Acta Oceanol. Sin., 2023, Vol. 42, No. 10, P. 67–74
https://doi.org/10.1007/s13131-023-2206-6
http://www.aosocean.com
E-mail: ocean2@hyxb.org.cn

Strip segmentation of oceanic internal waves in SAR images based on TransUNet

Kaituo Qi[1], Hongsheng Zhang[1]*, Jiaojiao Lu[1], Yinggang Zheng[2]*, Zhouhao Zhang[1]

[1] College of Ocean Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

[2] Translational Research Institute of Brain and Brain-Like Intelligence, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai 200434, China

Received 15 January 2023; accepted 13 April 2023

Abstract

The development of oceanic remote sensing artificial intelligence has made possible to obtain valuable information from amounts of massive data. Oceanic internal waves play a crucial role in oceanic activity. To obtain oceanic internal wave stripes from synthetic aperture radar (SAR) images, a stripe segmentation algorithm is proposed based on the TransUNet framework, which is a combination of U-Net and Transformer, which is also optimized. Through adjusting the number of Transformer layer, multi-layer perceptron (MLP) channel, and Dropout parameters, the influence of over-fitting on accuracy is significantly weakened, which is more conducive to segmenting lightweight oceanic internal waves. The results show that the optimized algorithm can accurately segment oceanic internal wave stripes. Moreover, the optimized algorithm can be trained on a microcomputer, thus reducing the research threshold. The proposed algorithm can also change the complexity of the model to adapt it to different date scales. Therefore, TransUNet has immense potential for segmenting oceanic internal waves.

Key words: oceanic internal waves, deep learning, stripe segmentation, synthetic aperture radar, TransUNet

Citation: Qi Kaituo, Zhang Hongsheng, Lu Jiaojiao, Zheng Yinggang, Zhang Zhouhao. 2023. Strip segmentation of oceanic internal waves in SAR images based on TransUNet. Acta Oceanologica Sinica, 42(10): 67–74, doi: 10.1007/s13131-023-2206-6

1 Introduction

Oceanic internal waves are a wave phenomenon occurring in a considerable part of ocean, which significantly impact the dynamic processes of the ocean and the safety of oceanic engineering structures (Lavrova et al., 2014). Therefore, it is necessary to accurately determine the location of oceanic internal waves.

The microwave-band synthetic aperture radar (SAR) could observe underwater oceanic internal waves from tens to hundreds of meters. Owing to the change of ocean surface flow field caused by the propagation of internal waves modulates the distribution of ocean surface micro-scale waves, the ocean surface roughness changed, which appears as bright and dark stripes in SAR images (Alpers, 1985). Oceanic internal waves often generally propagate in the form of wave packets and internal solitary waves. In the last 40 years, SAR has been widely used in earth remote sensing. The amount of remote sensing data is increasing, along with the demand for data processing. Therefore, to determine the locations of oceanic internal waves accurately, it is necessary to develop an automated segmentation methodology for oceanic internal waves in SAR images.

Several scholars have studied the detection of oceanic internal waves. For example, Wang et al. (2019) extracted internal waves from an unmanned aerial vehicle (UAV) image based on a principal component analysis network, and detected them using a support vector machine classifier. The method proposed by Wang et al. can only mark the approximate regions of internal waves, but cannot easily identify the precise locations quantitatively. Similarly, Bao et al. (2020) used faster regions with a convolutional neural network to detect oceanic internal waves in SAR images, and the results showed that the approximate region of the internal waves could be effectively identified. Zhang et al. (2020) recognized internal waves based on the K-means clustering algorithm, which has the advantages of application convenience and fast convergence, however, has the disadvantages of high requirements for data sets and limited scope of application. Zheng et al. (2021) proposed an integrated algorithm for detecting and recognizing of oceanic internal waves. Li et al. (2020) employed an optimized U-Net to obtain information on internal waves from Himawari-8 satellite images. They aggregated the three-layer output of the convolutional layer in the U-Net encoder into one layer and thus obtained good results for internal wave extraction. Zheng et al. (2022) proposed a segmentation algorithm for oceanic internal wave stripes in SAR images based on SegNet. However, SegNet cannot fully utilize the relationship of global context, which leads to the segmentation results not enough fine.

In 2012, the deep learning model (AlexNet) adopted by Krizhevsky et al. (2017) won the Image Net image recognition competition, which once again attracted significant attention from the academic community. Subsequently, deep learning has rapidly developed. For example, Simonyan and Zisserman (2015) proposed deep convolutional neural networks (CNN), Shelhamer et al. (2017) proposed fully convolutional networks (FCN), and based on FCN, Ronneberger et al. (2015) achieved the seg-

Foundation item: The National Natural Science Foundation of China under contract No. 51679132; the Science and Technology Commission of Shanghai Municipality under contract Nos. 21ZR1427000 and 17040501600.
*Corresponding author, E-mail: hszhang@shmtu.edu.cn; ingopro@126.com

mentation of medical images using a U-Net fully convolutional network. The CNN assumes that the elements are independent of each other. However, many elements are connected in reality. Therefore, Zaremba et al. (2014) proposed the Recurrent Neural Networks (RNNs), which is recursive along the sequence direction. However, in RNNs, the processing of data is in series from one level to the next, making it difficult to accelerate and parallelize the training process. Vaswani et al. (2017) proposed a parallel computing Transformer, which is composed of a multi-head attention mechanism and feed-forward neural network, and mainly used in the Sequence-to-Sequence natural language processing. Inspired by the robust performance of Transformer, researchers extended Transformer to the field of computer vision and have achieved great success in this field. Dosovitskiy et al. (2021) proposed a Vision Transformer (ViT) for image classification, which achieved good results when directly applied to image slice sequences. The above mentioned studies elucidate how Transformer is a valuable tool in image detection and segmentation.

