Precise and Fast Segmentation of Offshore Farms in High-Resolution SAR Images Based on Model Fusion and Half-Precision Parallel Inference

Chuang Yu, *Member, IEEE*, Yunpeng Liu, *Member, IEEE*, Xin Xia, Deyan Lan, Xin Liu, and Shuhang Wu

***Abstract*—In aquaculture, using high-resolution synthetic aperture radar (SAR) images to precisely segment offshore farms is helpful for reasonable layout planning and statistics of breeding density. However, conventional segmentation methods tend to have low accuracy and slow inference speed. Therefore, we propose a novel, precise, and fast segmentation scheme for offshore farms in high-resolution SAR images based on model fusion and half-precision parallel inference. Specifically, we propose several new high-performance improved UNet++ methods and reasonably fuse the test results. At the same time, a simulated annealing strategy and a morphological closing operation are introduced to improve the segmentation accuracy. In addition, we find that resizing the images to 256 × 256 pixels is better than 512 × 512 pixels for this task, which not only has higher segmentation accuracy but can also increase the inference speed by nearly 13%. Furthermore, a novel half-precision parallel inference strategy is proposed, which can fully utilize the GPU and increase the inference speed by 72.6%. Compared with some state-of-the-art methods, the proposed scheme that merges two improved UNet++ achieves superior accuracy with a frequency weighted intersection over union of 0.9876 and a single image inference time of 0.0218 s on the high-resolution SAR offshore farm dataset.**

***Index Terms*—Half-precision parallel inference, improved UNET++, SAR images, segmentation, simulated annealing.**

I. Introduction

SYNTHETIC aperture radar (SAR), as an active imaging system, has a certain surface penetration capability and can work in all-weather and day–night conditions. It is widely used in marine monitoring, resource exploration, mapping, military, and other fields [1]–[5]. SAR image segmentation aims to assign a label to each pixel, which is the basis for many SAR applications (e.g., crop yield estimation [6] and change detection [7]). Due to the characteristics of imaging, SAR images contain a lot of speckle noise. At the same time, inference speed is another important metric for segmentation tasks. Therefore, it is challenging to achieve more precise and faster segmentation of SAR images.

Image segmentation methods can be mainly divided into nondeep-learning-based methods and deep-learning-based methods. For nondeep-learning-based methods, considering the difference between visible images and SAR images, some excellent methods for visible images [8]–[11] often cannot be directly applied to SAR images [12], [13]. To minimize the negative impact of speckle noise on SAR image segmentation, nondeep-learning-based methods for SAR image segmentation include level set [14], [15], clustering [16], [17], graph cuts [18], [19], active contour model [20], [21], edge-based scheme [22]–[25], region-based scheme [26], [27], and hybrid scheme [28], [29]. However, the above-mentioned nondeep-learning-based methods usually lack robustness, which makes it very easy to have a large number of false detections and missed detections in complex scenes.

Deep-learning-based methods can not only adapt to different environments but also have the advantages of high accuracy, fast inference, and self-learning of parameters. Their powerful feature extraction ability and good generalization ability have been verified in many fields [30]–[32]. At the same time, most deep-learning-based segmentation networks can be directly transferred to different source image tasks and achieve excellent results. FCN [33] is one of the earliest deep learning methods applied in image segmentation and achieves relatively excellent segmentation performance. Later, a large number of deep-learning-based segmentation networks are successively proposed [34]–[38]. Olaf *et al.* proposed U-Net [34], which adopts an encoder–decoder structure and merges the low-level features (rich detailed information) with the high-level feature map (rich semantic information) through skip connections. Subsequently, based on U-Net, LinkNet [35] and UNet++ [36] are proposed. UNet++ introduces a built-in U-Net collection with variable depth and redesigns the jumper in U-Net to achieve a better segmentation effect. To better extract the global context for more reliable scene recognition, Zhao *et al.* proposed PSPNet [37]. It uses pyramid pooling to extract multiscale information and then aggregates the context of different regions. Based on PSPNet,

Manuscript received March 6, 2022; revised April 30, 2022 and May 16, 2022; accepted June 1, 2022. Date of publication June 10, 2022; date of current version June 24, 2022. This work was supported by the Innovation Project of Equipment Development Department–Information Perception Technology under Grant E01Z040601. *(Corresponding author: Yunpeng Liu.)*

Chuang Yu is with the Key Laboratory of Opto-Electronic Information Processing and the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, and also with the Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Sheyang 110169, China, and also with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: yuchuang@sia.cn).

Yunpeng Liu, Xin Xia, Deyan Lan, Xin Liu, and Shuhang Wu are with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China (e-mail: ypliu@sia.cn; xiaxin_krist@163.com; landeyan@sia.cn; liuxin1@sia.cn; wushuhang@sia.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3181355

Li *et al.* further proposed PAN [38], which uses feature pyramid attention to focus on extracting useful features and global attention upsampling to supplement low-level information with high-level information. It is undeniable that some recent studies [39]–[41] have achieved excellent results on the multiclass segmentation task of SAR images. However, considering that the studied task contains only one class of objects, an overly complex network is not cost-effective. At the same time, the networks proposed in [34]–[38], can already achieve high segmentation accuracy and better scalability for the studied tasks.

For the deep-learning-based segmentation methods, an excellent feature extraction network can better extract the features of the target and effectively suppress the interference of the background [42]. In the early days, Simonyan *et al.* proposed VGG19 [43]. However, the network requires a large amount of calculation. Therefore, Szegedy *et al.* proposed Inception v1 [44], which uses the inception structure, pointwise convolutional, and mean pooling to reduce the calculation. Then, they subsequently proposed Inception v2 [45], Inception v3 [45], Inception v4 [46], and Inception-ResNet [46]. To improve the performance of Inception v3, Chollet *et al.* proposed Xception [47]. The network adopts a depthwise separable convolution, which improves the effectiveness of the model without increasing the network complexity. To solve the phenomenon of gradient disappearance or explosion caused by the superposition of network layers, He *et al.* proposed the ResNet [48]. Subsequently, Jie *et al.* introduced the attention mechanism into ResNet and proposed SE-Net [49]. In addition, Tan *et al.* proposed EfficientNet [50], which proposes a multidimensional hybrid model scaling method. Radosavovic *et al.* proposed RegNet [51], which combines the advantages of neural architecture search with manual design.

For SAR image segmentation, nondeep-learning-based methods often heavily rely on handcrafted features. However, due to the speckle noise in SAR images, it is difficult to extract the desired handcrafted features. Considering the strong learning ability of the deep learning segmentation method and the importance of the feature extraction network, we explore some new high-performance improved networks by combining multiple segmentation methods with different feature extraction networks. Subsequently, the test results of several excellent improved networks are reasonably fused to achieve more precise segmentation. At the same time, considering that the model easily falls into the local optimum and the result has the phenomenon of missing segmentation in some small regions, we introduce a simulated annealing strategy and a morphological closing operation to improve the segmentation accuracy. In addition, considering that the image source studied is SAR and the research object is the offshore farms distributed in blocks, we consider that a reasonable downsampling operation can eliminate the noise of the SAR images while losing little target information. It not only relatively increases the receptive field of the network but also increases the inference speed of the network. Therefore, we propose to resize the images to 256×256 pixels instead of the general 512×512 pixels. Furthermore, considering that inference speed is another important metric for segmentation tasks, we consider that the commonly used single-image testing method wastes many GPU resources. To fully utilize the GPU, we propose a novel half-precision parallel inference strategy, which can achieve a huge increase in inference speed while maintaining accuracy. It has excellent practical application value. It is worth noting that the proposed scheme is applied to the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition and achieves excellent results. In summary, we propose a precise and fast segmentation scheme for offshore farms in high-resolution SAR images based on model fusion and half-precision parallel inference, which has excellent performance and application prospects. The contributions of this article are as follows.

