

Deep-sea Nodule Mineral Image Segmentation Algorithm Based on Pix2PixHD

Wei Song^{1,2,3}, Haolin Wang¹, Xinping Zhang¹, Jianxin Xia^{4,*}, Tongmu Liu⁵ and Yuxi Shi⁶

¹School of Information and Engineering, Minzu University of China, Beijing, 100081, China

²Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resource, Guangzhou, 510300, China

³National Language Resource Monitoring & Research Center of Minority Languages, Beijing, 100081, China

⁴School of Ocean Science, China University of Geosciences, Beijing, 100191, China

⁵Department of Buoy Engineering, South China Sea Marine Survey and Technology Center, Guangzhou, 510300, China

⁶Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, New Jersey, 07102, USA

*Corresponding Author: Jianxin Xia. Email: jxxia@vip.sina.com

Received: 12 January 2022; Accepted: 10 March 2022

Abstract: Deep-sea mineral image segmentation plays an important role in deep-sea mining and underwater mineral resource monitoring and evaluation. The application of artificial intelligence technology to deep-sea mining projects can effectively improve the quality and efficiency of mining. The existing deep learning-based underwater image segmentation algorithms have problems such as the accuracy rate is not high enough and the running time is slightly longer. In order to improve the segmentation performance of underwater mineral images, this paper uses the Pix2PixHD (Pixel to Pixel High Definition) algorithm based on Conditional Generative Adversarial Network (CGAN) to segment deep-sea mineral images. The model uses a coarse-to-fine generator composed of a global generation network and two local enhancement networks, and multiple multi-scale discriminators with same network structures but different input pictures to generate high-quality images. The test results on the deep-sea mineral datasets show that the Pix2PixHD algorithm can identify more target minerals under certain other conditions. The evaluation index shows that the Pix2PixHD algorithm effectively improves the accuracy rate and the recall rate of deep-sea mineral image segmentation compared with the CGAN algorithm and the U-Net algorithm. It is important for expanding the application of deep learning techniques in the field of deep-sea exploration and mining.

Keywords: Deep-sea mineral; image segmentation; generative adversarial network



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

In recent years, with the development of the global economy and the continuous growth of the population, the demand for mineral resources has been continuing to rise, the depletion of land mineral resources is difficult to meet the needs of industrial production, and the contradiction between supply and demand has become increasingly prominent. In this context, many countries have turned their attention to the ocean [1]. The vast ocean has become a new development zone for humanity and the commanding heights of international strategic competition. The deep-sea mineral resources account for three-quarters of the world and are almost undeveloped. Although there are a large number of nodule minerals in the deep seas of all oceans, large-scale mining has not yet been achieved considering factors such as the difficulty of shooting, the level of economic cost, technical feasibility, and the accuracy of resource assessment. After evaluating the mineral resources of the sea area, some scientific research teams tried to sample in areas with high mineral density and obvious benefits [2,3]. As one of the methods to efficiently evaluate the distribution of deep-sea mineral resources, the deep towing system uses non-contact photography. Without damaging the seabed environment, the images and videos detected by the Ocean Floor Observation System (OFOS) for mineral resource analysis have become the key to resource assessment. Among them, it is particularly important to segment the mineral image to locate and identify the mineral, which is directly related to the efficiency of the subsequent mining work.

The accuracy of the deep-sea nodule mineral segmentation algorithm based on image processing technology is directly related to the accuracy of resource evaluation. Therefore, segmentation accuracy is the primary goal of the deep-sea mineral image segmentation algorithm. In order to save unnecessary economic and time costs, it is particularly important to evaluate the distribution of deep-sea resources effectively. The segmentation of nodule minerals is one of the most challenging tasks. First of all, the size, morphology, and density of polymetallic nodule minerals are varied, and they may even block each other. Secondly, the captured seabed images have problems such as overexposure, low contrast, blur, uneven light and so on. At present, there may be two main problems in current image segmentation algorithms to detect the distribution of deep-sea polymetallic nodules minerals: over-segmentation or under-segmentation of images in different sea areas. In recent years, with the development of deep learning in the field of image segmentation, different image segmentation models play an important role in their specific fields. Therefore, U-Net and some GAN [4] network models have been used to evaluate the performance of deep-sea nodule mineral image segmentation.

The main contributions of this article include: (1) The effect of hyper parameters on the segmentation results is studied on the seabed mineral image data set. (2) It is the first time successfully using the Pix2PixHD method to segment deep-sea mineral images.

The rest of this paper is organized as follows: Section 2 describes related work of image segmentation. Section 3 introduces the proposed method. Section 4 shows the experiments and analysis. Section 5 presents the conclusion.