According to the advantages of Transformer in segmentation, this study proposes an algorithm for stripe segmentation of oceanic internal waves based on TransUNet. The proposed algorithm has the advantages of both high-resolution spatial information of U-Net and feature learning from global contexts by Transformer. The algorithm can identify specific internal wave stripes and has a wide range of applications. This study optimized the Transformer layer, multi-layer perceptron (MLP) channel, and dropout parameter of the TransUNet model to render it more suitable for lightweight oceanic internal wave segmentation. The second section introduces the datasets and the model proposed in this study; the third and fourth sections contain the experimental results, analysis, discussion, respectively, and the conclusion is drawn at the end.

2 Data and methods

2.1 *Data*

2.1.1 *Data sources*

In this study, the Environmental Satellite (Envisat), the first European Remote Sensing Satellite (ERS-1), the second European Remote Sensing Satellite (ERS-2), and the Advanced Land Observing Satellite (ALOS) were used as sources of SAR image data. The red polygon in Fig. 1 represents the coverage area of the downloaded image data.

2.1.2 *Data set and annotation*

The acquisition steps of the oceanic internal wave SAR images and corresponding label datasets were as follows.

(1) Download 5 817 SAR images, and select the data with oceanic internal wave stripes.

(2) The single-look complex (SLC) image has more speckle noise, it is thus necessary to perform multi-look processing and filtering on the original data. Multi-look processing averages the resolution of image along the distance direction and azimuth direction to suppress the speckle noise of the SAR image. Gamma Map filtering with a window of 5 is used to reduce speckle noise in SAR images. Subsequently, the geocoding is used to convert the SAR image slant range projection into geographic coordinate projection. The scattering intensity information is normalized by radiometric calibration. The results of the pre-processing are shown in Fig. 2. The imaging time is March 16, 1998. The center point coordinates are near 6.796 3°N, 97.327 3°E. It is a single look complex image with vertical polarization mode.

(3) Convert the images into TIFF format to obtain the intensity value and geographic information of each pixel of the images.

(4) As shown in the red rectangular box in Fig. 2, the image data of all TIFF formats are randomly cut to increase the amount of data and saved in JPG format.

(5) Label the cropped JPG format data set with LabelMe (Russell et al., 2008) visualization software, and generate label data in PNG format.

(6) Binarize the label data; the grey value of the stripe is selected as 255, and the grey value of the background is selected as 0.

Figure 3 shows a flowchart of the image data processing. Labelme software was used to label each pixel of each image in the dataset with the corresponding category label, and then the pixel sample was converted into a binary image and 703 images were produced. The crop dataset consists of randomly cropped image data.

2.2 *Method and optimization*

Chen et al. (2021) proposed the TransUNet algorithm for semantic segmentation, which is a deep neural network structure

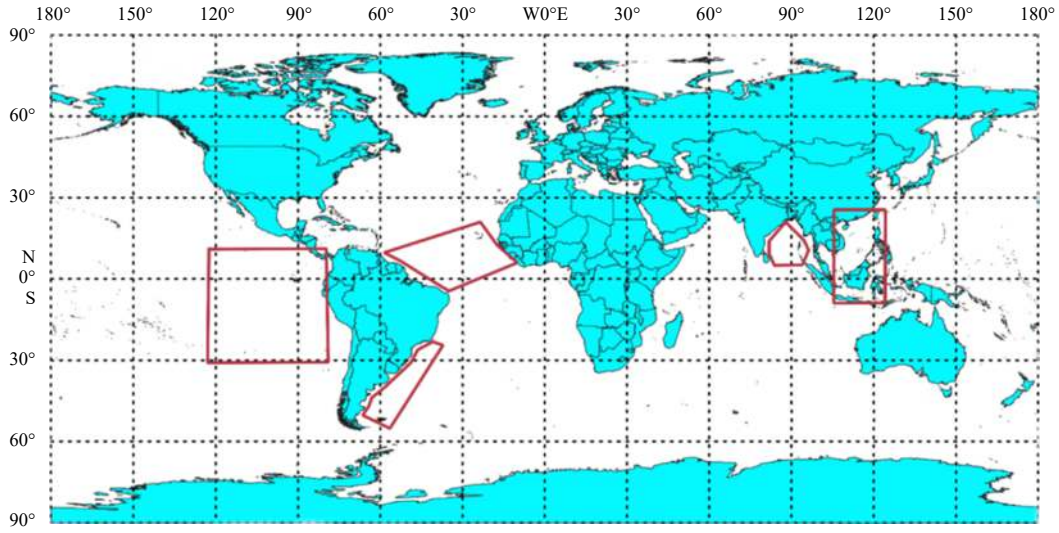

Fig. 1. Synthetic aperture radar (SAR) data area information.

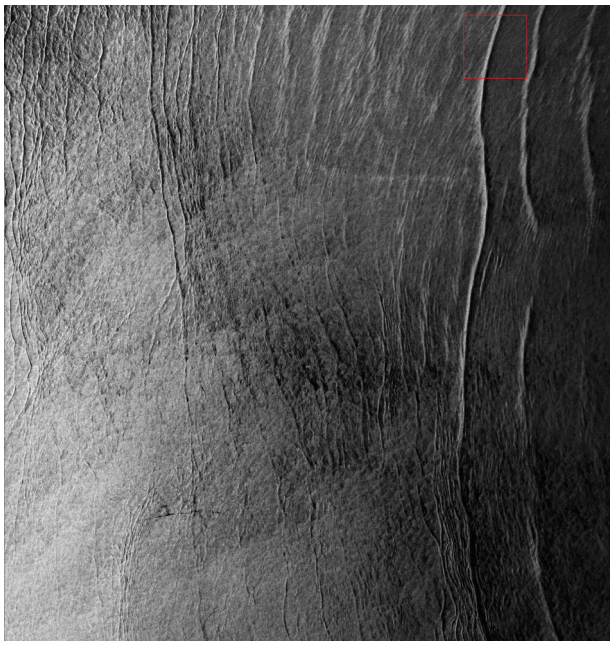

Fig. 2. Oceanic internal wave from synthetic aperture radar image of ERS-2 satellite.

that combines U-Net and Transformer. TransUNet has both the high-resolution spatial information of CNN and the global context advantage of Transformer coding; thus, it exhibits superior performance in the field of image segmentation (Chen et al., 2021).