1) We propose several new high-performance improved UNet++ methods and reasonably fuse the test results to improve segmentation accuracy.
2) A novel half-precision parallel inference strategy is proposed, which can fully utilize the GPU and achieve a huge increase in inference speed.
3) We find that it is better to resize the images to 256×256 pixels instead of the general 512×512 pixels for the studied task, which has higher segmentation accuracy and faster inference speed.
4) A simulated annealing strategy and a morphological closing operation are introduced to improve segmentation accuracy.

The rest of this article is arranged as follows. In Section II, the proposed scheme and each component module are introduced in detail. Section III introduces the dataset and experimental settings, verifies the improvement of the proposed scheme by comparative experiments, and analyzes the results of the experiment. Finally, Section IV concludes this article.

II. Method

A. Proposed Scheme

As shown in Fig. 1, the proposed scheme can be divided into model training and model application. In the model training part, first, the training samples are uniformly resized to 256×256 pixels and data augmentation operations are performed. While meeting the input size of the model, it expands the training set and enhances the robustness of the generated model [52]. Then, the preprocessed images are input into the improved UNet++ methods for training. The simulated annealing strategy is adopted in network training to constrain updating the learning rate. Finally, the validation set is used to select the best models generated by different methods. In the model application part, first, the test samples are resized to 256×256 pixels. Second, the test samples are made into a sample set in batches, which are sequentially input into the trained models. It is worth noting that both the produced sample set and the trained models have undergone half-precision data format conversion. Then, the test results generated by different models are merged, and the size of the image is restored to the original image. Finally, the morphological closing operation is adopted, and the required binary segmentation images are output.

B. Improved UNet++ Methods

Due to the limited feature extraction ability of the original UNet++ network, we replace the encoding part of UNet++

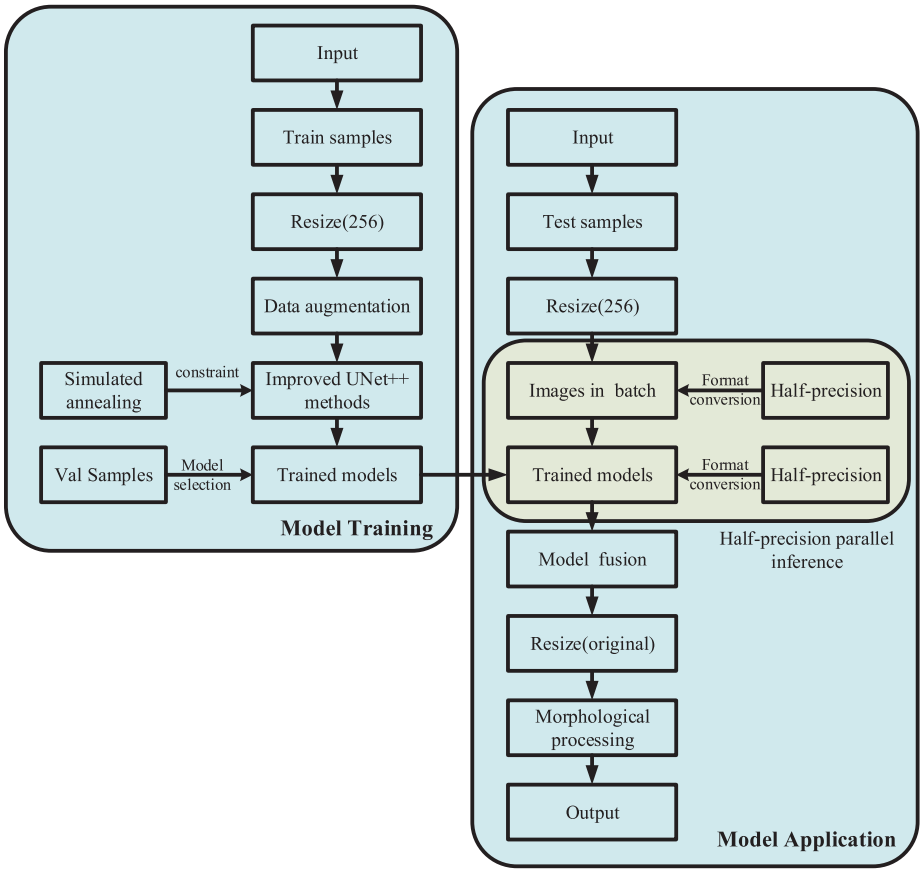

Fig. 1. Overall scheme.

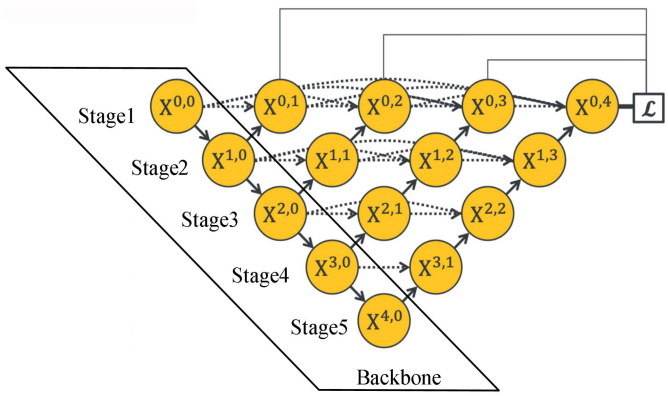

Fig. 2. Network structure of the improved UNet++ method. The parallelogram part is the improvement, which replaces the output of the encoding part of the original UNet++ with the five-stage feature maps output by different high-performance backbones.

with some excellent feature extraction networks (backbones) [36], [53]. As shown in Fig. 2, we combine some high-performance backbones (SE-Net [49], Effientnet [50], RegNet [51], ResNet50 [48], ResNet101 [48], Xception [47], Inception-ResNet [46], and Inception v4 [46]) as the encoding part of UNet++.

Inspired by the feature pyramid network (FPN), which extracts multiscale feature maps [54], the encoding part of UNet++ extracts and outputs five-stage feature maps similar to other high-performance backbones. Therefore, we consider replacing the encoding part of UNet++ with some high-performance backbones. Then, multiple improved UNet++ methods with higher performance are constructed. At the same time, our proposed networks are based on UNet++. UNet++ is an improvement based on U-Net, which is widely used for its excellent performance on segmentation tasks. UNet++ introduces a built-in variable depth U-Net [34] collection and redesigns the jumper in U-Net. The structure can obtain better segmentation performance for objects of different sizes while realizing flexible feature fusion. In addition, training variable depth U-Net sets embedded in the UNet++ architecture can stimulate collaborative learning between U-Nets.

C. Model Fusion

Since using a single network model for segmentation will cause missed detection and false detection, we propose a model fusion strategy. Different network models extract different features from the same image, and their segmentation effects on the same region vary [55]. Therefore, fusing the segmentation results of two or more network models can reduce false segmentation. As shown in Fig. 3, since each pixel value of the output image is a probability value belonging to the target, the final segmentation result needs to be binarized. By observing

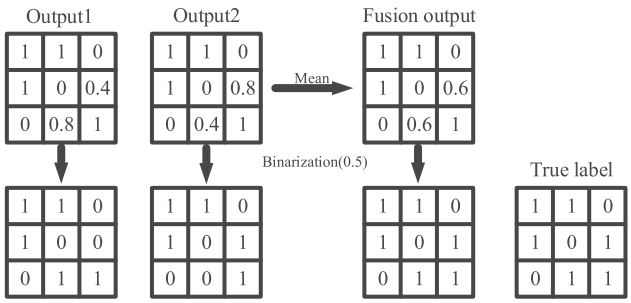

Fig. 3. Fusion of the two models. "Mean" denotes averaging the corresponding point values of the input image. "Binarization (0.5)" denotes binarization processing with 0.5 as the threshold.

