Decoding the Partial Pretrained Networks for Sea-Ice Segmentation of 2021 Gaofen Challenge

Jian Kang, *Member, IEEE*, Fengyu Tong, Xiang Ding, Sijiang Li, Ruoxin Zhu, Yan Huang, *Member, IEEE*, Yusheng Xu, *Member, IEEE*, and Ruben Fernandez-Beltran, *Senior Member, IEEE*

***Abstract*—Sea-ice segmentation is of great importance for environmental research, ship navigation, and ice hazard forecasting. Remote sensing (RS) images have been a unique data source for rapid and large-scale sea-ice monitoring. The 2021 Gaofen Challenge has offered a track of sea-ice segmentation based on optical RS images. For the initial competition, our team ranked 3rd place (`deepjoker`) in the accuracy leaderboard and the solution has been the most efficient algorithm to achieve a segmentation score above 97.79%. In this article, we briefly introduce our three strategies of the achievement including: 1) decoding the partial pretrained networks which can simultaneously capture the complex boundaries of sea ices and decrease the computational cost without the performance drop; 2) employing the classwise Dice loss for solving the gradient vanishing problem when most ground-truth maps are backgrounds; and 3) replacing the commonly exploited decoder with the one proposed by Silva *et al.* (2021). The main contributions are twofold: 1) an efficient and effective sea-ice segmentation method is proposed and 2) the gradient vanishing problem of binary Dice loss is investigated under some scenarios and solved by introducing its classwise version. Comparison and ablation experiments demonstrate the effectiveness of the proposed method with respect to other commonly adopted deep segmentation models.**

***Index Terms*—Gaofen (GF) challenge, gradient vanishing problem, sea-ice segmentation, semantic segmentation.**

I. Introduction

As THE effect of global warming, amounts of sea ices have disappeared in the past decades. Although the decreased sea-ice cover has led to the opening of new pathways for shipping, isolated and floating sea ices are still potential risks and hazards to the international shipping through those areas [1]. Compared to other sources of sea-ice images, such as phones, the nadir-looking from satellites has offered unique way for rapid and large-scale sea-ice monitoring through the acquired remote sensing (RS) images, as shown in Fig. 1. Therefore, segmenting sea ices from RS images is of great importance for ice hazard forecasting, ship navigation, environmental research, and other related topics [2].

The sea-ice segmentation aims at pixel-wisely labeling the sea-ice area with 1 and the background area with 0, which belongs to the binary segmentation problem. Conventional methods for segmenting sea ices are mostly model-driven, such as graph cut segmentation [3] on the back-scattering intensities from synthetic aperture radar (SAR) images [4], Markov random field segmentation based on image textures [5], and snake segmentation on detected ice pixels [6]. Although those methods have demonstrated prominent results for segmenting sea ices, they cannot effectively preserve the accuracy and efficiency as latest satellites provide more fine-grained and huge-volume RS images.

More recently, the fast development of deep learning methods, such as convolutional neural networks (CNNs), has significantly promoted the state of the art performances of RS imagery interpretation including sea-ice segmentation [7]–[17]. Based on CNNs, the key point for extracting sea ices is learning to discriminate pixels from sea ices and the background. The technical solution for solving such a binary segmentation problem is also general to other related tasks, such as building segmentation, cloud detection, and road extraction [18]–[22]. Commonly, the CNN models applied for the segmentation task are consisted of three parts: 1) CNN architectures, which decide the overall network construction; 2) encoders, which hierarchically encode multiscale features from the input images; 3) decoders, which decode the multiscale features into binary masks. For architectures, the most widely exploited ones include fully convolutional networks (FCN) [23], SegNet [24], UNet [25], DeepLab series [26], [27], etc. Researchers favor to choose the networks which achieve the state of the art performances on the ImageNet dataset [28], such as ResNets [29] and EfficientNets [30], as the encoders within segmentation architectures. To gradually increase the spatial dimensions of the encoded features, the upsampling operation and convolutional blocks are adopted in the design of decoder structures. In order to optimize the CNN

Manuscript received March 5, 2022; revised May 23, 2022; accepted June 1, 2022. Date of publication June 10, 2022; date of current version June 15, 2022. This work was supported in part by the Jiangsu Province Science Foundation for Youths under Grant BK20210707 and in part by the National Natural Science Foundation of China under Grant 62101371. *(Corresponding author: Ruoxin Zhu.)*

Jian Kang, Fengyu Tong, Xiang Ding, and Sijiang Li are with the School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: kangjian_1991@outlook.com; iamfytong@163.com; d1554758569@163.com; 878976631@qq.com).

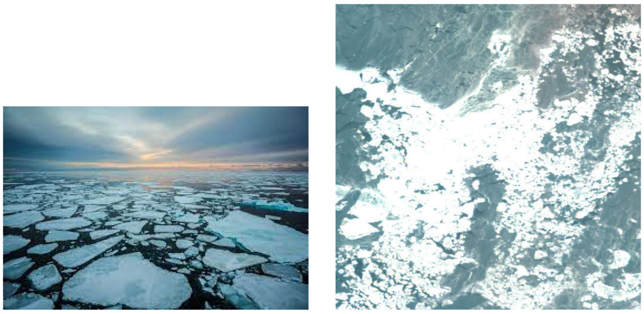
Ruoxin Zhu is with the State Key Laboratory of Geo-Information Engineering, Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China (e-mail: ruoxin.zhu@tum.de).

Yan Huang is with the State Key Lab of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 211100, China, and also with the Purple Mountain Laboratory, Nanjing 211100, China (e-mail: yellowstone0636@hotmail.com).

Yusheng Xu is with the Photogrammetry and Remote Sensing, Technical University of Munich, 80333 Munich, Germany, and also with the College of Surveying and Geo-Informatics, Tongji University, 200092 Shanghai, China (e-mail: yusheng.xu@tum.de).

Ruben Fernandez-Beltran is with the Department of Computer Science and Systems, University of Murcia, 30100 Murcia, Spain (e-mail: rufernan@uji.es).

Digital Object Identifier 10.1109/JSTARS.2022.3180558



(a) (b)

Fig. 1. Sea ices observed from different sensors. (a) Phones (Credit: Google) and (b) satellites.

models for binary segmentation, sigmoid activation is applied on the last layer and losses such as binary cross entropy (BCE), Dice, or their combined version, are commonly exploited for learning the CNN parameters.