TransUNet consists of an encoder and corresponding decoder, as shown in Fig. 4. The TransUNet encoder has a CNN-Transformer hybrid architecture. A CNN was used to encode the image into feature maps. Because the Transformer is an attention mechanism for Sequence-to-Sequence prediction, the features of the images need to be serialized. In this study, the input feature map was divided into a sequence of flattened 2D patches, and each patch size was $P \times P$. The sequence length was $N = (HW)/P^2$ (H and W are the length and width of the input image, respectively). The vectorized x_p (patch) was mapped into a potential D-dimensional embedding space using trainable linear projection. To encode the spatial information, specific location information was added to each patch, as follows:

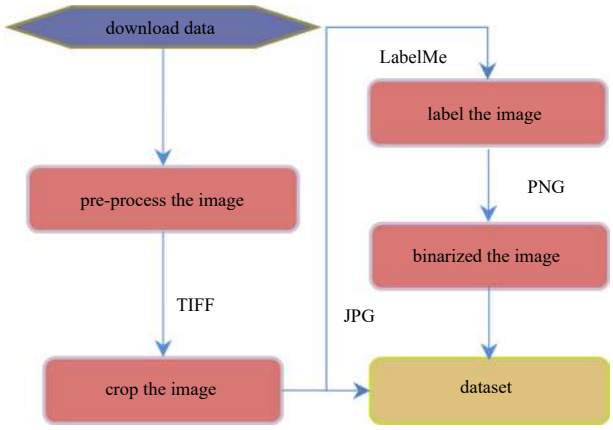

Fig. 3. Steps for collecting the ocean internal wave synthetic aperture radar data sets.

$$z_0 = \left[x_p^1 \boldsymbol{E}; x_p^2 \boldsymbol{E}; \ldots; x_p^N \boldsymbol{E}\right] + \boldsymbol{E}_{pos}, \quad (1)$$

where z_0 is the initial input to the Transformer; $\boldsymbol{E}$ represents the embedded projection; and $\boldsymbol{E}_{pos}$ represents the embedded position.

There are l-layer Transformers in the encoder, each of which

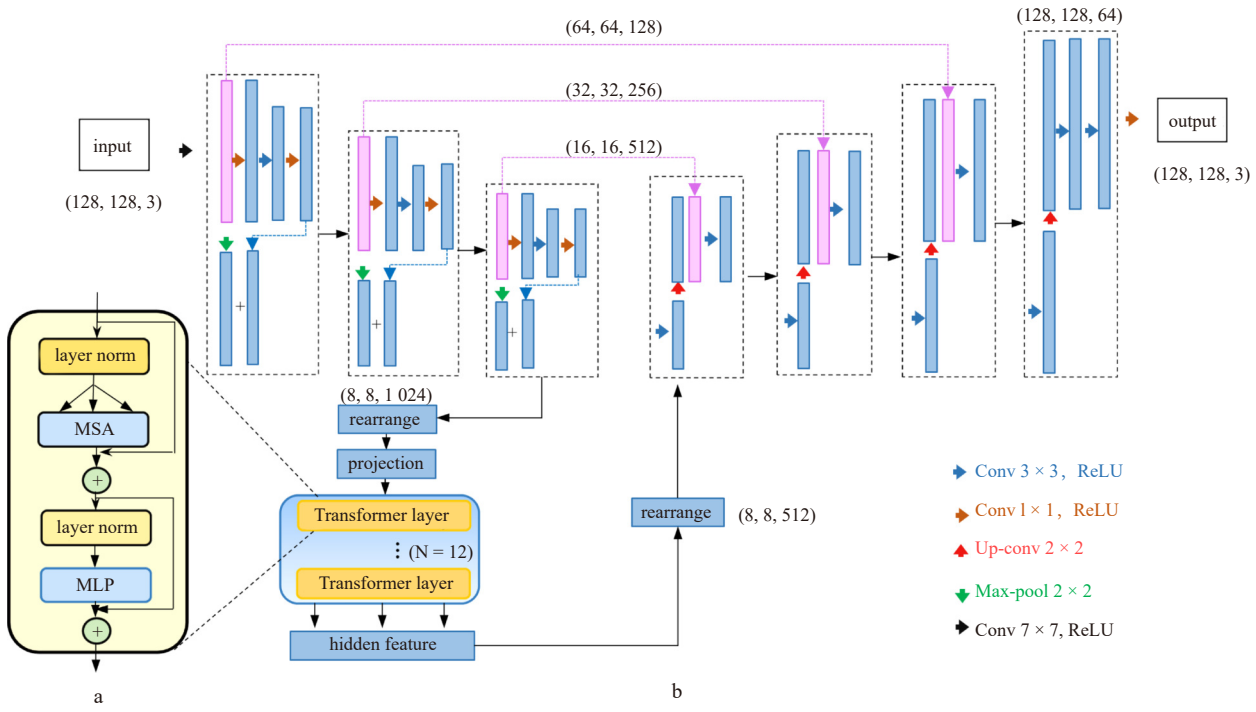

Fig. 4. TransUNet framework structure, Transformer structure (a) and Cross-network structure (b). The parameters in the bracket are image size, image size and dimension, respectively. MSA: multihead self-attention; MLP: multi-layer perceptron; Conv: convolution kernel; ReLU: activation function.

consists of a multihead self-attention (MSA) and a MLP block. Thus, the output of the l-th layer can be expressed as:

$$\boldsymbol{Z}'_l = \mathrm{MSA}\left(\boldsymbol{L}_\mathrm{N}\left(\boldsymbol{Z}_{l-1}\right)\right) + \boldsymbol{Z}_{l-1}, \tag{2}$$

$$\boldsymbol{Z}_l = \mathrm{MLP}\left(\boldsymbol{L}_\mathrm{N}\left(\boldsymbol{Z}'_l\right)\right) + \boldsymbol{Z}'_l, \tag{3}$$

where $\boldsymbol{L}_\mathrm{N}$ is the layer normalization operator and $\boldsymbol{Z}_l$ is the encoded image feature. The structure of the Transformer layer is illustrated in Fig. 4a.