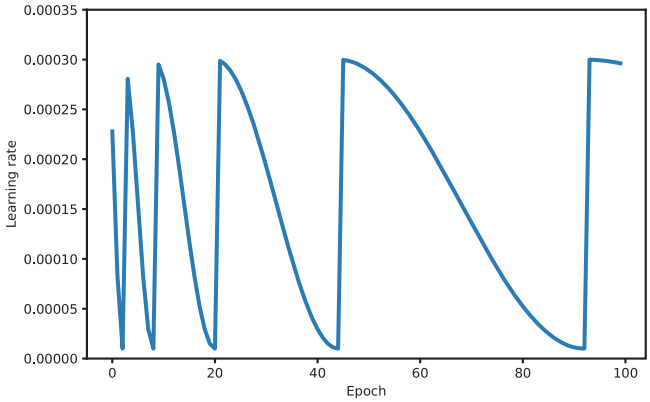

Fig. 4. Curve of the learning rate when using the simulated annealing strategy.

the output of the two single models, false segmentation occurs. The result of model fusion is consistent with the true label.

D. Simulated Annealing Strategy

The deep learning method easily falls into a local optimum during network training. Optimization methods, such as Momentum [56], Adagrad [57], and Adam [58], can make the adjustment of internal parameters more reasonable, reduce the occurrence of overfitting, and minimize the loss function. However, the problem cannot be completely solved if only optimization methods are used. Therefore, we introduce a simulated annealing strategy [53], [59].

From Fig. 4, on the one hand, the change trend of the learning rate is corrugated using the simulated annealing strategy. When the network weight falls into the local optimum, the initial learning rate makes the network weight update step larger, which helps to jump out of the local optimum. On the other hand, every two epochs of the initial learning rate interval increase by multiples, which provides a sufficient number of iterations for the weight to converge to the global optimum.

E. Morphological Closing Operation

To further realize the refined segmentation of test samples, we perform a morphological closing operation on the image after segmentation and binarization processing [60], [61]. As

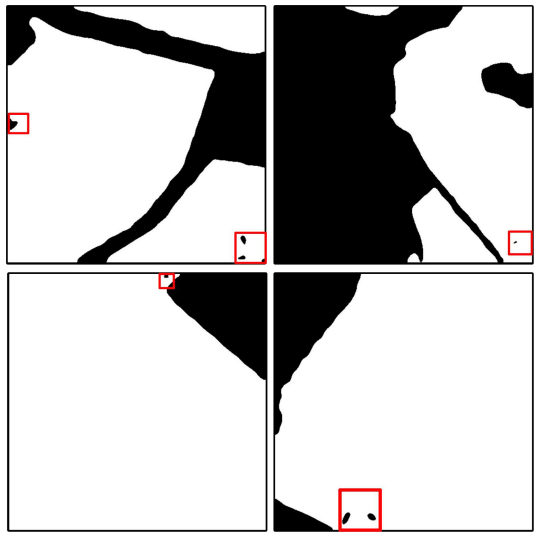

Fig. 5. Binary images after segmentation. The white areas are the target areas, and the black areas in the red frames are the missed detection areas.

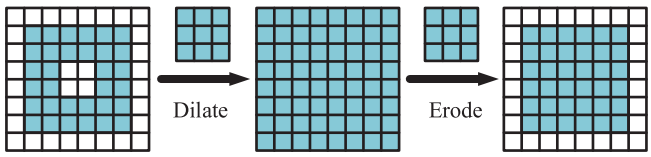

Fig. 6. Morphological closing operation.

shown in Fig. 5, there are holes in some target areas of the segmented binary images. We consider that the phenomenon is caused by the large amount of noise in the SAR images and the inconspicuous features. The boundary information of the target area is obvious, but the characteristics of the internal area and the background area are highly similar.

To reduce the missed detection inside the target area, we perform a morphological closing operation on the segmented binary image. Fig. 6 demonstrates the use of a 3×3 kernel function to perform a morphological closing operation. By performing the morphological dilation operation first and then the morphological erosion operation, the inner area can be filled.

F. Half-Precision Parallel Inference

The inference speed is another important metric for evaluating performance. To achieve faster segmentation speed, we propose a novel half-precision parallel inference strategy. It can fully improve the utilization rate of the GPU under the premise of ensuring accuracy. Fig. 7 shows the test images are first subjected to batch splice and data format conversion operations to generate half-precision batch samples. Then, the batch samples are input into the half-precision model to obtain the batch segmentation images. Finally, the batch segmentation images are split to obtain the required binary segmentation images. Using multiple parallel inferences can make full use of the computing power of the GPU. At the same time, converting the input images and the trained models into a half-precision format can reduce the

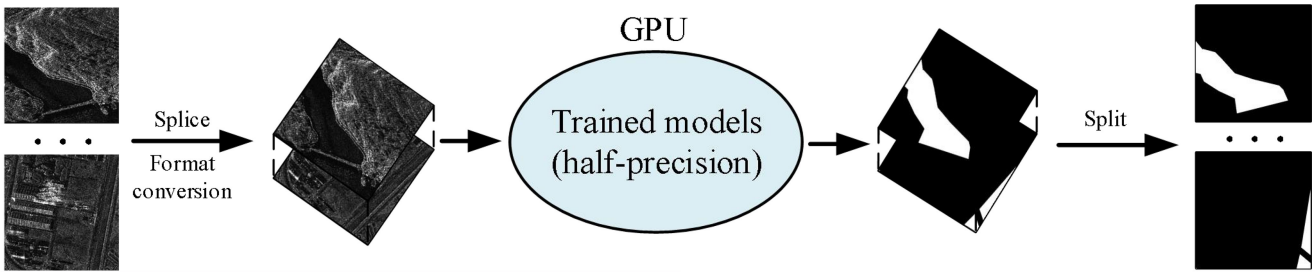

Fig. 7. Half-precision parallel inference strategy.

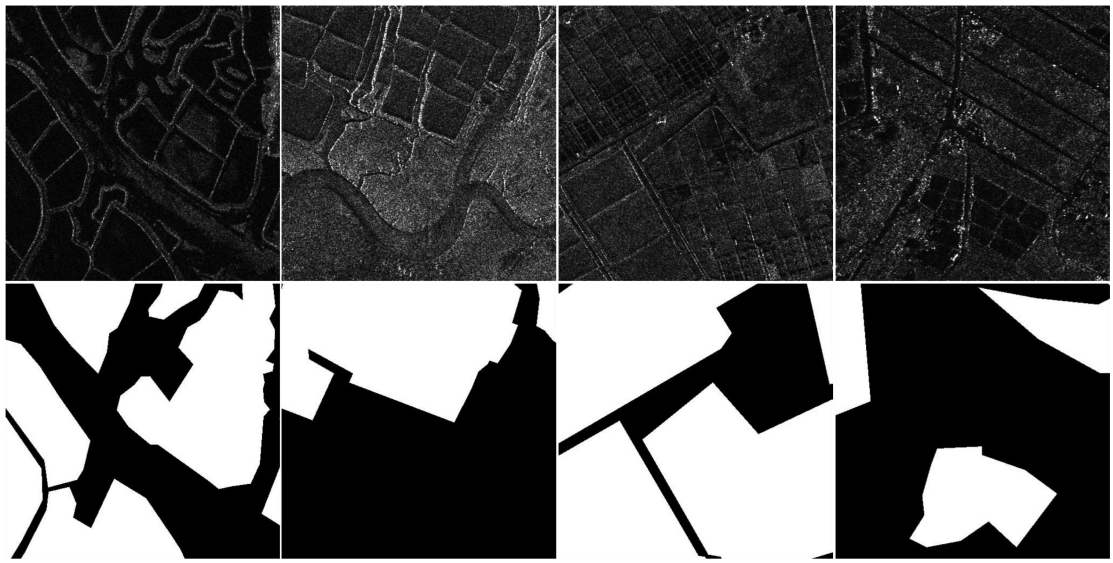

Fig. 8. Some samples are displayed from the high-resolution SAR offshore farm dataset. The first row is the original images, and the second row is the ground truth images corresponding to the first row.

consumption of GPU memory, and more images can be tested per round.

