "Attention is all you need," *In Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[28] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, "Non-local neural networks," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, 2018.

[29] Han. Hu, Gu. Jiayuan, Zheng. Zhang, Dai. Jifeng, and Wei. Yichen, "Relation networks for object detection," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3588–3597, 2018.

[30] Hengshuang. Zhao, Yi. Zhang, Shu. Liu, Jianping. Shi, Chen. Change Loy, Dahua. Lin, and Jiaya. Jia, Psanet, "Point-wise spatial attention network for scene parsing," *In Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 267–283, 2018.

[31] Fu, Jun, Jing. Liu, Haijie. Tian, Yong. Li, Yongjun. Bao, Zhiwei. Fang, and Hanqing. Lu. "Dual attention network for scene segmentation," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3146-3154, 2019.

[32] J. Hu, L. Shen, G. Sun, "Squeeze-and-Excitation Networks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.