The output result of the l-th layer Transformer is the encoding feature of a two-dimensional patch sequence in D-dimensional space. To restore spatial order, the size of the encoded features should be reshaped to the format before serialization. The decoder is shown in Fig. 4b. The activation function in the decoder was ReLU. As shown in Fig. 4b, the encoded features were up-sampled; that is, 2 × 2 convolution kernels were used to reduce the number of feature channels by half; the encoded features were then aggregated with the features of the corresponding resolution on the skip-connections, and the output results were subjected to two 3 × 3 convolution operations. The above process was performed three times. Finally, the step of aggregating features with the corresponding resolution on the skip connections was omitted, and a 1 × 1 convolutional layer was used for image classification.

The dice similarity coefficient (DSC) loss function of TransUNet is a measure of set similarity that is typically used to evaluate the similarity between two samples. The DSC loss function is defined as follows:

$$\mathrm{DSC}_\mathrm{loss} = 1 - \frac{2F_\mathrm{P}}{F_\mathrm{P} + 2T_\mathrm{P} + F_\mathrm{N}}, \tag{4}$$

where F_P is the false positive; T_P is the true positive; and F_N is the false negative.

In this study, when the original TransUNet was used to segment oceanic internal wave stripes, it was found that the segmentation effect was slightly worse than that of CNN. Dosovitskiy et al. (2021) experiment showed that the Transformer's accuracy was better than that of the CNN only when the data reached a mega-scale. To make TransUNet more suitable for the segmentation of lightweight oceanic internal wave data and solve the overfitting problem caused by the high complexity of the model, this study analyzed the two modules of the Transformer layer and MLP channel, which mainly affect TransUNet complexity.

The effects of the Transformer layer and MLP channel on DSC accuracy are shown in Table 1. For lightweight oceanic internal wave data, the accuracy increased primarily and then decreased with an increase in the Transformer layer. When there were 2 layers, the average accuracy of the 5 MLP channels was the highest. The model displayed good segmentation performance and was advantageous in terms of computational efficiency. When there were more than 8 layers, the segmentation accuracy decreased rapidly. When the number of MLP channels was 512, the average accuracy was the highest. To determine the influence of the MLP channel and Transformer layer on the training process more clearly, performance analysis charts with 2 layers and 512 channels were drawn.

Table 1. The effects of Transformer layer and multi-layer perceptron (MLP) channel on dice similarity coefficient accuracy (%)

	Transformer layer				
	1	2	4	8	12
MLP channel 128	84.13	84.18	82.35	82.64	75.22
MLP channel 256	84.15	83.75	83.49	82.50	72.92
MLP channel 512	83.73	84.57	84.19	82.61	73.73
MLP channel 768	84.09	83.95	80.37	82.49	77.2
MLP channel 1 024	84.00	84.89	80.54	82.66	76.6

Figure 5 shows the effect of the Transformer layer on the loss rate. When the Epoch is 100, the sequence of convergence speed is represented by the black line for one layer, the red line for two layers, the blue line for four layers, the green line for eight layers, and the purple line for 12 layers. At this time, the accuracy decreased as the number of Transformer layers increased. When the Epoch was 200, the accuracy of the two layers was the best. From the perspective of training duration, the sequence of the time-consuming order is represented by the purple line for 12 layers, the black line for one layer, the red line for two layers, the green line for eight layers, and the blue line for four layers. As mentioned above, when there are 2 Transformer layers, the efficiency and accuracy of TransUNet training is preferable.

The effect of the MLP channel on the loss rate is illustrated in Fig. 6. The smaller the MLP channel, the more unstable is the loss rate in the later training period. When the Epoch is 200, the ac-

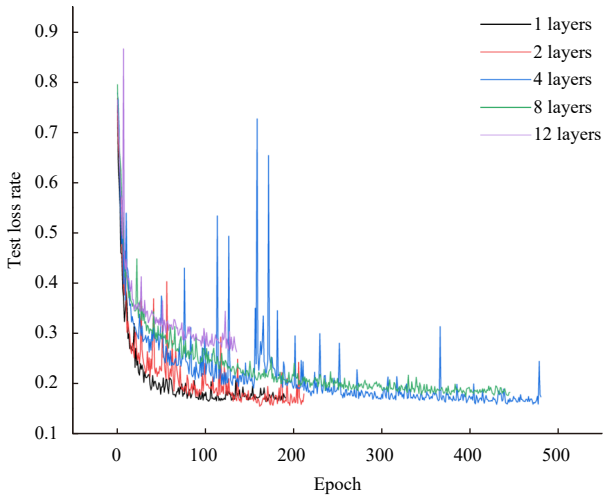

Fig. 5. The effect of Transformer layer on loss rate.

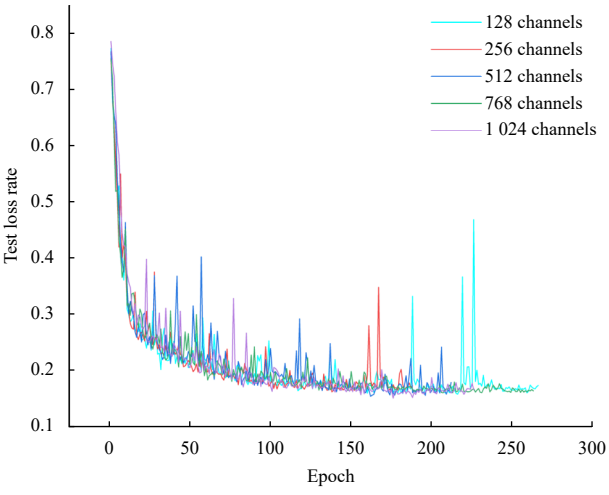

Fig. 6. The effect of multi-layer perceptron (MLP) channel on loss rate.

curacy of all five lines basically achieve optimal value and are close to each other. Therefore, segmented oceanic internal wave stripes were used to analyze the influence of the MLP channel on the accuracy.

The test set segmentation results for the different MLP channels are shown in Fig. 7. When the MLP channel increased from 128 to 768, the contours of the segmentation results were precise and complete for high-contrast images, and the segmentation results were increasingly complete for low-contrast images (such as the 8th plot of Fig. 7e). When the number of channels was 1 024, some segmented stripes were broken. Therefore, the segmented results were better when the number of channels was 768.