G. Loss Function and Loss Optimization

We combine binary cross-entropy loss with dice loss [62] as the final loss function, which can promote the rapid convergence of the model. Its expression is shown as follows:

$$\varepsilon(Y,P) = -\frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N}\left(y_{n,c}\log p_{n,c} + \frac{2y_{n,c}p_{n,c}}{y_{n,c}^2 p_{n,c}^2}\right) \quad (1)$$

where $p_{n,c} \in P$ and $y_{n,c} \in Y$ denote the target labels and predicted probabilities of the cth and nth pixels in the batch processing, respectively. Y and P denote the ground truth images and prediction results of the test images, respectively. C and N denote the number of classes and pixels, respectively.

III. Experiments

A. Dataset

The high-resolution SAR offshore farm dataset [63] is derived from HiSea-1 and Gaofen-3 with a resolution of 1–3 m. The dataset contains a total of 4000 images with sizes ranging from 512 to 2048 pixels, and the size of most samples is 512×512 pixels. The scene covers China's southeast coastal area, and targets include common large-scale offshore farms. In the local experiment, we divide the 4000 samples into a training set, a test set, and a validation set at a ratio of 3:1:1. It is worth noting that considering that the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition provides an additional unobtainable test set for scoring, we divide these 4000 samples into a training set and a validation set at a ratio of 4:1 in the competition. It allows the generated model to learn more samples. Fig. 8 shows part of the sample images and their true labels. The SAR images contain a large amount of noise, and the features are not obvious. The dataset is available online.[1]

B. Experimental Environment and Parameter Settings

The operating system and GPU are Ubuntu18.04 and RTX 2080Ti 11G, respectively. The epochs and batch sizes are 100 and 8, respectively. The optimizer, learning rate, momentum, and weight decay are AdamW [64], 0.0003, 0.9, and 0.0005,

[1][Online]. Available: http://gaofen-challenge.com/challenge/dataset/3

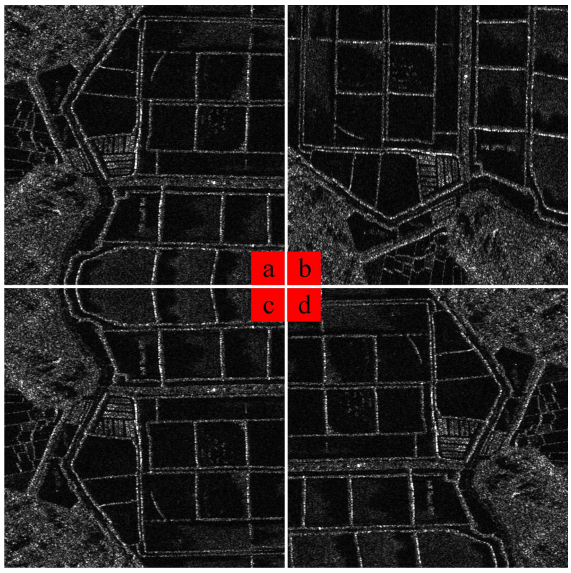

Fig. 9. Data augmentation. (a) Original image. (b) Original image rotated 90° counterclockwise. (c) Original image flipped vertically. (d) Original image flipped horizontally.

TABLE I
PERFORMANCE COMPARISON OF VARIOUS SEGMENTATION METHODS

Methods	FWIoU	Time on GPU/s
FCN(EfficientNet) [33], [50]	0.9858	0.0446
U-Net(EfficientNet) [34], [50]	0.9834	0.0447
LinkNet(EfficientNet) [35], [50]	0.9847	0.0449
PSPNet(EfficientNet) [37], [50]	0.9799	**0.0219**
PAN(EfficientNet) [38], [50]	0.9839	0.0451
UNet++(EfficientNet) [36], [50]	**0.9861**	0.0465

TABLE II
COMPARISON OF THE FWIOU UNDER DIFFERENT METHODS AND LOSS FUNCTIONS

Methods	binary cross-entropy loss	dice loss	combined loss
U-Net(EfficientNet)	0.9831	0.9793	**0.9834**
LinkNet(EfficientNet)	0.9838	0.9811	**0.9847**
UNet++(EfficientNet)	0.9857	0.9804	**0.9861**

respectively. The initial rounds of simulated annealing are 3. The number of epochs between the two restarts increases by multiples. It is worth noting that for the model uploaded to the final competition of the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation," the GPU used is RTX teslaV100 32G. The epochs and batch sizes are 200 and 32, respectively.

The optimization of weights in deep learning networks requires a large number of training samples. As shown in Fig. 9, the distribution of offshore farms is not affected by rotation and flip transformations. We randomly use rotation transformation and flip transformation to enhance and expand the training samples. It can reduce overfitting and improve the robustness and generalization of the generative models [65].

For the evaluation metric, to better evaluate the performance of the proposed model and keep it consistent with the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition, we adopt frequency weighted intersection over union (FWIoU) [66].

C. Selection of the Segmentation Method

To achieve precise segmentation of offshore farms in high-resolution SAR images, we compare some excellent segmentation methods. It is worth noting that to rigorously compare the performance differences between the methods instead of the performance improvement brought by the feature extraction network, we uniformly use EfficientNet to replace the feature extraction network of various methods (FCN [33], U-Net [34], LinkNet [35], PSPNet [37], PAN [38], and UNet++ [36]). All methods in Table I use the simulated annealing strategy, and the images are uniformly resized to 256×256 pixels for input. "(EfficientNet)" denotes using EfficientNet to replace the feature extraction network part of the corresponding method.

From Table I, UNet++ achieves the best segmentation accuracy with an FWIoU of 0.9861. Compared with PSPNet, PAN, FPN, U-Net, and LinkNet, UNet++ increases FWIoU by 0.0062 (from 0.9799 to 0.9861), 0.0022 (from 0.9839 to 0.9861), 0.0003 (from 0.9858 to 0.9861), 0.0027 (from 0.9834 to 0.9861), and 0.0014 (from 0.9847 to 0.9861), respectively. It is worth noting that the purpose of our research is to achieve refined segmentation of offshore farms in high-resolution SAR images. On the one hand, the FWIoU of most deep learning methods is above 98%, and small improvements in FWIoU will become more difficult and meaningful. On the other hand, when considering the actual application for counting the area and density of farms, the improvement in accuracy will help farmers perform more refined management and reduce farming accidents. In addition, for the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition, we find that the distribution of the test set of the competition is not completely consistent with the local dataset. The experimental results show that the test set of the competition will amplify the difference in segmentation accuracy between the methods and the gap can be as high as 0.03. For single image inference time, although UNet++ is 0.0246 s more than PSPNet, the accuracy of PSPNet is significantly lower. Except for PSPNet, the other five methods have little difference in single image inference time. In summary, UNet++ has relatively optimal performance.

D. Effect Verification of the Loss Function

We consider that combining binary cross-entropy loss with dice loss as the final loss function has better performance. To fully verify this hypothesis, we carry out comparative verification on LinkNet, U-Net, and UNet++. All methods in Table II use the simulated annealing strategy, and the images are uniformly resized to 256×256 pixels for input.

