POAC'23
Glasgow, United Kingdom

Proceedings of the 27th International Conference on
Port and Ocean Engineering under Arctic Conditions
12-16 June 2023, Glasgow, United Kingdom

Study of semantic segmentation on sea ice image based on deep learning

Xiaodong Chen[1,2], Yunhan Ma[1,2], Shunying Ji[1,2]
1. Dalian University of Technology, Dalian, China
2. State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, Dalian, China

ABSTRACT

For the video images from the Arctic environment, it is difficult to recognize sea ice with properly due to the random shape of sea ice and complicated scenarios. To im prove the accuracy of ice recognition, the deep learning method is applied on the ima ge processing while the semantic segmentation is built on the Deeplab V3+ network. In this paper, the method of supervised training is adopted for semantic segmentation, which is all trained with self-made datasets, thus the weight model is obtained. In th e verification, in addition to human observation and comparison, the average intersecti on ratio is also introduced as an evaluation standard. The manually labeled results are used as a reference and compared with the segmentation output results. The average intersection ratio is calculated to evaluate the recognition results. From the results, th e recognition accuracy of the four recognition objects of sea ice, sky, ship and sea w ater background has reached the expected value, with an average intersection over uni on of 94.69%. It shows the feasibility and effectiveness of the deep convolution neura l network in the semantic segmentation task of sea ice images. In addition, the metho d is less affected by sea ice distribution, shooting brightness, sea ice type and other variables, and can provide high recognition accuracy in various scenarios.

KEY WORDS: Sea ice recognition, video camera image, semantic segmentation, deep learning, Deeplab V3+

1. INTRODUCTION

Sea ice poses a major threat to ship navigation in icy regions, and the use of video imager for sea ice identification can effectively enhance navigation safety. In high-latitude areas, sea ice is an important component of the ocean and studying the Sea Ice Concentration (SIC), which refers to the percentage of sea ice area within a sea area, is particularly important. (Su et al., 2013). Sea ice concentration is not only used to describe sea ice-related characteristics but also used for studying global climate change (Laxon et al., 2013) Therefore, sea ice concentration can provide a deep understanding of sea ice generation, growth, and persistence.Understanding the sea ice concentration can provide guidance and assistance for various maritime activities. The distribution of sea ice concentration is of great value for navigation and climate modeling in ice-affected waters (Aldenhoff et al., 2016). Capturing relevant data on sea ice concentration is crucial for important applications such as ship route planning, marine weather forecasting, marine disaster prediction, and marine resource management (Liu et al., 2016). Therefore, the influences of sea ice on ship navigation is comprehensive and multi-faceted, and the recognition and measurement of sea ice are crucial for navigation safety.

In general sea ice identification utilizes the difference in grayscale values between sea ice and other media in images to segment and recognize sea ice. Toyota et al.(2011) proposed an ice-water segmentation algorithm based on global thresholding by analyzing the brightness histogram of input images, combined with manual adjustments to handle adherent fragmented ice edges. Watershed algorithm combined with manual refinement can be also applied to recognize sea ice from the background in images (Blunt et al., 2012). This method can effectively identify sea ice targets in SAR images (Ijitona T. B. et al., 2014). The region merging algorithm employed in this method effectively avoids the over-segmentation problem caused by the traditional watershed algorithm. Zhang et al.(2015) proposed an ice floe contour extraction method based on the gradient vector flow (GVF), which effectively solves the problem of fragmented ice edge adhesion in traditional thresholding methods. In addition to thresholding methods, sea ice image recognition can also be achieved through manual annotation methods, which have become mature approaches.

However, traditional image processing methods usually can only solve binary classification tasks and are more susceptible to variations in image brightness. It becomes challenging to extract features under poor lighting conditions, leading to indistinct boundaries and even mixed segmentation when identifying ice contours. Moreover, in binary images, there are typically only two categories: sea ice and other objects, lacking the ability to understand complex sea ice image scenes. Kalke et al., (2018) mentioned in their study on support vector machine learning for digital river ice images that the current simple and widely used technique is setting a constant threshold and applying it to the recognition process of multiple images. However, due to the inaccuracies in automatic threshold selection, thresholding methods often produce inaccurate results. On the other hand, manual operations can result in high variability in recognition results, as different individuals may have different judgments regarding the ice-water boundary, requiring significant labor for annotation. Furthermore, the thresholds obtained by these two methods are usually subjective and only applicable to images with very similar histograms (i.e., images with little color or brightness variation). As a result, these methods often yield impractical or undesirable thresholding results, significantly deviating from reality.