A Dropout (Srivastava et al., 2014) was used to solve the over-fitting problem. Dropout is referred to as random inactivation. During forward propagation, the activation values of some neurons are stopped with a certain probability (p), which can significantly reduce over-fitting. This study analyzed the influence of p on the segmentation results.

The effect of Dropout on the loss rate is illustrated in Fig. 8. When p is 1, the fitting effect of TransUNet is poor; with the gradual decrease of p, the fitting effect improves; and when p is between 0.05 and 0.2, the loss rate stabilizes. Therefore, Dropout significantly reduces over-fitting and improves the training accuracy.

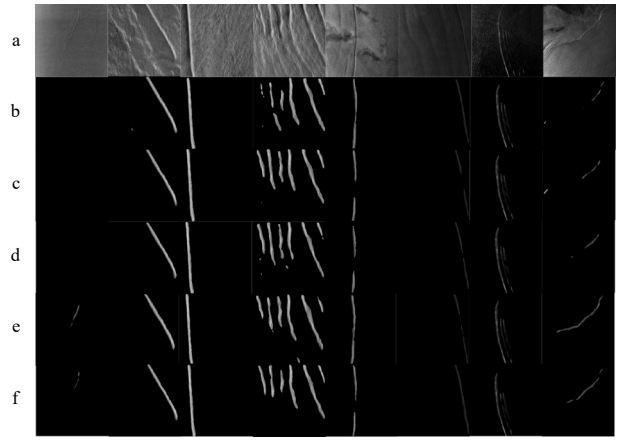

Fig. 7. Test set segmentation results of different multi-layer perceptron (MLP) channels. a is the original image; MLP channels are 128 (b), 256 (c), 512 (d), 768 (e), and 1 024 (f).

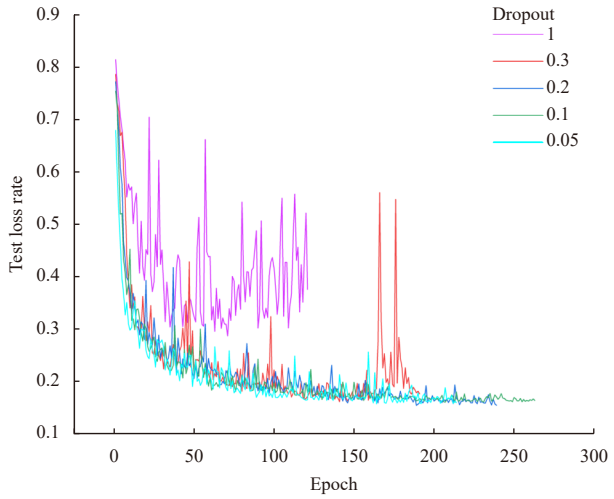

Fig. 8. The effect of Dropout on loss rate.

3 Results and discussion

3.1 *Implementation details*

To train the TransUNet network, the datasets were divided into three parts: training, validation, and testing, as listed in Table 2. As shown in Table 2, when the data distribution ratio was 8:1:1, after the model tests, the DSC loss value of the test set was reduced to the lowest level, and the model effect was better. The original image resolution size was 256 × 256, and it was resized to 128 × 128 in the model.

For the model, based on previous experience (Chen et al., 2021) and the model tests, the SGD optimizer with learning a rate of 0.001, momentum of 0.9, and weight decay 1×10^{-4} was used for training. All tests were conducted using a Precision Tower 7920 with a single NvidiaRTX3060GPU. Windows was used as the operating system. The algorithm was implemented on the Sentinel Application Platform and Python Platforms.

3.2 *Model evaluation*

The model tests were carried out on BatchSize 8, 16 and 32. It was found that more time would be taken and convergence was not unfavourable if the BatchSize was too small, and that the gradient direction of different batches did not change, making it easy to fall into local minimum if the batch size was too large. Therefore, the BatchSize was set to 16. The tests also indicated that the size of the PatchSize has little effect on the accuracy. According to the Dosovitskiy et al. (2021) setting, the PatchSize was set to 16.

To better segment the lightweight oceanic internal wave stripes, TransUNet was optimized to reduce the degree of over-fitting. In the experiment, the Transformer layer was changed to 2, the MLP channel was set to 768, and p of dropout was set to 0.2. Other default parameters were as follows: BatchSize was set to 16, PatchSize was set to 16, and default training Epochs was set to 500. The patience was set to 50 to prevent over-fitting (training was stopped if the loss rate of the test set did not decrease 50 consecutive times).

The model performance analysis of the original TransUNet and optimized TransUNet is shown in Fig. 9. As shown in Fig. 9a, at the 379th Epoch, the original TransUNet achieved the best training model; the DSC loss value of the training set was reduced to 0.267, and the loss value of the test set was reduced to 0.228. As shown in Fig. 9b, at the 189th Epoch, the optimized TransUNet achieved the best training model, the DSC loss value of the training set was reduced to 0.136, and the loss value of the test set was reduced to 0.154. The optimized TransUNet was found to be more suitable for segmenting lightweight oceanic internal wave stripes in terms of both accuracy and efficiency.