The above-mentioned strategies are widely chosen for solving binary segmentation tasks in the framework of deep learning. However, for some cases in RS, they are not optimal and can lead to the *gradient vanishing problem*, which will be explained in Section III-C with details. In this article, we propose an efficient deep learning based binary segmentation method and uncover such problem in some specific scenarios. As a case study for validating the proposed method, our solution achieves the 3rd place (team name: `deepjoker`) during the first round of sea-ice segmentation task in 2021 Gaofen (GF) Challenge[1] with 97.79% frequency weighted intersection over union (FWIoU). Specifically, the following three strategies are carefully designed for this challenge.

1) Differently to other methods, we decode the partial pretrained networks rather than the full ones which are commonly exploited.
2) We replace the last layer activated by the sigmoid function with the one activated by the softmax function and adopt the classwise Dice loss rather than the normal Dice loss for binary segmentation.
3) We adopt the CNN block proposed in [31] as the decoder in replace of the vanilla one.

To this end, the main contributions of this article can be summarized as follows.

1) We propose a binary segmentation method which efficiently decodes the partial pretrained networks and achieve the 3rd place during the first round of sea-ice segmentation task in 2021 Gaofen Challenge.
2) We uncover a scenario of binary segmentation in RS where the sigmoid function and Dice loss are not suitable and the gradient vanishing problem happens. Moreover, a simple and effective solution is proposed for tackling it.

The rest of this article is organized as follows. Section II presents some related work from the perspectives of semantic segmentation and binary segmentation in RS. Section III introduces the overall method proposed for sea-ice segmentation in the 2021 Gaofen Challenge. Section IV demonstrates the conducted experiments and analyzes the associated results. Finally, Section V concludes this article.

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

II. Related Work

A. Semantic Segmentation

Semantic segmentation can be formulated as classifying pixels with object categories [32]. In RS, those categories are generally referred to land-use or land-cover classes. Recent advances of RS image segmentation techniques have been achieved with the development of deep learning methods. An end-to-end semantic segmentation network is proposed in [33], where detecting semantically meaningful boundaries is involved for refining segmentation results. Kampffmeyer *et al.* [34] investigated the class imbalance that existed in RS segmentation task and proposed a novel method for identifying small geospatial objects. Li *et al.* proposed [35] a novel UNet-like architecture with DownBlock and UpBlock structures for sea–land segmentation. By simultaneously training the segmentation and edge detection networks, Cheng *et al.* [36] proposed an edge-aware CNN architecture for sea–land segmentation in RS harbor images. Chen *et al.* [37] proposed symmetrical dense-shortcut mechanism for semantic segmentation in very high resolution RS images. Audebert *et al.* [38] investigated different fusion strategies for semantic segmentation based on multimodal and multiscale RS images. By integrating several auxiliary tasks, such as input image reconstruction and distance transform inference, Diakogiannis *et al.* [39] proposed a novel CNN architecture, i.e., ResUNet-a, and a loss function based on the Dice loss for segmenting high-resolution aerial images. Sun *et al.* introduced a boundary-aware segmentation method for high-resolution RS images when a small amount of labels are available [40]. By adaptively capturing global correlations of space, channel, and category, Niu *et al.* [41] proposed a hybrid multiple attention network for segmenting high-resolution aerial images. For boosting the segmentation performance on SAR images, Shi *et al.* [42] created a well-annotated multimodality RS dataset consisted of GF-3 and Google Earth optical images, along with object-level vector data.

B. Binary Segmentation in RS

Differently to semantic segmentation, binary segmentation aims at discriminating one specific land-use or land-cover class from RS images, such as buildings and roads. For example, an FCN-based segmentation CNN architecture with the multiscale feature aggregation strategy is proposed in [43] for learning fine-grained building contours. By optimizing the selective spatial pyramid dilated network with a L-shape weighting loss, Jing *et al.* [20] proposed a novel method for characterizing building footprints in SAR images. To improve the performance of building boundaries, Ziaee *et al.* [44] modified the conventional Pix2Pix framework with the proposed two generators which generate localization and boundary features. Shamsolmoali *et al.* [45] proposed a novel adversarial spatial pyramid network for applying structural domain adaptation for synthetic image

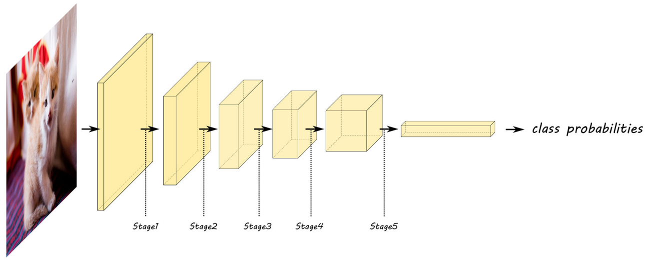


Fig. 2. Typical structure of the pretrained CNN networks.

generation and road segmentation. Similarly, to solve the label-deficient problem in road segmentation, Zhang *et al.* [46] investigated a stagewise unsupervised domain adaptation method based on adversarial self-training scheme to segment roads in high-resolution RS images. Mohajerani *et al.* [47] proposed a filtered Jaccard loss function for better segmenting foreground objects when they are absent in RS images. He *et al.* [22] proposed a deformable context feature pyramid module which can effectively adapting multiscale features extracted from RS images and a boundary-weighted loss function for cloud detection.

III. Methodology

The proposed method contains two main novel strategies for achieving efficient and effective sea-ice segmentation including: 1) efficiently decoding the partial pretrained networks and 2) replacing the binary Dice loss with its classwise version for avoiding the gradient vanishing problem in some scenarios. In the following, we describe all these components in detail.

A. Pretrained CNN Backbone

As pretrained CNN backbones have sufficient features to be reused in the challenge, we adopt the transfer learning strategy which exploits state of the art pretrained CNN backbones on ImageNet as the encoder subnetworks [48]. For the later ensemble learning purpose, Res2Net50 [49], EfficientNet-B4 [30], and dual path networks (DPN) [50] are exploited as the backbones. As shown in Fig. 2, these networks can be all represented in a common structure, which is composed of several feature encoding stages. The spatial dimensions of the output features from the current stage are all half of those from one stage earlier. Generally, the early stages often capture detailed structure information of the input images while the last several stages encode high-level semantic information with larger receptive fields.