TABLE III
COMPARISON OF THE FWIOU UNDER DIFFERENT METHODS AND WHETHER TO USE THE SIMULATED ANNEALING STRATEGY

Methods	No simulated annealing	simulated annealing
U-Net(EfficientNet)	0.9713	**0.9834**
LinkNet(EfficientNet)	0.9717	**0.9847**
UNet++(EfficientNet)	0.9719	**0.9861**

TABLE IV
COMPARISON OF THE FWIOU UNDER DIFFERENT METHODS AND DIFFERENT RESIZING STRATEGIES

Methods	512	256
U-Net(EfficientNet)	0.9828	**0.9834**
LinkNet(EfficientNet)	0.9840	**0.9847**
UNet++(EfficientNet)	0.9842	**0.9861**

The experimental results presented in Table II verify that combining binary cross-entropy loss with dice loss as the final loss function has the best performance. At the same time, the results of binary cross-entropy loss and combined loss are similar, and they are both significantly better than dice loss. Taking UNet++(EfficientNet) as an example, compared with using dice loss, the FWIoU using combined loss can be improved by 0.0057, which is significantly improved. Since the improvement of the loss function only affects the network training, it will not have any impact on the inference speed. Therefore, using combined loss is an optimal choice.

E. Effect Verification of the Simulated Annealing and Resizing Strategy

We consider that using a simulated annealing strategy for updating the learning rate can avoid the model falling into a local optimum. To fully verify this hypothesis, we carry out comparative verification on LinkNet, U-Net, and UNet++. In Table III, the images are uniformly resized to 256×256 pixels for input.

From Table III, compared with not using the simulated annealing strategy, the FWIoU improves by more than 0.01. This shows that using the simulated annealing strategy for updating the learning rate can significantly improve the segmentation accuracy. At the same time, it will not affect the inference speed.

In addition, we consider that the segmentation performance of uniformly resizing the images to 256×256 pixels is better than that of 512×512 pixels. Therefore, we conduct experiments on LinkNet, U-Net, and UNet++. The experimental results are shown in Table IV. All methods use the simulated annealing strategy; "256" and "512" denote that the images are uniformly resized to 256×256 pixels and 512×512 pixels for input, respectively.

From Table IV, compared to uniformly resizing the images to 512 $\times$512 pixels for training and testing, resizing images to 256 $\times$256 pixels can improve the FWIoU. We consider that there are two main reasons. On the one hand, the studied data are SAR images and the influence of noise is more serious.

TABLE V
COMPARISON OF THE SINGLE IMAGE INFERENCE TIME UNDER DIFFERENT METHODS AND DIFFERENT RESIZING STRATEGIES

Methods	512	256
U-Net(EfficientNet)	0.0503	**0.0447**
LinkNet(EfficientNet)	0.0494	**0.0449**
UNet++(EfficientNet)	0.0567	**0.0465**

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT BACKBONES AS THE FEATURE EXTRACTION NETWORK OF UNET++

Methods	FWIoU	Time on GPU/s
UNet++ [36]	0.9819	0.0180
UNet++(ResNet50) [36], [48]	0.9834	0.0238
UNet++(ResNet101) [36], [48]	0.9814	0.0294
UNet++(Xception) [36], [34]	0.9836	0.0202
UNet++(Inception-ResNet) [36], [46]	**0.9842**	0.0492
UNet++(Inception v4) [36], [46]	0.9841	0.0509
UNet++(RegNet) [36], [51]	0.9836	0.0287
UNet++(SE-Net) [36], [49]	**0.9853**	0.0425
UNet++(EfficientNet) [36], [50]	**0.9861**	0.0465

By downsampling the input images, small disturbances, such as noise, can be reduced. On the other hand, the offshore farm targets in the images are distributed in blocks. The downsampling operation can make the size of the input images smaller without losing the target. The network's receptive field for image features is relatively enlarged, thereby improving robustness and generalization.

There will be differences in the inference speed when using different resizing strategies. We conduct comparative experiments on LinkNet, U-Net, and UNet++. From Table V, compared with resizing the images to 512×512 pixels, when resizing images to 256×256 pixels, LinkNet speeds up by 9.1% (from 0.0494 to 0.0449), U-Net speeds up by 11.1% (from 0.0503 to 0.0447), and UNet++ speeds up by 18.0% (from 0.0567 to 0.0465). In summary, resizing images to 256×256 pixels not only achieves better segmentation accuracy but also increases the inference speed by nearly 13%.

F. Effect Verification of Model Fusion

From the above-mentioned experimental results, compared with PSPNet, PAN, LinkNet, FPN, and U-Net, UNet++ has relatively better segmentation performance. However, the segmentation ability of a single model is limited. Therefore, we propose using model fusion to further improve the segmentation accuracy. The models that can be fused need to have different understandings of images. Additionally, to further explore the performance of different backbones replacing the encoding part of UNet++, we conduct experiments on different backbones (ResNet50 [48], ResNet101 [48], Xception [47], Inception-ResNet [46], Inception v4 [46], RegNet [51], SE-Net [49], and EfficientNet [50]). The experimental results are shown in

TABLE VII
PERFORMANCE COMPARISON OF DIFFERENT COMBINATIONS OF IMPROVED UNET++ MODELS

Methods	FWIoU	Time on GPU/s
UNet++(Inception-ResNet)[36, 46]	0.9842	0.0492
UNet++(SE-Net)[36, 49]	0.9853	0.0425
UNet++(EfficientNet)[36, 50]	0.9861	0.0465
UNet++(Inception-ResNet and SE-Net)[36, 46, 49]	0.9864	0.0816
UNet++(Inception-ResNet and EfficientNet)[36, 49, 50]	0.9867	0.0858
UNet++(SE-Net and EfficientNet)[36, 49, 50]	**0.9876**	0.0792
UNet++(Inception-ResNet , SE-Net and EfficientNet)[36, 46, 49, 50]	**0.9876**	0.1182

Table VI. The backbone in parentheses denotes the feature extraction network of UNet++.

From Table VI, the single image inference time of the improved UNet++ methods of the top three segmentation accuracies is relatively high. However, the main purpose of the model fusion strategy is to improve the segmentation accuracy. We initially tentatively merge three improved UNet++ methods (Inception-ResNet, SE-Net, and EfficientNet). To further explore the combination of the three improved methods, we conduct experiments on the performance of the different combinations. The experimental results are shown in Table VII.

From Table VII, with the fusion of models, the segmentation accuracies will indeed improve. The segmentation accuracy of the two improved UNet++ methods (SE-Net and EfficientNet, 0.98756) and the three improved UNet++ methods (Inception-ResNet, SE-Net, and EfficientNet, 0.98761) are almost equal. From the single image inference time, it can be found that when the models are fused, the time is close to the sum of the time of each model. It is worth noting that from the test results of the uploaded model in the preliminary stage of the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition, the single model has a low score. When using two models for fusion, the FWIoU can increase by 0.01 or more. When more models are fused, the increase in FWIoU begins to slow down significantly. Therefore, we decide to adopt two improved UNet++ methods (SE-Net and EfficientNet) as the preferred fusion scheme.

The balance between the inference speed and the segmentation accuracy can be based on different application scenarios and evaluation metrics. For example, for the preliminary stage of the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition, accuracy is used as the only evaluation metric. At that time, the fusion of multiple models will have a relatively excellent accuracy improvement effect. For the final stage of the competition, the inference speed is another important evaluation metric. It is necessary to make a reasonable balance between the number of models for fusion and the inference speed.

To compare the segmentation effects of various schemes more intuitively, we show the segmentation effects of UNet++, UNet++(SE-Net), UNet++(EfficientNet), and UNet++(SE-Net and EfficientNet), as shown in Fig. 10. By observing the second to fifth rows in Fig. 10, the improved UNet++ has a significantly better segmentation effect than the original UNet++.