3.3 *Visual analysis*

To evaluate the segmentation effect of the optimized TransUNet, the segmentation results of the original TransUNet, optimized TransUNet, and U-Net models are shown in Fig. 10. The DSC segmentation accuracies were 0.770, 0.846, and 0.828, respectively. When the stripes were segmented by the original TransUNet, most of the wave packets were broken and incom-

Table 2. Model training performance at different rates

Rate	Training loss	Test loss
6:2:2	0.179 7	0.319 1
7:1.5:1.5	0.190 2	0.280 6
8:1:1	0.157 5	0.203 9
9:0.5:0.5	0.156 4	0.258 5

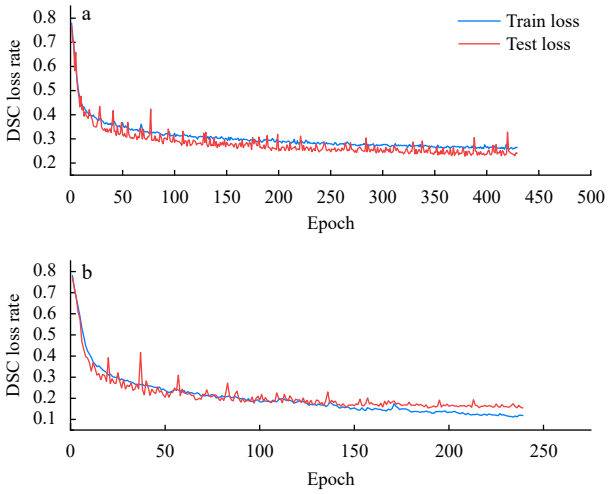

Fig. 9. Model performance analysis of the original TransUNet (a) and the optimized TransUNet (b). DSC: dice similarity coefficient.

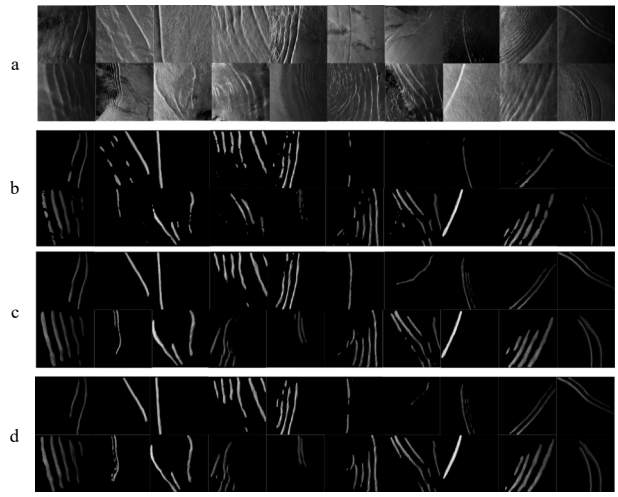

Fig. 10. Qualitative comparison of different approaches by visualization. The original image (a), different approaches by original TransUNet (b), optimized TransUNet (c) and U-Net (d).

plete, and the internal solitary waves could not even be segmented. When the stripes segmented using U-Net were relatively complete, the wave packets could be effectively segmented. However, images with the internal solitary waves and low-contrast encounter the problem of incomplete segmentation, as shown in the 6th plot of Fig. 10d. With optimized TransUNet, segmentation results were complete and unbroken, whether for wave packets or the internal solitary waves. The tests demonstrate that the optimized TransUNet can more precisely segment lightweight oceanic internal wave stripes and retain detailed position information.

The segmentation results of the entire SAR image with a resolution of 4 903 × 5 151 pixels are shown in Fig. 11. As shown in Fig. 11, for segmenting large-scale internal wave images, the proposed algorithm can only segment the internal waves on the right side of the image. However, the recognition effect of the subtle internal wave stripes on the left side of the image is slightly insufficient. Therefore, the subtle internal waves in Fig. 11 are cropped and segmented in this study. The small-scale SAR image segmentation results are shown in Fig. 12. As shown in Fig. 12, the segmentation results of two small-scale images are good. The segmentation effect is better if the resolution of the input image is small-scale; however, it is probably encountered that the oceanic internal waves cannot be identified if the resolution of the input image is large-scale. Therefore, the proposed algorithm has advantages in recognition of ocean internal waves, the effect of segmentation, however is affected by the resolution of the input image. Because the sample resolution trained in this study is 128 × 128, it is better for the recognition of small-scale image. If a computer with a more robust graphics processing unit is employed to directly train large-scale SAR images, the final segmentation accuracy will be improved and the oceanic internal wave positions will be obtained more conveniently.

The segmentation results of the entire SAR image with a resol-

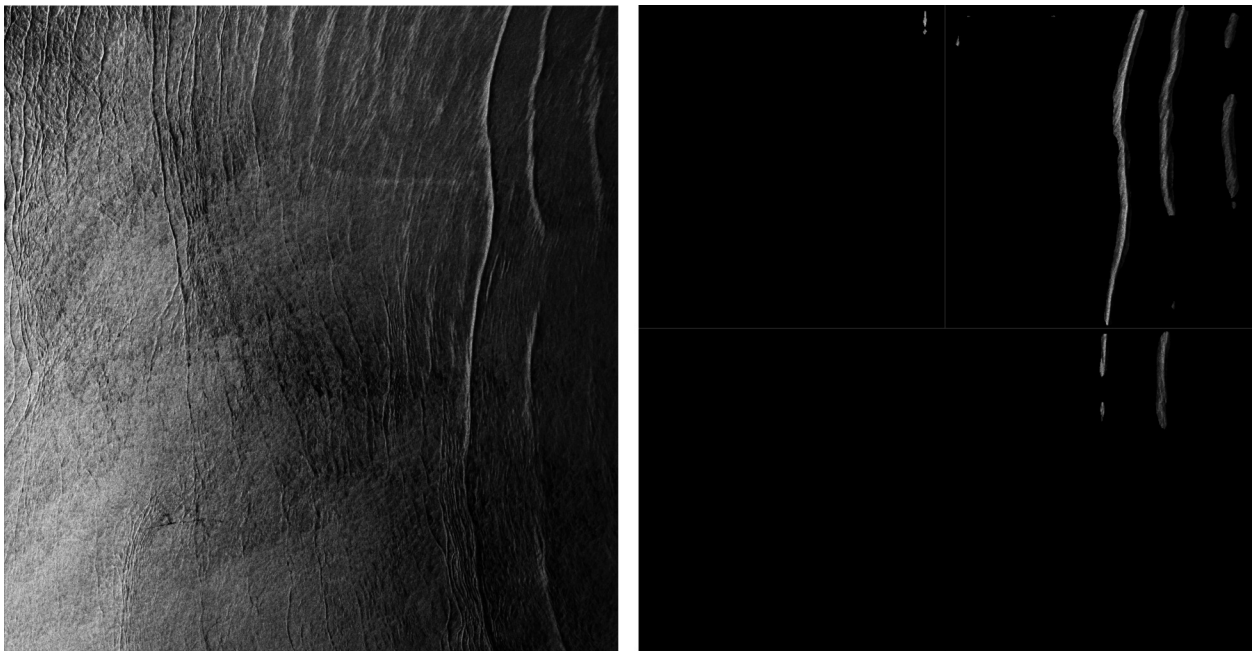

Fig. 11. Results of the entire synthetic aperture radar image segmentation (resolution size is 4 903 × 5 151).