B. Proposed Segmentation Architecture

According to the provided challenge dataset present in Fig. 8, it can be observed that precisely segmenting the sea ices is challenging, since the spatial coverage of them is complicated. To effectively learn their detailed boundary structures, we decide to reuse the early-stage features from the pretrained networks, as they capture the detailed spatial information of the input images. Differently with other methods, the 5th encoding blocks of the pretrained networks are omitted here, with the consideration of

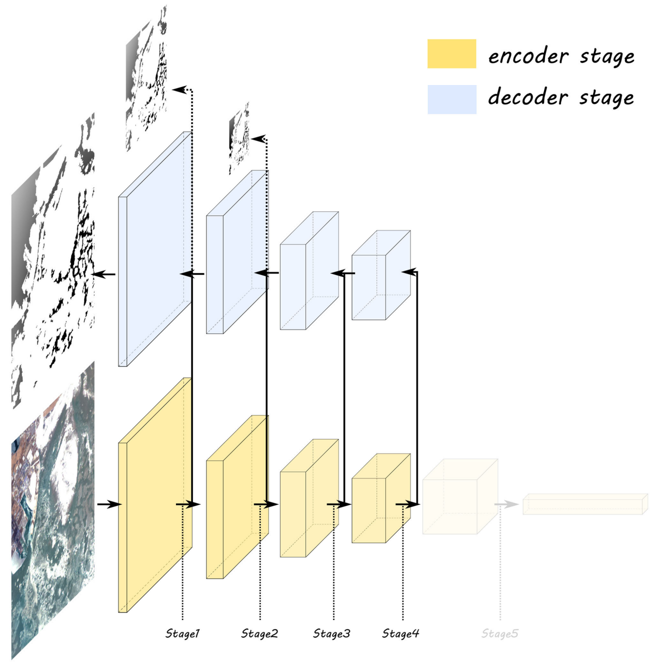


Fig. 3. Proposed segmentation architecture for sea-ice segmentation. Differently to other methods utilizing the pretrained CNN networks, we remove the fifth encoding stage for achieving a relatively light-weight segmentation network.

the balance between the computational cost and segmentation performance. For one convolutional kernel of the spatial size $K \times K$, its convolutional operation on the features $\mathbf{F}$ with the size of $C \times H \times W$ is of the computational cost

$$\mathcal{O}(K \times K \times C \times H \times W). \tag{1}$$

With the stage increasing, the channel numbers of feature maps also become larger. In addition, the feature maps of the 5th stage generally capture high-level semantics of the input images, while they cannot be of great help for localizing the boundary structures of sea ices at pixel-level. Taken these into account, we directly decode the features output from the 4th stage in order to achieve a relatively light-weight CNN with lower computational cost than decoding the full pretrained networks. As illustrated in Fig. 3, a segmentation network with a U-shape structure is proposed which can simultaneously speed up the inference and preserve the detailed spatial information of sea ices. The output channel numbers of four decoders are $\{128, 64, 32, 16\}$ in the top-down direction. Moreover, we exploit the *deep supervision* strategy [51] to further improve the multilevel feature discrimination through the supervision on the latent representations within the networks. To achieve this, we additionally add two more segmentation heads on top of second and third last decoder outputs.

Conventional decoders for segmentation are usually composed of the stacked Conv[3×3]-BN-ReLu modules. To improve the efficiency, we adopt the convolutional block proposed in [31], which exploits the residual learning scheme with the depthwise and pointwise convolutions involved. In addition, the concurrent spatial and channel squeeze and excitation (SCSE) block [52], [53] (as shown in Fig. 4) is integrated to refine the

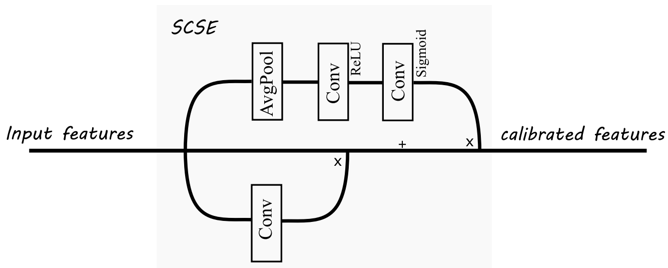


Fig. 4. Concurrent SCSE block exploited for calibrating the input features along the spatial and channel dimensions.

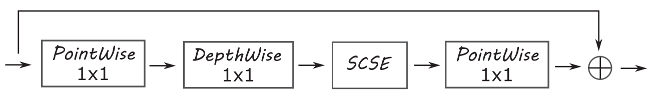


Fig. 5. Adopted CNN block [31] in the decoder.

features both along the spatial and channel dimensions. The overall structure of the convolutional block is illustrated in Fig. 5.

C. Loss Function

For the binary segmentation task, commonly exploited losses in the literature are BCE, Dice, or their combinations. Compared to the pixel-based loss, i.e., BCE, Dice loss is more emphasized on learning precise region predictions with the following formula:

$$L_{\text{Dice}} = 1 - \frac{2\sum_i p_i y_i + \epsilon}{\sum_i y_i + \sum_i p_i + \epsilon} \tag{2}$$

where y_i and p_i denote the pixel-wise ground-truth and the predicted probability value output from the sigmoid activation, respectively, and ϵ is often set as 1×10^{-6} to avoid unstable division. For most cases, Dice loss can be optimized to align the prediction and the associated ground-truth maps. However, as pointed out in [47], there exists a defect when there are no class 1 in ground-truth maps, i.e., $y_i = 0, \forall i$. In such case, the dice loss is rewritten as

$$L_{\text{Dice}} = 1 - \frac{\epsilon}{\sum_i p_i + \epsilon}. \tag{3}$$

Considering a toy case shown in Fig. 6, the ground-truth map is of the size with 3×3 pixels and all of them are 0, i.e., $y_i = 0$. Let us assume two kinds of predictions: 1) each pixel is with the probability of 0.01 for class 0 and 2) each pixel is with the probability of 0.9 for class 0. Clearly, the second prediction is correct, since the probabilities of class 0 are all above 0.5. However, the Dice losses for the two cases are both approximated to 1. It indicates the networks cannot be optimized when the whole ground-truth map is all 0. For this challenge, since most training images are without sea ices, there will be no effects on the network optimization when those images are fed. To solve this issue, we propose to exploit a simple and effective loss, which extends L_{Dice} in a classwise manner

$$L_{\text{Cls-Dice}} = 1 - \frac{1}{2} \sum_{c \in \{0,1\}} \frac{2\sum_i p_i^c y_i^c + \epsilon}{\sum_i y_i^c + \sum_i p_i^c + \epsilon} \tag{4}$$