TABLE VIII
IMPACT OF MORPHOLOGICAL CLOSING OPERATIONS WITH DIFFERENT KERNEL FUNCTION SIZES ON SEGMENTATION PERFORMANCE

Methods	FWIoU	Time on GPU/s
No kernel	0.98761	0.0792
3×3 kernel	**0.98764**	0.0796
5×5 kernel	0.98762	0.0797
7×7 kernel	0.98754	0.0799

The original UNet++ is more likely to misjudge nontargets between target regions as targets. By observing the fourth to sixth rows in Fig. 10, it can be seen that the model fusion scheme does have a better segmentation effect on the boundary of the target than a single network. At the same time, by observing the sixth and seventh columns in Fig. 10, it can be seen that various methods have better segmentation effects for some simple scenes. In addition, by observing the eighth column in Fig. 10, it can be seen that various methods have misjudgments for some complex scenes. We think the possible reason is the high similarity to the target.

G. Effect Verification of the Morphological Closing Operation

To further improve the accuracy of segmentation, we introduce a morphological closing operation. Considering that the edge information of the research target is obvious but the interior area is similar to the background area, it is prone to missed detection. At the same time, to study the impact of different sizes of kernel functions on performance, we perform a morphological closing operation on the fusion of two improved UNet++ methods (SE-Net and EfficientNet). The experimental results are shown in Table VIII.

From Table VIII, the morphological closing operation is used, and the additional time consumed by the operation is almost negligible. In addition, using a morphological closing operation with a kernel size of 3×3 or 5×5 has a slight improvement in FWIoU, which shows effectiveness. When uploading the model to the "2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation" competition for testing, its accuracy can be increased by approximately 0.005, and a good improvement effect is achieved. The results further illustrate the effectiveness of the morphological closing operation, and the proposed scheme has stronger robustness and generalization in complex scenarios.

Fig. 11 shows the segmentation effect of the final scheme, which combines two improved UNet++ methods (SE-Net and EfficientNet) and a morphological closing operation with a 3×3 kernel. Although SAR images have the characteristics of multiple noises and inconspicuous features, there is a small part of false detection areas between two adjacent target areas in some images. The segmentation boundaries of most images are relatively accurate. This shows that the overall segmentation effect of the proposed scheme is very good.

H. Effect Verification of Half-Precision Parallel Inference

The inference speed is one of the important metrics to evaluate the performance of the proposed scheme. To improve the speed of inference, in addition to the resizing strategy, we also propose

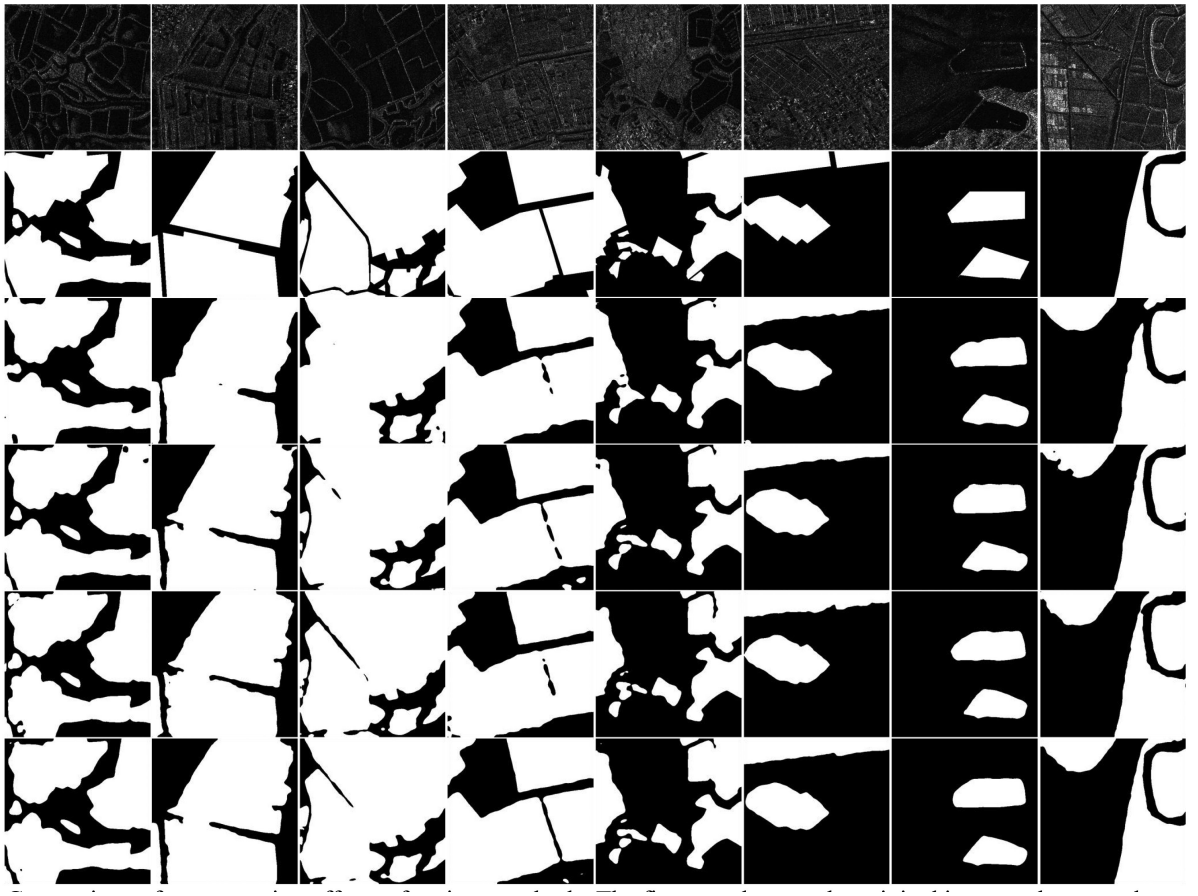

Fig. 10. Comparison of segmentation effects of various methods. The first row denotes the original images; the second row denotes the ground truth images; and the third to sixth rows denote the segmentation results of UNet++, UNet++ (SE-Net), UNet++ (EfficientNet), and UNet++ (SE-Net and EfficientNet), respectively.

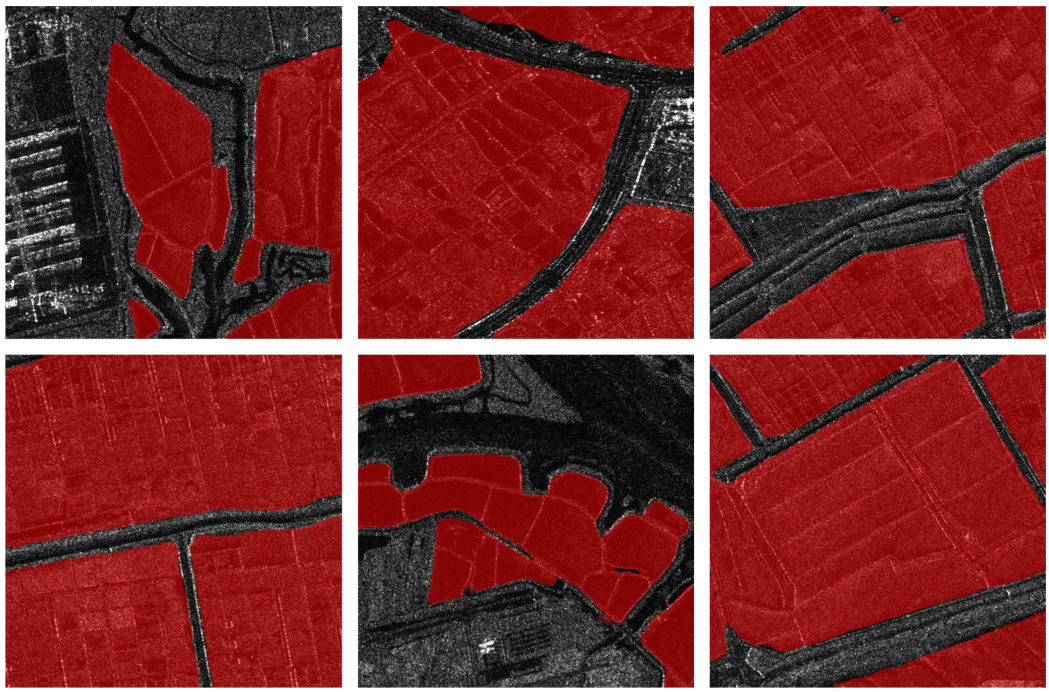

Fig. 11. Segmentation effect of the proposed scheme on the test sample.