In recent years, deep learning-based methods have made rapid advancements in the field of computer vision, especially in image semantic segmentation tasks. Convolutional neural networks (CNNs) can automatically learn features from images and explore high-level semantic information, thereby achieving accurate scene understanding across different images. For images with ambiguous features that require manual design and lack intuitive physical meanings, deep learning models can achieve better results with large-scale training data. Particularly in image recognition, deep learning has reduced the error rate by about 30% and made significant progress. Compared to traditional neural networks, deep neural networks have made major improvements and can effectively mitigate training difficulties (such as gradient vanishing) through layer-wise pretraining. These mainstream semantic segmentation networks have achieved outstanding results in fields such as autonomous driving, remote sensing, and medical imaging (Hesamian et al., 2019). However, there is limited research on sea ice image recognition based on convolutional neural networks for on-site monitoring, while most scholars focus on predicting the area and concentration of sea ice in SAR images (Cooke et al., 2019; Wang et al., 2016).

To further improve the accuracy of sea ice recognition in video images, this paper proposes a convolutional neural network method based on the DeeplabV3+ semantic segmentation network framework. The method is applied to the measured image data from the Arctic expedition of the Xuelong vessel. It aims to identify sea ice, seawater, sky, and ship elements in different scenarios and evaluate the accuracy of the recognition results using the mean intersection over union. The method also explores factors that influence the recognition accuracy and calculates the sea ice concentration.

2.Semantic Segmentation Method for Sea Ice Images

2.1 Training Process of Image Segmentation Model

In this paper, DeeplabV3+ is used as the underlying network for training semantic segmentation weights to detect sea ice images. The specific implementation process is shown in Figure 1. Firstly, the dataset is established: video information of ice conditions in the navigation area is collected using a camera mounted on the ship. Different sea ice images are extracted from the navigation videos and manually labeled to create the dataset samples. Then, the training process is conducted. The dataset is inputted into the DeeplabV3+ network for iterative training. The number of training iterations depends on whether the loss function and the average intersection over union (mIoU) of the validation set converge. After training, the corresponding model weight parameter files are saved. Finally, the validation of the trained model is performed: the test images in the validation dataset are inputted into the trained sea ice recognition model, using the trained weight files. The prediction results are compared with the ground truth to evaluate the model accuracy.

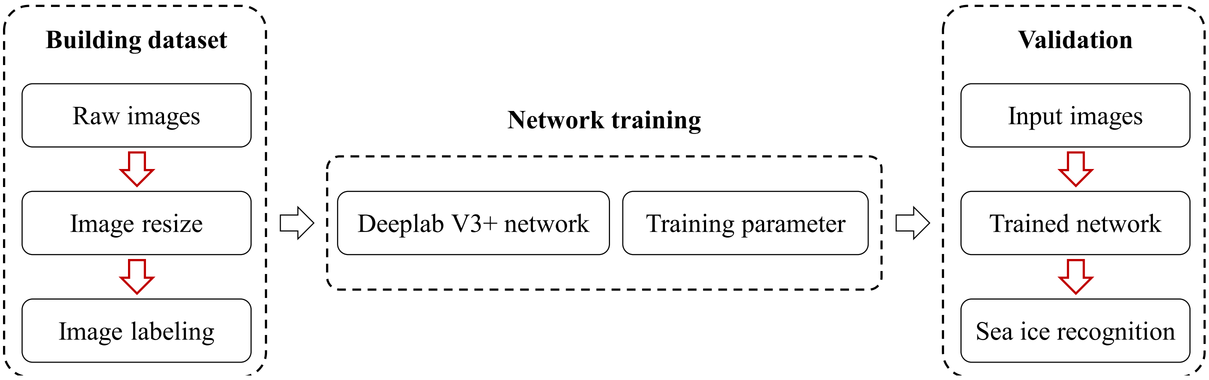


Fig.1 The training process for sea-ice recognition model

2.2 DeeplabV3+ Model

DeeplabV3+ is considered a new pinnacle in semantic segmentation due to its robust on image recognition. Previous approaches for deep neural networks and image semantic segmentation commonly utilized either Spatial Pyramid Pooling (SPP) or encoder-decoder structures. The former encodes multi-scale contextual information by using filters to detect incoming feature maps at a high rate and with multiple effective receptive fields, while the latter progressively reads spatial data to obtain clearer object boundaries. DeeplabV3+ integrates the advantages of both methods by optimizing segmentation results and introducing a more efficient and simplified decoder to upgrade DeeplabV3, particularly in terms of object boundaries. By further developing the original model and applying depth-wise separable convolutions to the spatial pyramid pooling and decoder modules, a faster and more powerful encoder-decoder network, DeeplabV3+, is generated.