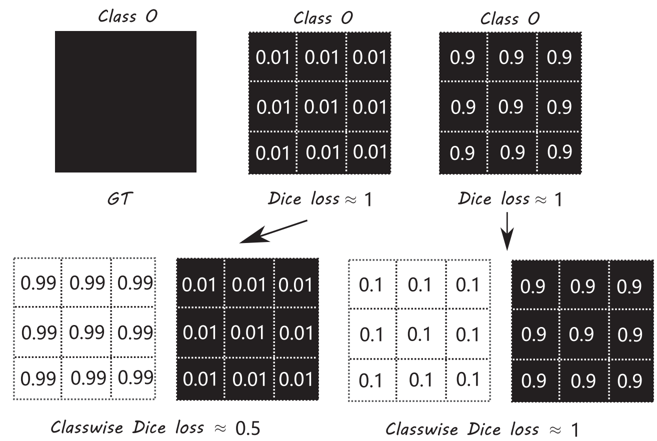


Fig. 6. Toy example for explaining the signification difference of loss values between the Dice and classwise Dice losses when the ground-truth map is background.

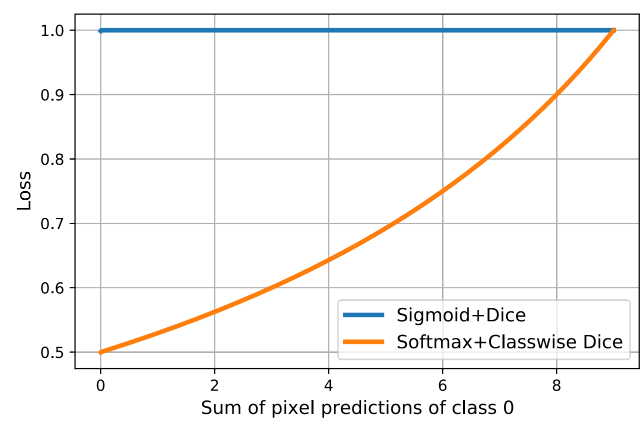


Fig. 7. Loss values versus sum of pixel predictions of class 0 in the toy example.

where c indicates the class label. It is important to note that the predictions p_i^c are activated by softmax function, although it is binary. Take the same toy example as above, the classwise Dice losses for two predictions are different, i.e., 0.5 and 1 for both 0.01 and 0.9 probability of class 0, respectively. Moreover, for the toy example, we plot the Dice and classwide Dice losses with respect to the sum of prediction probability of class 0 in Fig. 7. It can be observed that no matter what predictions of networks make, the loss values are all around 1 when ground-truth maps are all 0. In a comparison, classwise Dice loss gradually increases when the predictions are becoming worse. Such gradient vanishing problem can be further observed from the following formulas:

$$\frac{\partial L_{\text{Dice}}}{\partial p_i} = \frac{\epsilon}{(p_i + \epsilon)^2} \quad \forall y_i = 0 \tag{5}$$

$$\frac{\partial L_{\text{Cls-Dice}}}{\partial p_i^c} = \frac{1}{2}\left(\frac{\epsilon}{(p_i^c + \epsilon)^2} - \frac{2+\epsilon}{(2 - p_i^c + \epsilon)^2}\right) \quad \forall y_i^c = 0. \tag{6}$$

Since ϵ is almost 0, $\partial L_{\text{Dice}}/\partial p_i$ will be vanished when all y_i are 0. Differently, for $\partial L_{\text{Cls-Dice}}/\partial p_i^c$, the term $2 + \epsilon/(2 -$

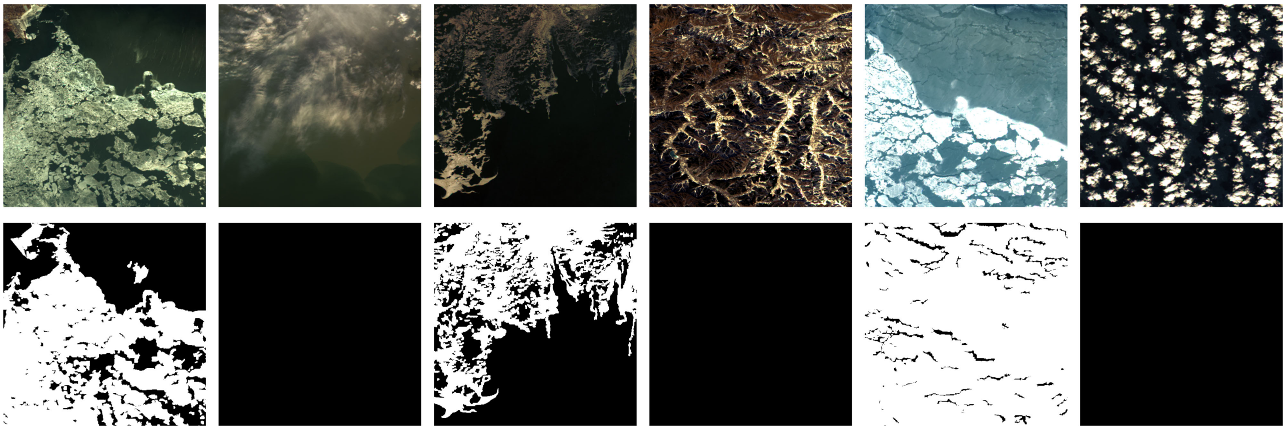


Fig. 8. Some released images with the associated ground-truth maps during the first round of challenge.

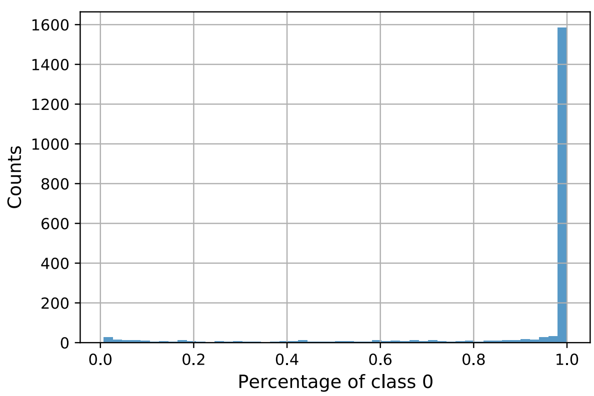


Fig. 9. Image counts under different pixel number ratios of class 0.