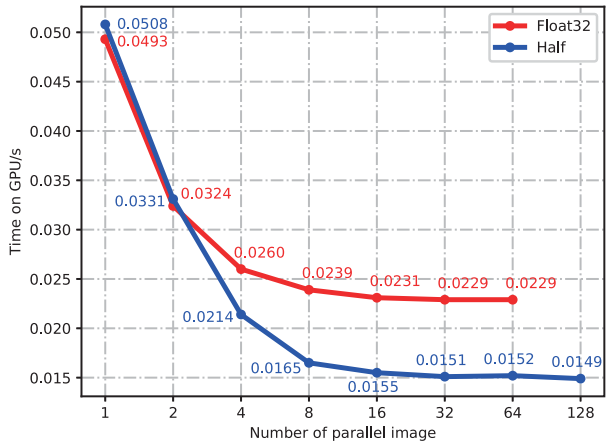

Fig. 12. The changes in the number of parallel computing images and the comparison of the single image inference times under different data formats.

a novel half-precision parallel inference strategy. As shown in Fig. 12, we compare the single image inference times of the different data formats (Float32 and Half) with different numbers of parallel test images on the improved UNet++ (EfficientNet).

From Fig. 12, both curves decline first and then tend to be stable, and the blue curve is lower. Compared with a single image for testing, using full-precision (Float) and half-precision (Half) parallel inference can increase the speed by up to 53.5% (from 0.0493 to 0.0229) and 70.7% (from 0.0508 to 0.0149), respectively. On the one hand, the result shows that the parallel inference method can significantly reduce the single image inference time. On the other hand, it verifies that half-precision has better acceleration performance. When the half-precision is adopted, the GPU memory space consumed by the models and images is greatly reduced, which facilitates the rapid exchange of data. It also explains that when the number of parallel computing images is 128, the GPU memory will be insufficient if the "Float32" is used, whereas the half-precision can be calculated. When the test image is only 1 or 2 each time, the inference speed using "Float32" is faster than "Half." We consider that the utilization rate of GPU in this situation is low, there is less congestion in data exchange, and the calculation optimization of GPU for "Float32" is better than that of "Half." However, when the number of test images increases each time, the utilization rate of the GPU gradually increases and the ratio of data exchange time is relatively higher. Therefore, our final scheme uses two improved UNet++ methods (SE-Net and EfficientNet), a simulated annealing strategy, a resizing strategy, a morphological closing operation, and a half-precision parallel inference strategy. The experimental results show that the scheme has an FWIoU of 0.9876 and the single image inference time is only 0.0218 s. Compared with the scheme that does not use the half-precision parallel inference strategy, the speed increases by 72.6% (from 0.0796 to 0.0218). At the same time, compared with using only one improved UNet++ method (EfficientNet) without half-precision parallel inference and morphological closing operation, the inference speed of the proposed final scheme increases by 53% (from 0.0465 to 0.0218).

IV. Conclusion

To achieve precise and fast segmentation of offshore farms in high-resolution SAR images, we propose a scheme based on model fusion and half-precision parallel inference. By comparing the performances of multiple excellent deep learning segmentation methods and feature extraction networks, SE-Net and EfficientNet are used as the feature extraction modules to replace the encoding part of UNet++. Then, the results generated by the improved UNet++ methods are fused, which can achieve more precise segmentation. To further improve the segmentation accuracy, we validate the performance of the simulated annealing strategy and the morphological closing operation and incorporate them into the proposed scheme. Considering that the inference speed is another important metric of segmentation performance, on the one hand, we uniformly resize the images to 256×256 pixels, which can improve the inference speed by nearly 13%. On the other hand, we propose a novel half-precision parallel inference strategy that can improve the inference speed by 72.6%. The experimental results show that the FWIoU and single image inference times are 0.9876 and 0.0218 s using the final scheme on the high-resolution SAR offshore farm dataset. It is worth noting that the proposed final scheme merges two improved UNet++ models, which can be adjusted reasonably for different accuracy and inference speed requirements.

References

[1] F. Ma, F. Zhang, Q. Yin, D. L. Xiang, and Y. S. Zhou, "Fast SAR image segmentation with deep task-specific superpixel sampling and soft graph convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2021, Art. no. 5214116, doi: 10.1109/TGRS.2021.3108585.

[2] F. Chen, A. H. Zhang, H. Balzter, P. Ren, and H. Y. Zhou, "Oil spill SAR image segmentation via probability distribution modeling," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 533–554, Jan. 2022.

[3] A. Stokhom, T. Wulf, A. Kucik, R. Saldo, J. Buus-Hinkler, and S. M. Hvidegaard, "AI4SeaIce: Toward solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 4304013, doi: 10.1109/TGRS.2022.3149323.

[4] P. Chen, H. Zhou, Y. Li, B. X. Liu, and P. Liu, "Shape similarity intersection-over-union loss hybrid model for detection of synthetic aperture radar small ship objects in complex scenes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9518–9529, Oct. 2021, doi: 10.1109/JSTARS.2021.3112469.

[5] F. Sharifzadeh, G. Akbarizadeh, and Y. S. Kavian, "Ship classification in SAR images using a new hybrid CNN-MLP classifier," *J. Indian Soc. Remote Sens*, vol. 47, no. 4, pp. 551–562, Apr. 2019.

[6] K. Clauss, M. Ottinger, P. Leinenkugel, and C. Kuenzer, "Estimating rice production in the Mekong delta, Vietnam, utilizing time series of Sentinel-1 SAR data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 73, pp. 574–585, Dec. 2018, doi: 10.1016/j.jag.2018.07.022.

[7] L. Zhou, G. Cao, Y. Li, and Y. Shang, "Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3478–3488, Aug. 2016.

[8] S. Y. Chien, Y. W. Huang, B. Y. Hsieh, S. Y. Ma, and L. G. Chen, "Fast video segmentation algorithm with shadow cancellation, global motion compensation, and adaptive threshold techniques," *IEEE Trans. Multimedia*, vol. 6, no. 5, pp. 732–748, Oct. 2004.

[9] H. Y. Lee, N. Codella, M. D. Cham, J. W. Weinsaft, and Y. Wang, "Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac MRI," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 905–913, Apr. 2010.

[10] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[11] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1309–1321, Sep. 2001.

[12] J. Wang, T. Zheng, P. Lei, and X. Bai, "Ground target classification in noisy SAR images using convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4180–4192, Nov. 2018.

[13] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, Mar. 2020, Art. no. 107432.

[14] A. M. Braga, R. Marques, F. Rodrigues, and F. Medeiros, "A median regularized level set for hierarchical segmentation of SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1171–1175, Jul. 2017.

[15] Y. F. Wu, C. J. He, Y. Liu, and M. T. Su, "A backscattering-suppression-based variational level-set method for segmentation of SAR oil slick images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5485–5494, Dec. 2017.

[16] J. C. Fan and J. Wang, "A two-phase fuzzy clustering algorithm based on neurodynamic optimization with its application for PolSAR image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 72–83, Feb. 2018.