DeeplabV3+ introduces a large number of dilated convolutions in the encoder, which increases the receptive field without losing information, enabling each convolution output to encompass a larger range of information. The following is an illustrative diagram of dilated convolutions, which perform cross-pixel sampling during feature extraction. DeeplabV3+ mainly focuses on the architecture of the model, introducing the ability to control the resolution of the features extracted by the encoder, thereby balancing accuracy and computation time.

This new neural network is highly effective on semantic image segmentation datasets and achieves 89% performance on the test set without any post-processing. In summary, the characteristics of DeeplabV3+ are as follows:

1) DeeplabV3+ is a novel encoder-decoder structure that utilizes the powerful DeeplabV3 as the encoder module and an updated, simplified, and efficient decoder module.

2) In the encoder-decoder structure of DeeplabV3+, the precision of the extracted encoder features can be controlled arbitrarily using convolutional methods to balance accuracy and runtime, which is not possible in existing encoder-decoder models.

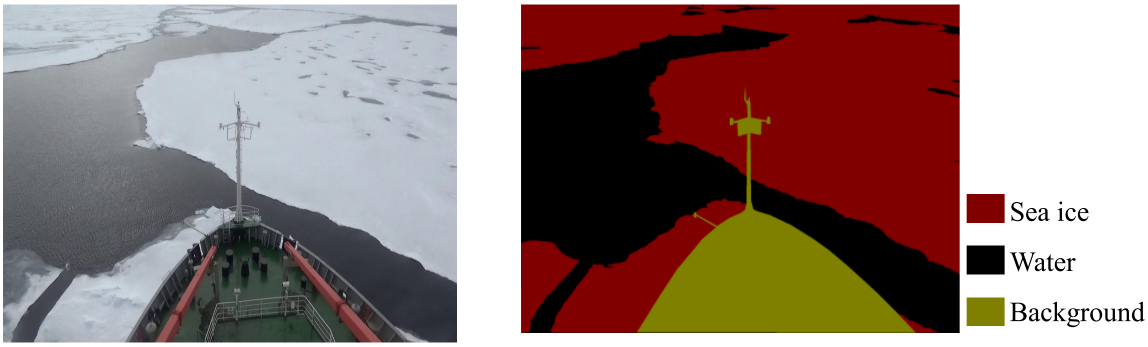
3) Detailed analysis of design choices and model variations can be provided.

4) Functionality can be implemented based on Tensorflow's model.

3. Application of Sea Ice Image Recognition Based on Xuelong's Cruise Observation Data

3.1 Description of images dataset

The accuracy of sea ice image collection is generally limited by the precision of the camera itself and the principles of photography. Therefore, image collection is mainly suitable for sea ice parameter experiments within the range of 10-1000 meters, which is a common engineering range. Within this range, measurements of sea ice concentration, thickness, ice block size, ice motion speed, and types of ice damage have higher effectiveness and accuracy. In this study, the tested parameter is sea ice concentration, which refers to the proportion of sea ice area in a certain area of the sea surface. In image monitoring, the measured digital images are binarized, and the sea ice concentration is calculated by computing the ratio between sea ice and non-sea ice pixels. To improve the accuracy of sea ice concentration calculation, image collection is usually conducted by increasing the monitoring angle and range. Figure 3 shows some sea ice images and their corresponding annotation results from the training set. There are four semantic classes: sea ice, water, sky, and ships. The dataset consists of 670 images, with a validation-to-training ratio of 1:9. This means that 603 images are selected as the training set, and the remaining 67 images are used as the validation set. Due to the small scale of this experimental dataset, data augmentation techniques such as horizontal flipping and brightness adjustment are applied to expand the training set and reduce network overfitting before training.