$p_i^c + \epsilon)^2$ can avoid such a issue. To this end, the joint loss containing cross entropy (CE), classwise Dice, and boundary (BD) losses [54] is exploited for this contest.

IV. Experiments

A. Experimental Setup

The organizers provided 1500 RGB optical RS images with spatial sizes ranging from 512×512 to 2048×2048 for training the sea-ice segmentation network and 1000 images for validating the inference performance during the first round. Some provided samples are illustrated in Fig. 8, along with the labeled segments. It can be observed that accurately segmenting the sea-ice areas is challenging due to the isolated and complex-structured blocks of sea ices with different sizes. Moreover, as shown in Fig. 9, we calculate the histograms of the ratio between the number of background pixels and the total number of pixels in the scene and observe that a large amount of training images are without sea ices. Such a specific characteristic of the dataset will lead to the gradient vanishing problem when the Dice loss is applied for training the networks. It will be thoroughly described in Section III-C.

We randomly split the provided images into the training and test sets with the ratio of $9:1$. As introduced above, Res2Net50 [49], EfficientNet-B4 [30], and DPN [50] are employed as CNN backbones and their results are ensembled through a majority voting strategy. For data augmentation, HorizontalFlip and RandomRotate90 are adopted and the images are normalized with the mean and standard deviation values of 127.5 and 31.875, respectively. Stochastic gradient descent (SGD) optimizer is utilized to train the segmentation model with a initial learning rate of 5×10^{-3} and progressively adjusted by a polynomial scheduler. The networks are trained for a total number of 200 epochs with a minibatch size of 8. All the experiments are implemented in PyTorch [55] and carried out on an NVIDIA RTX3090 GPU. To validate the effectiveness of the proposed method, we compare it with several widely-adopted methods for segmentation including: 1) UNet [25]; 2) PSPNet [56]; 3) feature pyramid network (FPN) [57]; 4) LinkNet [58]; 5) DeepLabV3 [26]; and 6) DeepLabV3+ [27]. The evaluation metrics are frequency FWIoU, intersection over union (IoU), and overall accuracy (OA). The detailed calculations of them can be found in [59].

TABLE I
Evaluation of the Proposed Method Compared With Several Commonly-Exploited Segmentation Methods Under Different Metrics (%)

Method	FWIoU	IoU	OA
UNet[25]	98.25	92.32	99.10
PSPNet[56]	97.39	88.69	98.65
FPN [57]	98.07	91.57	99.01
LinkNet [58]	97.96	91.10	98.95
DeepLabV3 [26]	97.87	90.70	98.90
DeepLabV3+ [27]	98.02	91.32	98.98
Proposed	**98.75**	**94.51**	**99.37**

B. Experimental Results

1) Comparison to State of the Art Methods: Table I demonstrates the segmentation results on the test set manually split from the data provided during the first contest round. It can be observed that the proposed method can outperform UNet by a margin of 0.5% for the metric FWIoU and achieve the best result

TABLE II
SEGMENTATION ACCURACY AND EXECUTION TIME EVALUATED ON TOP-10 TEAMS IN THE LEADERBOARD DURING THE FIRST ROUND (% AND SECOND)

Method	FWIoU	Time
glotwo	97.9354	6033
分割的都对	97.9246	933
deepjoker	97.7907	136
宝略科技	97.7360	367
追风少年冲冲冲	97.7190	800
WIDEA	97.6998	2456
天竺鼠车车	97.6956	607
BIT_505	97.6534	3495
MDISL-lab-team000	97.6137	4553
冲冲冲呀	97.5757	5982

TABLE III
SEGMENTATION RESULTS EVALUATED ON THE PROPOSED METHOD WHEN DIFFERENT NUMBERS OF ENCODER STAGES ARE DECODED (%)

Backbone	#Stage		FWIoU	IoU	OA	#Para	Time
	4	5					
Res2Net50	✓		98.43	93.07	99.20	31.8M	4.41
		✓	98.32	92.59	99.14	56.6M	5.68
EfficientNet-B4	✓		98.57	93.70	99.27	18M	5.09
		✓	98.29	92.46	99.12	19.9M	6.68
DPN	✓		98.60	93.85	99.29	15.6M	3.12
		✓	98.01	91.30	98.98	21.5M	8.07

among all the considered methods. To visualize some predicted results of sea ices, we select the image examples associated with their ground-truth maps shown in Fig. 10. As illustrated in Fig. 10, the obtained sea-ice segments are compared among all the considered methods. The proposed method can achieve more accurate boundary predictions in all the examples than others. The plausible reason is that the boundary-loss term is exploited for emphasizing the boundary areas. In addition, the proposed method is compared with other teams' solutions during the initial phase and their results are illustrated in Table II. Although all the considered methods can achieve comparable results with slightly different FWIoU, the proposed method is ranked at the third place and the execution time is much shorter than the other methods, which can validate its efficiency and effectiveness for the sea-ice segmentation.

2) Ablation Study: Effect of number of encoder stages: In the proposed method, we only decode the extracted features until the fourth stage of the encoder, since the features from the last stage mainly capture the semantics of whole images and may not be profitable enough for learning complex sea-ice boundaries. For evaluation, the different number of stages within the three pretrained networks are decoded and the associated segmentation performances are illustrated in Table III. It can be observed that decoding the features from four stages can achieve higher segmentation accuracy than exploiting all the stage features. Moreover, by removing the fifth stage, more lightweight networks and faster image inference can be achieved.

Effect of $L_{\mathrm{Cls-Dice}}$: As analyzed above, the gradient vanishing problem can be avoided when the softmax activation and classwise Dice loss are utilized. To verify this, we replace them with the sigmoid activation and Dice loss, and calculate their

TABLE IV
SEGMENTATION PERFORMANCE COMPARISON BETWEEN THE DICE AND CLASSWISE DICE LOSSES (%)

Activation	Loss term	FWIoU	IoU	OA
Sigmoid	L_{Dice}	98.35	92.77	99.16
Softmax	$L_{\mathrm{Cls-Dice}}$	**98.75**	**94.51**	**99.37**

TABLE V
SEGMENTATION PERFORMANCE COMPARISON WHEN THE PROPOSED METHOD WITH AND WITHOUT THE DEEP SUPERVISION STRATEGY (%)

Deep supervision		FWIoU	IoU	OA
w/	w/o			
	✓	98.51	93.43	99.24
✓		**98.75**	**94.51**	**99.37**

TABLE VI
SEGMENTATION PERFORMANCE COMPARISON BETWEEN DIFFERENT DECODERS (%)

Decoder	FWIoU	IoU	OA
Vanilla [25]	98.54	93.55	99.25
Proposed in [31]	**98.75**	**94.51**	**99.37**

TABLE VII
SEGMENTATION PERFORMANCE COMPARISON BETWEEN DIFFERENT LOSS TERMS (%)

Loss	Backbone	FWIoU	IoU	OA
CE	EfficientNet-B4	98.11	91.74	99.03
CE+Cls-Dice	EfficientNet-B4	98.41	93.00	99.10
CE+Cls-Dice+BD	EfficientNet-B4	98.57	93.07	99.20

segmentation results in Table IV. Since most training images are backgrounds, the Dice loss cannot make any affords in the network optimization. Thus, the associated segmentation accuracy is lower than the proposed method.