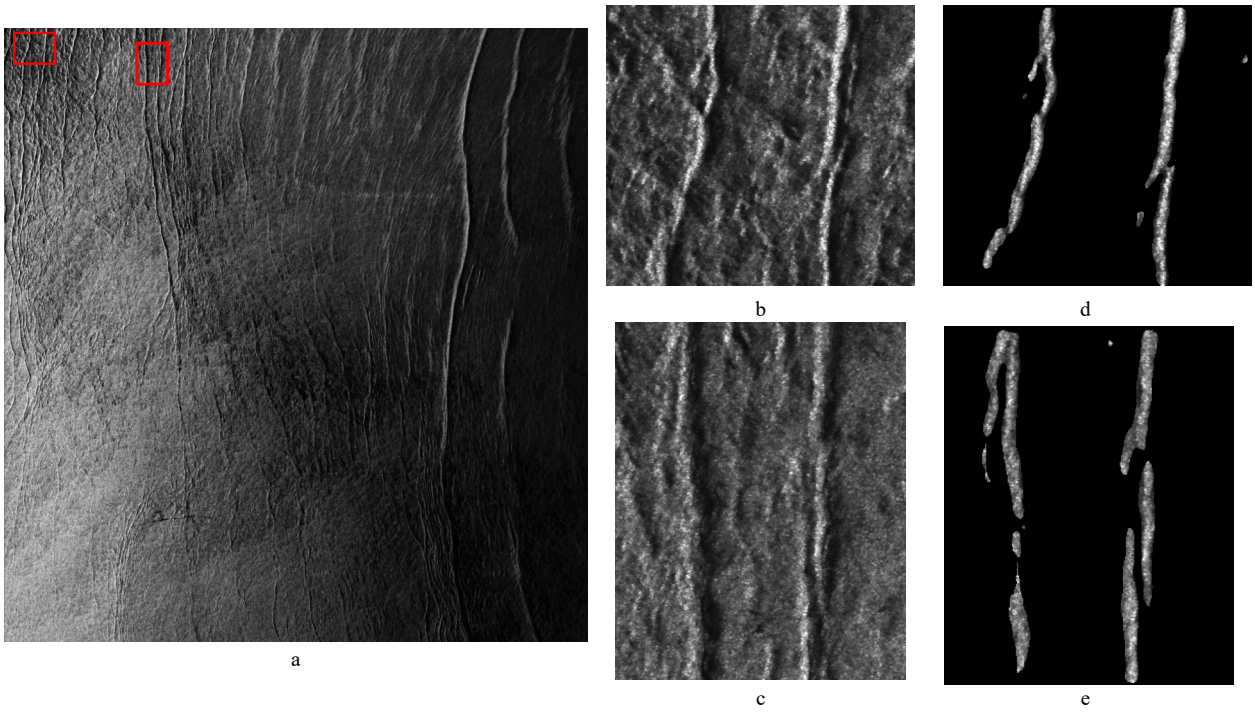

Fig. 12. Small-scale synthetic aperture radar image segmentation results (a), resolution sizes are 334 × 305 (b) and 282 × 348 (c). d is the segmentation result of b, e is the segmentation result of c.

ution of 9 501 × 7 915 pixels are shown in Fig. 13. The image date is February 8, 2007. The coordinates of the center point are near 6.385 1°N, 96.802 2°E. It is a single look complex product with HH polarization mode. As can be seen from Fig. 13, it is better to identify the bright stripes on the left of the image. While for the subtle internal waves on the right, some stripes cannot be segmented. On the one hand, the reason is the quality of SAR images; on the other hand, it is difficult for the human eye to accurately identify the subtle internal waves when labeling, resulting in the inability to identify the subtle internal wave stripes effectively. Therefore, it is understandable to have some incomplete segmentations in large-scale images.

4 Discussion

In this study, TransUNet was applied to segment oceanic internal waves, and the Transformer layer, MLP channel, and Dropout were optimized to solve the over-fitting problem. The segmentation results of the original TransUNet, optimized TransUNet, and U-Net were compared. It was found that the optimized TransUNet could better segment and retain location information. Li et al. (2020) employed an optimized U-Net to segment oceanic internal waves and achieved good results in detection thereof. Compared with Li et al. (2020) results, it can be found that the location of oceanic internal waves identified by the algorithm in this study were relatively accurate. Zheng et al. (2021) used SegNet to accurately segment the specific location of oceanic internal waves; however, the complexity of the SegNet model was fixed. The TransUNet selected in this study can adjust the model complexity by changing the number of Transformer layers, which is convenient for training data of different scales. Owing to the