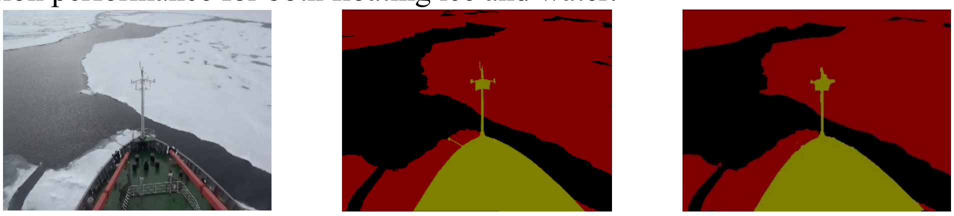


(a) Raw image (b) Labeled image

Figure.3 The segmentation of raw images

3.2 Image recognition result

After the network training is completed, the sea ice recognition of trained model is evaluated using the sea ice images from the validation set. Some of the predicted results are shown in Figure 4. In the figure, the raw image, ground truth value by labeling and recognized result with trained model are presented It can be observed that the trained sea ice semantic segmentation model effectively recognizes the input RGB images and shows good classification performance for the sea ice, water, and sky classes. In the scene 1(Fig.4(a)), the model accurately segments different shapes of floating ice, and particularly performs well in predicting fragmented ice. In the scene 2(Fig.4(b)), despite poor lighting conditions and the similarity in brightness between the sea ice at the far end of the horizon and the sky, the model still accurately locates the horizontal plane. In the scene 4(Fig.4(d)), the predicted results show the advantages to recognize the contours of floating ice, especially in identifying cracks and holes on the right side of the ice, which closely match the ground truth. However, it can be observed that the prediction accuracy of the horizontal plane is affected when there is significant occlusion such as heavy fog, as seen in the third row. In the fifth row, when a large area of sea ice is present in the field of view, the neural network demonstrates stable segmentation performance for both floating ice and water.



(a-1) Scene1: Raw image (a-2) Labeled image (a-3) Recognized image

(b-1) Scene 2: Raw image (b-2) Labeled image (b-3) Recognized image

(c-1) Scene 3: Raw image (c-2) Labeled image (c-3) Recognized image

(d-1) Scene 4: Raw image (d-2) Labeled image (d-3) Recognized image

(e-1) Scene 5: Raw image (e-2) Labeled image (e-3) Recognized image

Figure.4 The recognition results of images in different scenarios

3.3 Influence of concentration on sea ice recognition

To evaluate the accuracy of recognized results, Intersection over Union (IoU) and Mean Intersection over Union (MIoU) are applied as metrics. The two variable can be calculated as the following expressions:

$$IoU = \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{1}$$

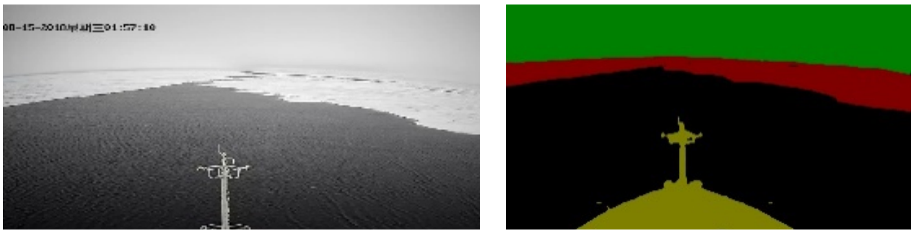
$$mIoU = \frac{1}{n} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{2}$$

where n represents the number of label categories; n_{ii} represents the pixels belongs to category in ground truth, t_i represents the total number of pixels belongs to category i in ground truth; n_{ji} represents the number of pixels in category j but is predicted into category i. After obtaining the segmentation result, the sea ice concentration can be determined based on the ratio of ice pixels to the total number of ice and sea pixels, using the following formula:

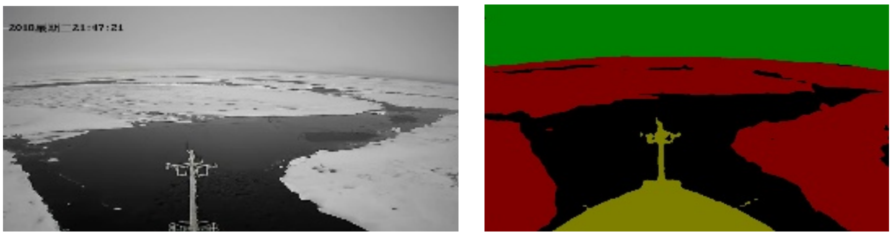
$$C_i = \frac{Pixel_{ice}}{Pixel_{ice} + Pixel_{water}} \tag{3}$$

The values of ice *mIoU* under different ice concentration is shown in Table 3. From the data, it can be observed that the predicted *mIoU* value is related to the distribution of ice in the images. When the ice concentration is low, such as in Scene 3 where the number of ice pixels in the ground truth is small, errors in the predicted ice segmentation will become critical on the calculation of *mIoU*. On the other hand, when there is a higher concentration of ice in the field of view, a few misclassified ice pixels will not have a significant influence on the *mIoU*.