Effect of deep supervision: Moreover, the ablation study regards to deep supervision is also conducted. As shown from Table V, with the constraint of deep supervision, the proposed method can achieve better segmentation performance than the one without it. By imposing the accurate segmentation predictions from multiscale output features, it can lead to more prominent results obtained from the last layer.

Effect of the decoder: We also compare segmentation performances when different decoders are adopted. Commonly, most work exploit the vanilla decoder proposed in [25]. The associated result is illustrated in Table VI. It can be seen that the decoder proposed in [31] can better capture the complex structures of sea ices. One plausible reason is that the residual design and the SCSE block applied in it can emphasize the learning of contributed features and inhibit others at the same time.

Effect of the joint loss: The joint loss exploited in the proposed method is composed of CE, Cls-Dice, and BD loss terms. In order to validate the effectiveness, we conduct ablation study on the loss configuration. As illustrated in Table VII, the best segmentation accuracy can be achieved by exploiting all of them. Compared to CE, Dice loss can effectively deal with the class-imbalance problem. As CE and Cls-Dice losses are region-based

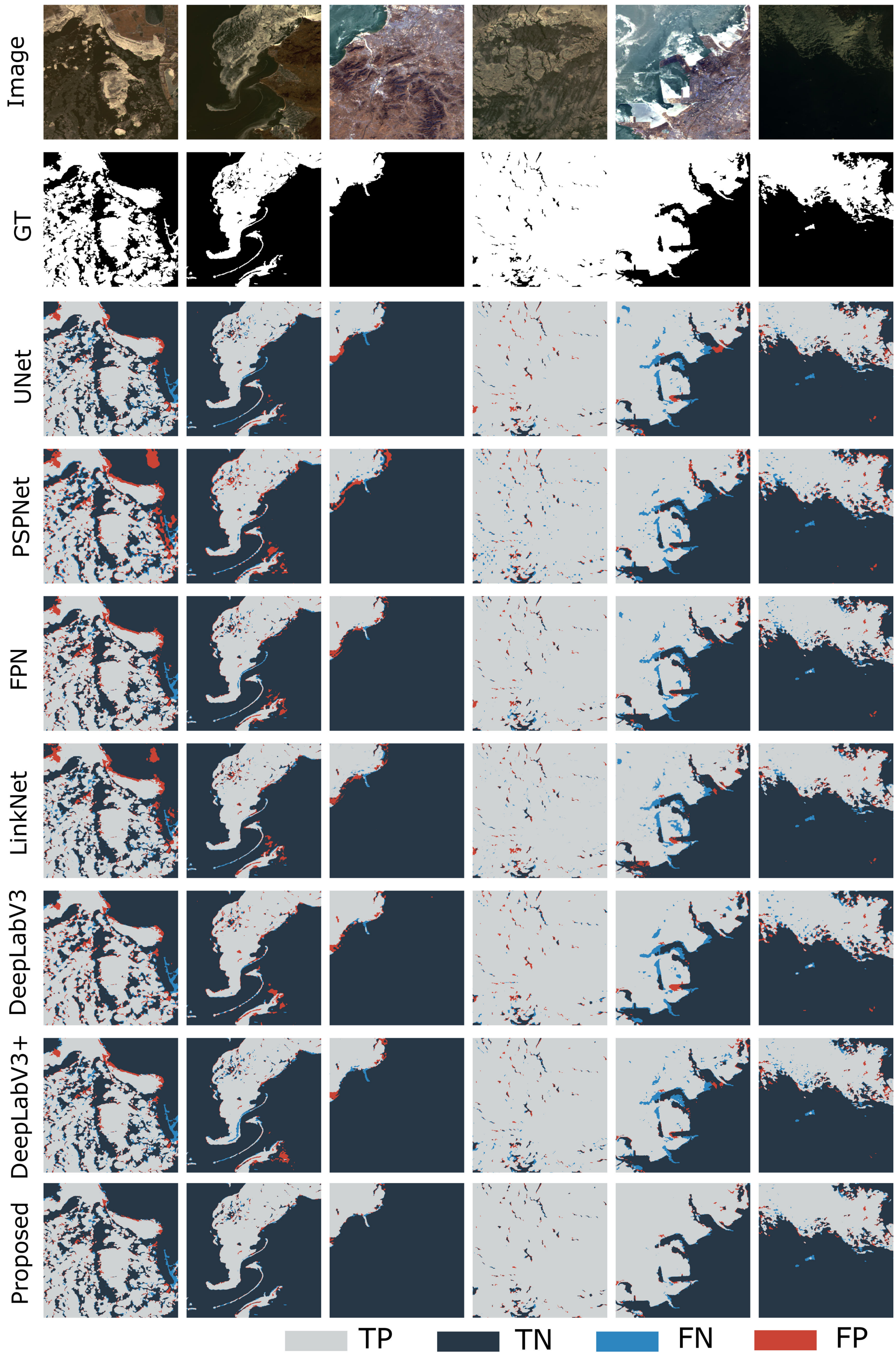


Fig. 10. Sea-ice predictions of the provided images based on the considered methods.

segmentation losses, BD loss is more emphasized on precisely learning the boundaries of segments, especially for extracting sea-ice areas with complex structures.