[17] G. Akbarizadeh and M. Rahmani, "A new ensemble clustering method for PolSAR image segmentation," in *Proc. 7th Conf. Inf. Knowl. Technol.*, May 2015, pp. 1–4.

[18] P. Salembier and S. Foucher, "Optimum graph cuts for pruning binary partition trees of polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5493–5502, Sep. 2016.

[19] W. K. Tan, J. Li, L. L. Xu, and M. A. Chapman, "Semiautomated segmentation of Sentinel-1 SAR imagery for mapping sea ice in Labrador coast," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1419–1432, May 2018.

[20] G. S. Xia, G. Liu, W. Yang, and L. P. Zhang, "Meaningful object segmentation from SAR images via a multiscale nonlocal active contour model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1860–1873, Mar. 2016.

[21] O. Germain and P. Refregier, "Edge location in SAR images: Performance of the likelihood ratio filter and accuracy improvement with an active contour approach," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 72–78, Jan. 2001.

[22] X. Yang and D. A. Clausi, "Evaluating SAR sea ice image segmentation using edge-preserving region-based MRFs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1383–1393, Oct. 2012.

[23] P.-L. Shui and Z.-J. Zhang, "Fast SAR image segmentation via merging cost with relative common boundary length penalty," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6434–6448, Oct. 2014.

[24] D. L. Xiang *et al.*, "Adaptive statistical superpixel merging with edge penalty for PolSAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2412–2429, Apr. 2020.

[25] M. Modava, G. Akbarizadeh, and M. Soroosh, "Hierarchical coastline detection in SAR images based on spectral-textural features and global-local information," *IET Radar Sonar Navig.*, vol. 13, no. 12, pp. 2183–2195, Dec. 2019.

[26] D. Xiang *et al.*, "Fast pixel–superpixel region merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9319–9335, Nov. 2021, doi: 10.1109/TGRS.2020.3041281.

[27] S. Fan, Y. Sun, and P. Shui, "Region-merging method with texture pattern attention for SAR image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 112–116, Feb. 2020, doi: 10.1109/LGRS.2020.2969321.

[28] E. Liu *et al.*, "SAR image segmentation based on hierarchical visual semantic and adaptive neighborhood multinomial latent model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4287–4301, Jul. 2016.

[29] F. Wang, Y. Wu, M. Li, P. Zhang, and Q. Zhang, "Adaptive hybrid conditional random field model for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 537–550, Jan. 2017.

[30] K. M. He, G. Gkioxari, D. P. Gkioxari, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.

[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv: 1804.02767*.

[32] C. Yu *et al.*, "Infrared small target detection based on multiscale local contrast learning networks," *Infrared Phys. Technol.*, vol. 123, 2022, Art. no. 104107.

[33] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[34] R. Olaf, F. Philipp, and B. Thomas, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Oct. 2015, pp. 234–241.

[35] C. Abhishek and C. Eugenio, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2017, pp. 1–4.

[36] Z. W. Zhou, M. Siddiquee, N. Tajbakhsh, and J. M. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[37] H. S. Zhao, J. P. Shi, X. J. Qi, X. G. Wang, and J. Y. Jia, "Pyramid scene parsing network," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6230–6239.

[38] H. C. Li, P. F. Xiong, J. An, and L. X. Wang, "Pyramid attention network for semantic segmentation," in *Proc. 29th Brit. Mach. Vis. Conf.*, Sep. 2018.

[39] Q. Garg, A. Kumar, N. Bansal, M. Prateek, and S. Kumar, "Semantic segmentation of PolSAR image data using advanced deep learning model," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 15365.

[40] H. X. Bi, F. Xu, Z. Q. Wei, Y. Xue, and Z. B. Xu, "An active deep learning approach for minimally supervised PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9378–9395, Nov. 2019.

[41] Z. Y. Yue, F. Guo, Q. X. Xiong, J. Wang, A. Hussain, and H. Y. Zhou, "A novel attention fully convolutional network method for synthetic aperture radar image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4585–4598, Sep. 2020.

[42] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, May 2015.

[44] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[45] C. Szegedy, V. Vanhoucke, S. Loffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2828–2826.

[46] C. Szegedy, S. Loffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 4278–4284.

[47] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1800–1807.

[48] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[49] H. Jie, S. Li, and S. Gang, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[50] M. X. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, Jun. 2019, pp. 6105–6114.

[51] L. Radosavovic, R. P. Kosaraju, R. Girshick, K. M. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 10425–10433.

[52] B. Zoph, E. D. Cubuk, G. Ghiasi, T. Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," in *Proc. 16th Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 566–583.

[53] C. Yu, Y. P. Liu, and X. Xia, "Precise segmentation of offshore farms in high-resolution SAR images based on improved UNet++," in *Proc. Int. Conf. Comput. Appl. Inf. Secur.*, 2021, pp. 26–34.

[54] T. Y. Lin, P. Dollar, R. Girshick, K. M. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[55] K. Jiang, Z. Y. Wang, P. Yi, G. C. Wang, K. Gu, and J. J. Jiang, "ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2734–2747, Oct. 2020.

[56] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, Jan. 1999.

[57] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Jul. 2011.

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, May 2015, pp. 1–13.

[59] L. LLya and H. Frank, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. 5th Int. Conf. Learn. Representations*, Apr. 2017.
[60] Y. X. Li, B. Peng, L. L. He, K. L. Fan, and L. Tong, "Road segmentation of unmanned aerial vehicle remote sensing images using adversarial network with multiscale context aggregation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2279–2287, Jul. 2019.
[61] C. Yu *et al.*, "Segmentation and density statistics of mariculture cages from remote sensing images using mask R-CNN," *Inf. Process. Agriculture*, 2021.
[62] F. Milletari, N. Navab, and S. A. Ahmdi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th IEEE Int. Conf. 3D Vis.*, Oct. 2016, pp. 565–571.
[63] "2021 Gaofen challenge on automated high-resolution earth observation image interpretation." 2021. [Online]. Available: http://gaofen-challenge.com
[64] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," 2017, *arXiv:1711.05101*.
[65] M. C. Chiu and T. M. Chen, "Applying data augmentation and mask R-CNN-based instance segmentation method for mixed-type wafer maps defect patterns classification," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 4, pp. 455–463, Nov. 2021.
[66] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft. Comput.*, vol. 70, pp. 41–65, Sep. 2018.

Chuang Yu (Member, IEEE) received the B.S. degree in network engineering from Hainan University, Haikou, China, in 2020. He is currently working toward the M.S. degree in pattern recognition and intelligent system at the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China.

His current research interests include cross-spectral image patch matching, infrared small target detection, and image segmentation.

Yunpeng Liu (Member, IEEE) received the Ph.D. degree in pattern recognition and machine intelligence from the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, in 2010.

He is currently a Professor with the Shenyang Institute of Automation, Chinese Academy of Sciences. His current research interests include cross-spectral image patch matching, image segmentation, infrared small target detection, small target tracking, and recognition based on the Riemannian manifold.

Xin Xia received the B.S. degree in network engineering from Hainan University, Haikou, China, in 2020. He is currently working toward the M.S. degree in pattern recognition and intelligent system at the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China.

His current research interests include computer version and image processing.

Deyan Lan received the B.S. degree in electronic information engineering from Tianjin University, Tianjin, China, in 2009, and the M.S. degree in signal and information processing from the University of Science and Technology of China, Hefei, China, in 2013.

Since 2017, he has been with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. His current research interests include machine learning and image processing.

Xin Liu received the B.S. degree in automation from the Hebei University of Technology, Tianjin, China, in 2016, and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2019.

Since 2019, he has been with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. His current research interests include deep learning, machine learning and image processing.

Shuhang Wu received the B.S. degree in automation from the Shandong University of Automation, Shandong, China, in 2014, and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2017.

Since 2017, she has been with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. Her current research interests include deep learning and computer vision.