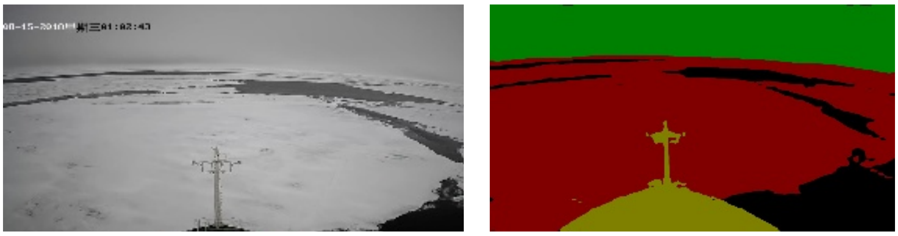
Therefore, the selection of 67 images from the validation set takes into account different ice distribution scenarios, providing a better basis for evaluating the segmentation performance of the model.



(a) Recognition result of image with low concentration, C_i=15%, $mIoU$=96.2%



(b) Recognition result of image with medium concentration, C_i=55%, $mIoU$=96.85%



(c) Recognition result of image with high concentration, C_i=81%, $mIoU$=97.2%

Figure.5 Recognition resolution of image with different sea ice concentration

The influence of ice concentration on recognition accuracy is shown in Fig.5 and Fig.6. It can be observed that the $mIoU$ of the validation set, which is an average weighted by other recognition objects besides ice, is slightly influenced by ice concentration but still shows an increasing trend. On the other hand, the ice-specific IoU is more significantly influenced by ice concentration, exhibiting a more pronounced increasing trend.

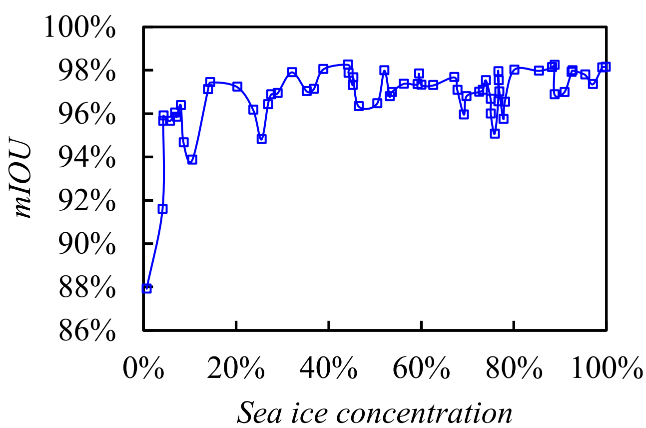
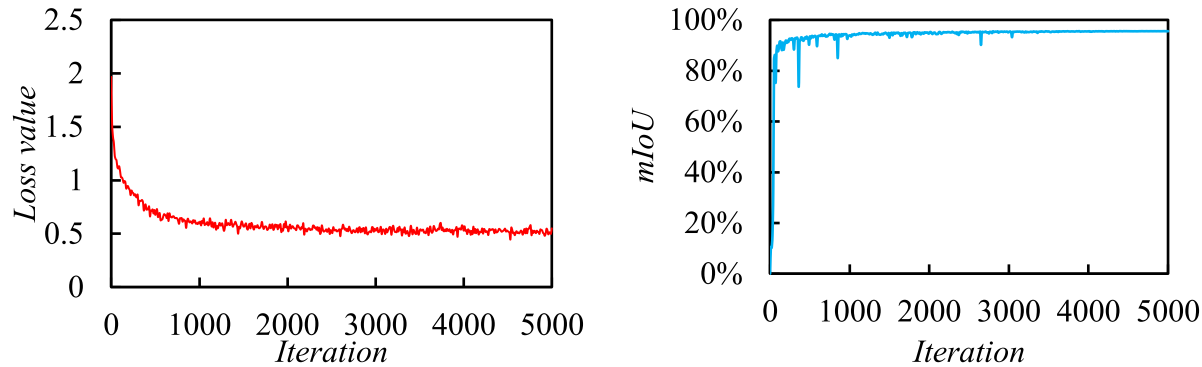


Figure 6 Correlation between mIoU and sea ice concentration.