3) Discussion: The extensive experiments carried out demonstrate the effectiveness of the proposed method and we also achieve a high rank on the enclosed test set during the first round of the contest. For the binary segmentation of objects with complex structures, the proposed method demonstrates the superior performance compared to other widely-adopted segmentation methods. Thus, the method also has the potential to be applied for other RS objects like sea ices, such as clouds. Moreover, we point out one extreme case that the Dice loss makes no contributions for the network optimization due to the gradient vanishing problem. To avoid such an issue, we propose a simple and effective solution which extends the Dice loss into a classwise manner. By analyzing its gradient, we can see that the gradient vanishing problem can be avoided. Thus, for the datasets where large number of images are backgrounds, the classwise Dice loss is suggested to be adopted. Last but not least, the proposed method offers a strong baseline for the binary segmentation challenge in RS.

V. Conclusion

In this article, we propose a method for sea-ice segmentation in 2021 GF Challenge, which achieves a FWIoU score of 97.79% and the 3rd place in the leaderboard during the first round of challenge. The proposed method contains three carefully designed strategies: 1) decoding the partial pretrained networks which can simultaneously capture the complex boundaries of sea ices and decrease the computational cost without performance drop; 2) employing the classwise Dice loss for solving the gradient vanishing problem when most ground-truth maps are backgrounds; and 3) replacing the commonly exploited decoder with the one proposed in [31]. Compared to other commonly-exploited methods for binary segmentation, the proposed method demonstrates significant performance improvement on the challenge dataset, which can also be served as a baseline method for other related binary segmentation tasks in RS.

References

[1] "Arctic open-water season grows," *Nature*, vol. 527, pp. 10–11, 2015. [Online]. Available: https://www.nature.com/articles/527010d#citeas

[2] A. S. Nagi, D. Kumar, D. Sola, and K. A. Scott, "RUF: Effective sea ice floe segmentation using end-to-end RES-UNET-CRF with dual loss," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2460.

[3] M. B. Salah, A. Mitiche, and I. B. Ayed, "Multiregion image segmentation by parametric kernel graph cuts," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 545–557, Feb. 2011.

[4] B. Hwang *et al.*, "A practical algorithm for the retrieval of floe size distribution of arctic sea ice from high-resolution satellite synthetic aperture radar imagery," *Elementa: Sci. Anthropocene*, vol. 5, 2017. [Online]. Available: https://online.ucpress.edu/elementa/article/

[5] D. A. Clausi and B. Yue, "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 215–228, Jan. 2004.

[6] Q. Zhang and R. Skjetne, "Image processing for identification of sea-ice floes and the floe size distributions," *IEEE Trans. Geosci. remote Sens.*, vol. 53, no. 5, pp. 2913–2924, May 2015.

[7] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[8] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[9] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, 2019.

[10] B. Dowden, O. De Silva, W. Huang, and D. Oldford, "Sea ice classification via deep neural network semantic segmentation," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11879–11888, May 2021.

[11] Y. Han, Y. Liu, Z. Hong, Y. Zhang, S. Yang, and J. Wang, "Sea ice image classification based on heterogeneous data fusion and deep learning," *Remote Sens.*, vol. 13, no. 4, p. 592, 2021. [Online]. Available: https://doi.org/10.3390/rs13040592

[12] H. Boulze, A. Korosov, and J. Brajard, "Classification of sea ice types in sentinel-1 SAR data using convolutional neural networks," *Remote Sens.*, vol. 12, no. 13, 2020, Art. no. 2165.

[13] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[14] J. Kang, R. Fernandez-Beltran, Z. Wang, X. Sun, J. Ni, and A. Plaza, "Rotation-invariant deep embedding for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5509713.

[15] X. Sun, Z. Wang, Y. Sun, W. Diao, Y. Zhang, and K. Fu, "Air-sarship-1.0: High-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019.

[16] X. Qiu, Z. Jiao, and L. Peng, "Sarmv3d-1.0: Synthetic aperture radar microwave vision 3d imaging dataset," *J. Radars*, vol. 10, no. 4, pp. 485–498, 2021.

[17] S. Fu, F. Xu, and Y.-Q. Jin, "Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks," *Sci. China Inf. Sci*, vol. 64, pp. 122301, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11432-020-3077-5#citeas

[18] J. Kang, Z. Wang, R. Zhu, X. Sun, R. Fernandez-Beltran, and A. Plaza, "Picoco: Pixelwise contrast and consistency learning for semisupervised building footprint segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10548–10559, 2021.

[19] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 161–173, 2021.

[20] H. Jing, X. Sun, Z. Wang, K. Chen, W. Diao, and K. Fu, "Fine building segmentation in high-resolution SAR images via selective pyramid dilated network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6608–6623, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9416747

[21] J. Kang, R. Fernandez-Beltran, X. Sun, J. Ni, and A. Plaza, "Deep learning-based building footprint extraction with missing annotations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 3002805.

[22] Q. He, X. Sun, Z. Yan, and K. Fu, "DABNET: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5601216.

[23] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. image Comput. Comput.- Assist. Intervention*, 2015, pp. 234–241.

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

[27] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[28] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

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

[30] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. PMLR*, 2019, pp. 6105–6114.

[31] J. L. Silva, M. N. Menezes, T. Rodrigues, B. Silva, F. J. Pinto, and A. L. Oliveira, "Encoder-decoder architectures for clinically relevant coronary artery segmentation," 2021, *arXiv:2106.11447*.

[32] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[33] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 158–172, 2018.

[34] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 1–9.

[35] R. Li *et al.*, "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018.

[36] D. Cheng, G. Meng, S. Xiang, and C. Pan, "FusionNet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5769–5783, Dec. 2017.

[37] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1633–1644, May 2018.

[38] N. Audebert, B. Le Saux, and S. Lefèvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 140, pp. 20–32, 2018.

[39] A. Fid, B. Fw, A. Pc, and W. C. Chen, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data - sciencedirect," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 94–114, 2020.

[40] X. Sun, A. Shi, H. Huang, and H. Mayer, "Bas ^{4} net: Boundary-aware semi-supervised semantic segmentation network for very high resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5398–5413, 2020.

[41] R. Niu, X. Sun, Y. Tian, W. Diao, K. Chen, and K. Fu, "Hybrid multiple attention network for semantic segmentation in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art no. 5603018.

[42] X. Shi, S. Fu, J. Chen, F. Wang, and F. Xu, "Object-level semantic segmentation on the high-resolution Gaofen-3 fusar-map dataset," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3107–3119, 2021.

[43] S. Wei, S. Ji, and M. Lu, "Toward automatic building footprint delineation from aerial images using CNN and regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2178–2189, Mar. 2020.