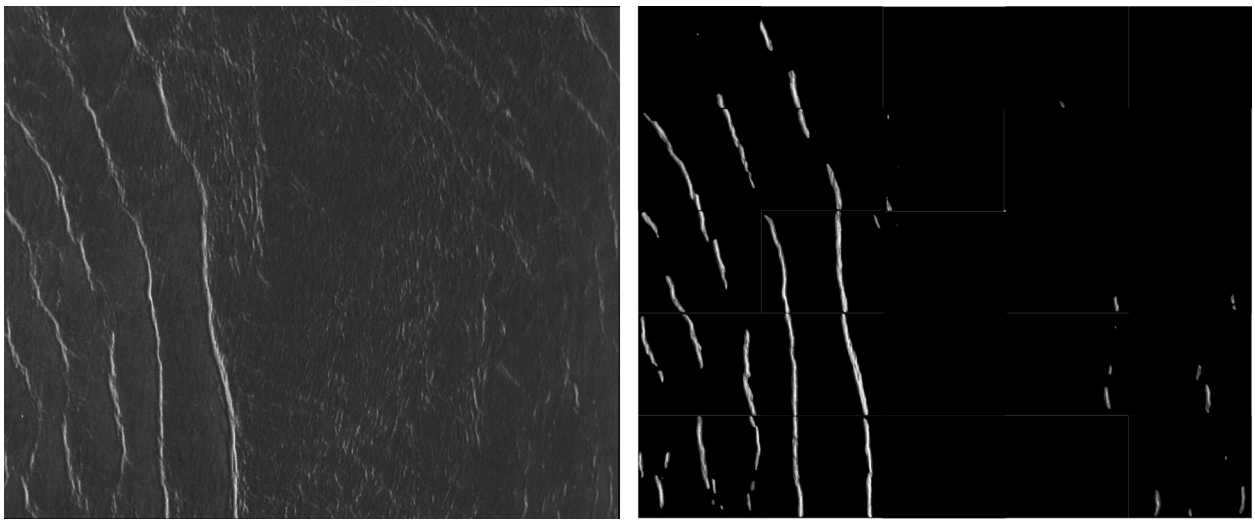

Fig. 13. Results of the entire synthetic aperture radar image segmentation.

limitations in computer performance, the effect of the present algorithm in identifying large-scale images is inadequate. Therefore, the proposed algorithm is suitable for small-scale image segmentation at present. With the advent of the big data era, TransUNet will have immense potential for oceanic internal wave segmentation.

5 Conclusions

Based on TransUNet, this study proposes an algorithm for lightweight oceanic internal wave stripe segmentation. In the proposed algorithm, the Transformer layer, MLP channel, and Dropout parameters that affect the complexity of the TransUNet model are optimized to effectively solve the over-fitting problem. The research results show that the optimized TransUNet can take advantage of the Transformer's global encoding of image feature sequences, creating more complete segmented results. The optimized algorithm can be trained on a microcomputer, which significantly reduces the research threshold. With the development of remote sensing technology, SAR image data will become increasingly available, and the original TransUNet will be at training massive data. The segmentation algorithm proposed in this study can obtain the specific location information of each internal wave in the image, which is helpful for the further study of oceanic internal waves in SAR images. The optimized model can effectively segment the small-scale images, while some mistaken segmentations for large-scale images are encountered. The small-scale images are trained in this study. If a more robust computer is applied to directly train large-scale images, the effect of large-scale image segmentation should be improved. In addition, it is difficult for the human eye to identify subtle internal waves when labelling accurately. Therefore, it is understandable to have some incomplete segmentation in large-scale images.

Acknowledgements

The authors are grateful to the websites of ESA and ASF, which are used to collect SAR images of the world and support Python platforms. The authors are also grateful to the Shanghai Frontiers Science Center of "Full Penetration" Far-Reaching Offshore Ocean Energy and Power.

References

Alpers W. 1985. Theory of radar imaging of internal waves. Nature, 314(6008): 245–247, doi: 10.1038/314245a0

Bao Sude, Meng Junmin, Sun Lina, et al. 2020. Detection of ocean internal waves based on faster R-CNN in SAR images. Journal of Oceanology and Limnology, 38(1): 55–63, doi: 10.1007/S00343-019-9028-6

Chen Jieneng, Lu Yongyi, Yu Qihang, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation. Preprint arXiv, https://arxiv.org/abs/2102.04306v1[2021-02-08/2022-07-29]

Dosovitskiy A, Beyer L, Kolesnikov A, et al. 2021. An image is worth 16×16 words: transformers for image recognition at scale. Preprint arXiv, https://arxiv.org/abs/2010.11929v2[2021-06-03/2022-07-29]

Krizhevsky A, Sutskever I, Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84–90, doi: 10.1145/3065386

Lavrova O Y, Mityagina M I, Serebryany A N, et al. 2014. Internal waves in the Black Sea: satellite observations and *in-situ* measurements. In: Proceedings of SPIE 9240, Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2014. Amsterdam, Netherlands: SPIE

Li Xiaofeng, Liu Bin, Zheng Gang, et al. 2020. Deep-learning-based information mining from ocean remote-sensing imagery. National Science Review, 7(10): 1584–1605, doi: 10.1093/NSR/NWAA047

Ronneberger O, Fischer P, Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 234–241

Russell B C, Torralba A, Murphy K P, et al. 2008. LabelMe: a database and web-based tool for image annotation. International Journal of Computer Vision, 77(1–3): 157–173, doi: 10.1007/s11263-007-0090-8

Shelhamer E, Long J, Darrell T. 2017. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640–651, doi: 10.1109/TPAMI.2016.2572683

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. Preprint arXiv, https://arxiv.org/abs/1409.1556v6[2015-04-10/2022-07-29]

Srivastava N, Hinton G, Krizhevsky A, et al. 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1): 1929–1958

Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.

Wang Shengke, Dong Qinghong, Duan Lianghua, et al. 2019. A fast internal wave detection method based on PCANet for ocean monitoring. Journal of Intelligent Systems, 28(1): 103–113, doi: 10.1515/JISYS-2017-0033

Zaremba W, Ilya S, Oriol V. 2014. Recurrent neural network regularization. Preprint arXiv, http://arXiv.org/abs/1409.2329v5[2015-02-19/2022-07-29]

Zhang Hao, Meng Junmin, Sun Lina, et al. 2020. Performance analysis of internal solitary wave detection and identification based on compact polarimetric SAR. IEEE Access, 8: 172839–172847, doi: 10.1109/ACCESS.2020.3025946

Zheng Yinggang, Zhang Hongsheng, Qi Kaituo, et al. 2022. Stripe segmentation of oceanic internal waves in SAR images based on SegNet. Geocarto International, 37(25): 8567–8578, doi: 10.1080/10106049.2021.2002430

Zheng Yinggang, Zhang Hongsheng, Wang Youqiang. 2021. Stripe detection and recognition of oceanic internal waves from synthetic aperture radar based on support vector machine and feature fusion. International Journal of Remote Sensing, 42(17): 6706–6724, doi: 10.1080/01431161.2021.1943040