3.4 Influence of training iteration on recognition

The accuracy of ice image recognition is also related to the number of training iterations. In this experiment, the 5000 training iterations were performed in the training process. The correlation between loss value and iteration is given in Fig.7. It shows that the loss value decreases significantly in the first 1000 iterations and becomes stable from iteration 2000 to 5000. Also, in the Fig.7(b), the *mIoU* value has a sudden increase within 1000 iteration and becomes stable at 95% after 1000 iteration. It means the iteration value has strong influences on the model accuracy. In the training procedure, it is critical to set a propriate iteration value to reach a reasonable loss value and *mIoU* value.



(a) The relationship between loss value and interation (b) The relationship between mIoU and interation

Figure 7 The influence of iteration on the trained model precision

CONCLUSIONS

In this study, a deep learning-based approach was employed using the DeeplabV3+ semantic segmentation framework to investigate the recognition capability of convolutional neural networks applied to sea ice images captured by on-board cameras. By predicting the presence of sea ice, seawater, and sky objects, the feasibility of semantic segmentation models in sea ice image recognition was demonstrated. The findings of this study are as follows:

(1) The trained sea ice semantic segmentation model effectively recognizes sea ice from raw images, producing predictions that closely align with ground truth, with an *mIoU* value of 96.92% on the validation set.

(2) The transfer learning approach employed in this study allows the network to efficiently learn from a small-scale dataset and achieve excellent segmentation results, demonstrating good generalization capability across different scenarios.

(3) The absence of incorporating shallow-level features in the DeeplabV3+ network structure results in less fine-grained predictions for some fragmented ice in the predicted images.

Further work should be done to evaluate the predictive capabilities of other mainstream semantic segmentation networks for sea ice images and include a wider range of sea ice image categories to expand the dataset. Incremental learning approaches should be employed to further enhance the reliability and accuracy of the models.

ACKNOWLEDGEMENTS

This study was supported financially by the National Key Research and Development Program of China (Grant nos.2018YFA0605902 and 2016YCF1401505) and the National Natural Science Foundation of China (Grant no. 41576179).

REFERENCES

Aldenhoff W, Berg A, Eriksson L E B. Sea ice concentration estimation from Sentinel-1 synthetic aperture radar images over the Farm Strait[C]. IEEE International Geoscience and Remote sensing symposium(IGARSS) Beijing, China: IEEE, 2016: 7675- 7677.

Blunt, J., Garas, V., Matskevitch, D., et al. Image analysis techniques for high Arctic, deepwater operation support[C]. OTC Arctic Technology Conference, Houston, Texas, USA, 2012.

Cooke, Colin LV, Andrea Scott K. Estimating sea ice concentration from sar: Training convolutional neural networks with passive microwave data[J]. IEEE Transactions on Geoscience and Remote Sensing. 2019, 57(7): 4735-4747.

Hesamian, Mohammad H, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. Journal of digital imaging 32.4 (2019): 582-596.

Ijitona T B, Ren J, Hwang P B. SAR sea ice image segmentation using watershed with intensity-based region merging[C]//2014 IEEE International Conference on Computer and Information Technology. IEEE, 2014: 168-172.

Kalke H, Loewen M. Support vector machine learning applied to digital images of river ice conditions[J]. Cold Regions Science and Technology, 2018,155(11):225-236.

Laxon S W, Giles K A, Ridout A L, et al. CryoSat 2 estimates of Arctic sea ice thickness and volume[J]. Geophysical Research Letters. 2013, 40(4): 732-737.

Liu Y H, Key J, Mahoney R. Sea and freshwater ice concentration from VIIRS on Suomi NPP and the future JPSS satellites[J]. Remote Sensing, 2016, 8(6): 523

Su J, Hao GH, Ye XX, Wang WB. The experiment and Validation of sea ice concentration AMSR-E retrieval algorithm in polar region[J]. Journal of Remote sensing. 2013, 17(3): 495-513.

Toyota T, Haas C, Tamura T. Size distribution and shape properties of relatively small sea-ice floes in the Antarctic marginal ice zone in late winter[J]. Deep Sea Research Part II: Topical Studies in Oceanography. 2011, 58(9-10): 1182-1193.

Wang, Lei, et al. Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study[J]. IEEE Transactions on Geoscience and Remote Sensing 54.8 (2016): 4524-4533.

Zhang Q, Skjetne R, Metrikin I, et al. Image processing for ice floe analyses in broken-ice model testing[J]. Cold Regions Science and Technology, 2015, 111: 27-38.