[44] A. Ziaee, R. Dehbozorgi, and M. Döller, "A novel adaptive deep network for building footprint segmentation," 2021, *arXiv:2103.00286*.

[45] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, and J. Yang, "Road segmentation for remote sensing images using adversarial spatial pyramid networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4673–4688, Jun. 2021.

[46] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art no. 5609413.

[47] S. Mohajerani and P. Saeedi, "Cloud and cloud shadow segmentation for remote sensing imagery via filtered Jaccard loss function and parametric augmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4254–4266, Apr. 2021.

[48] Z. Huang, C. O. Dumitru, Z. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution SAR images with deep transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 107–111, Jan. 2021.

[49] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. H. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.

[50] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," *Adv. neural inf. process. syst.*, vol. 30, 2017, *arXiv:1707.01629*.

[51] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.

[52] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2018, pp. 421–429.

[53] J. Kang *et al.*, "DisoptNet: Distilling semantic knowledge from optical images for weather-independent building segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[54] A. Bokhovkin and E. Burnaev, "Boundary loss for remote sensing imagery semantic segmentation," in *Proc. Int. Symp. Neural Netw.*, 2019, pp. 388–401.

[55] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 8026–8037, 2019.

[56] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[57] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[58] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.

[59] I. Ulku and E. Akagunduz, "A survey on deep learning-based architectures for semantic segmentation on 2d images," *Appl. Artif. Intell.*, 2022, pp. 1–45, *arXiv:1912.10230*.

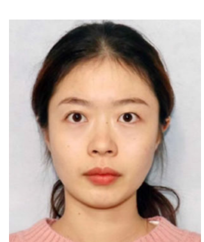
Jian Kang (Member, IEEE) received the B.S. and M.E. degrees in electronic engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2013 and 2015, respectively, and the Dr.-Ing. degree in signal processing for earth observation from Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

In August 2018, he was a Guest Researcher with the Institute of Computer Graphics and Vision (ICG), TU Graz, Graz, Austria. From 2019 to 2020, he was with the Faculty of Electrical Engineering and Computer Science, TU Berlin, Berlin, Germany. He is currently with the School of Electronic and Information Engineering, Soochow University, Suzhou, China. In particular, he is interested in intelligent SAR/InSAR data processing, and deep learning-based techniques for remote sensing image analysis. His research interests include signal processing and machine learning techniques, and their applications in remote sensing.

Dr. Kang obtained first place of the Best Student Paper Award in European Conference on Synthetic Aperture Radar (EUSAR) 2018, Aachen, Germany. His joint work was selected as one of the ten Student Paper Competition Finalists in Geoscience and Remote Sensing Symposium (IGARSS) 2020. He served as Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

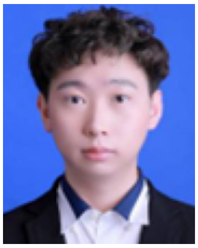
Fengyu Tong received the B.E. degree in optoelectronic information science and engineering from Nanjing Tech University, Nanjing, China, in 2021. He is currently working toward the M.E. degree in information and communication engineering with Electronic and Information Engineering School, Soochow University, Suzhou, China.

His research interests include deep learning algorithms for object detection in remote sensing data.



Xiang Ding received the B.E. degree in electronic engineering from the East China University of Science and Technology, Nanchang, China, in 2021. She is currently with the Electronic and Information Engineering School, Soochow University, Suzhou, China.

Her research interests include image denoising based on remote sensing data.



Sijiang Li received the B.Eng. degree in electronic engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2021. He is currently with the School of Electronic and Information Engineering, Soochow University, Suzhou, China.

His research interest includes remote sensing image processing based on depth learning.

Ruoxin Zhu received the B.E. and M.E. degrees in cartography and geographic information engineering from the Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China, in 2012 and 2015, respectively, and the Dr.-Ing. degree is cartography from the Chair of Cartography, Technical University of Munich (TUM), Munich, Germany, in 2020.

He is currently with the State Key Laboratory of Geo-Information Engineering, Xi'an Research Institute of Surveying and Mapping, Xi'an, China. His research interests include cartography and GIS. In particular, he is interested in social sensing and geostatistics based on spatio-temporal big data.

Yan Huang (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2013 and 2018, respectively.

From 2016 to 2017, he was a Visiting Ph.D. Student with the Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA, and with the Electrical and Systems Engineering Department, Washington University in St. Louis, St. Louis, MO, USA, from July 2017 to August 2018. He is currently an Associate Professor with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China. His research interests include machine learning, synthetic aperture radar, image processing, and remote sensing.

Yusheng Xu (Member, IEEE) was born in 1989. He received the B.S and M.E degrees in surveying and mapping from Tongji University, Shanghai, China, in 2011 and 2014, respectively, and the Ph.D. (Dr.-Ing., *summa cum laude*) degree in photogrammetry and the habilitation (venia legendi) in point cloud analysis from the Technical University of Munich (TUM), Munich, Germany, in 2019 and 2022, respectively.

From May 2022, he was with the College of Surveying and Geoinformatics in Tongji University. He has authored and coauthored more than 40 entries. His research interests include 3-D point cloud processing, image processing, and spaceborne photogrammetry.

Dr. Xu served as the Guest Editor or Editorial Board Member of several academic journals. He was also the Program Committee Member of International Society for Photogrammetry and Remote Sensing (ISPRS) Hannover Workshop 2017, the Scientific Committee Member of ISPRS Laser Scanning Workshop 2017 and 2019, and the Scientific Committee Member of ISPRS Workshop PIA2019+MRSS2019.

Ruben Fernandez-Beltran (Senior Member, IEEE) received the B.Sc. degree in computer science, the M.Sc. degree in intelligent systems, and the Ph.D. degree in computer science from the University Jaume I, Castellon de la Plana, Spain, in 2007, 2011, and 2016, respectively.

He is currently an Assisstant Professor with the Department of Computer Science and Systems, University of Murcia, Murcia, Spain. He has been a Visiting Researcher with the University of Bristol, Bristol, U.K., the University of Cáceres, Cáceres, Spain, Technische Universität Berlin, Berlin, Germany, and the Autonomous University of Mexico State, Toluca, NM, USA. His research interests include multimedia retrieval, spatio-spectral image analysis, pattern recognition techniques applied to image processing, and remote sensing.

Dr. Fernandez-Beltran was the recipient of the Outstanding Ph.D. Dissertation Award at Universitat Jaume I in 2017.