2 Related Work

In recent years, many image segmentation methods have been proposed. These methods can be roughly divided into two categories: traditional methods and deep learning based methods. Traditional methods have better interpretability and fast running speed, but the accuracy and precision are greatly affected by parameters, environment, size, morphology, color and so on. In recent years, deep learning based segmentation methods have become more and more popular. The main reason is that deep learning based methods have low requirements for image environment, good generalization, and has

advantages in the accuracy of image segmentation in specific fields, but the interpretability of the deep learning model is worse than traditional methods. And the running speed of the deep learning methods is much slower than many traditional methods. The two types of image segmentation methods are introduced below.

2.1 Traditional Methods

Traditional methods are mainly divided into threshold-based segmentation methods and cluster-based segmentation methods. Threshold segmentation is usually based on the difference in pixel values between the target to be segmented and the background in the image. The method of distinguishing the target from the background is by selecting an appropriate threshold. Deep-sea minerals can be considered as a single target regardless of the type. Therefore, selecting an appropriate threshold can segment the mineral particles from the original image. In 2018, Cho et al. [5] proposed the segment method for the mineral images collected in the KODOS (Korea Deep Ocean Study) area of the North Pacific for the first time. The method repeatedly performs the operations of enhancing contrast and median filtering until the edges of the image are clear, and finally uses threshold segmentation to separate the minerals. However, this method neither solves the problem of uneven illumination of the image nor eliminates the artifacts caused by noise, which will lead to poor final segmentation results. In 2017, Schoening et al. [6] proposed a CoMoNoD (Compact-Morphology-based polymetallic Nodule Delineation) method, which can realize fast processing of deep-sea mineral images with the support of GPU. However, this algorithm is easy to cause under-segmentation, excessive noise, and misjudge other objects as nodule minerals. The accuracy of resource assessment is limited. Mao et al. [7] proposed a segmentation method based on background gray value calculation in 2020, which effectively solved the problem of uneven illumination and morphological defects caused by factors such as sediment coverage in deep-sea nodule mineral images. It can repair the target in the image, morphology, and can segment deep-sea nodule mineral images in a short time. However, the morphology of nodule minerals segmented by this method will be distorted to a certain extent, which does not match the actual morphology of nodule minerals.

Clustering is an algorithm that combines pixels with similar characteristics such as texture, color, or gray value into the same category. Ji et al. [8] analyzed the “blur” problem in the nodule mineral image and concluded that the seabed photosensitive light intensity distribution is the essence of the degradation problem and solved the problem of poor focus and heavy hammer agitation in the deep-sea nodule mineral image. This method can remove the interference caused by uneven lighting, enhance the image sharpness, and improve the contrast effect due to the local blur caused by the seabed sediments and other factors. Schoening et al. [9,10] proposed Rapid PCCA (Rapid Pixel-Classification by Cluster Annotation) by refactoring the PCCA [11] code in 2013. The pre-processing, H²SOM mapping, and post-processing are allocated to three threads, which improves the utilization of CPU and GPU and cache efficiency in the PCCA algorithm. It only takes 0.3 s to segment a single image, making it possible to segment nodule minerals rapidly on scientific research ships in real-time. Schoening et al. [12] proposed a heuristic genetic algorithm named Evolutionary Tuned Segmentation using Cluster Co-occurrence and a Convexity Criterion (ES4C) to achieve fully automatic image segmentation in 2016. However, the ES4C algorithm is a heuristic algorithm, which means that the optimal solution may not be found. In addition, if multiple clusters form a clustering result of a class, it will cause error segmentation. Although initializing the size of nodule minerals can solve some of the problems, it is contrary to the fully automated method and still needs improvement.

2.2 Deep Learning Based Methods

Image segmentation methods based on deep learning mainly include CNN-based methods and GAN-based methods. CNN can extract richer image information through a deep network and has a good expression effect. With the development of CNN in the field of image feature extraction, in 2012, Cirosan first use CNN to challenge semantic segmentation tasks [13]. Cirosan adopted a sliding window method to take a small image patch centered on each pixel and input it into CNN to predict the semantic label of the pixel. But the disadvantage of this method is that it needs to traverse each pixel to extract the training and prediction patches, so the speed is very slow. In 2015, Ronneberger et al. [14] proposed a model based on CNN named U-Net. This model has achieved great success in the field of biological image segmentation. Good results can be achieved with 30 pictures. The U-Net network draws on the structure of FCN [15]. After the fifth stage of the convolutional layer, deconvolution and up-sampling are started, high-order features and low-order features are fused, and the prediction map is finally output. The input image is directly used to calculate the loss of pixel semantic classification. However, U-Net will perform repeated operations on the overlapping parts of the neighborhood, and it needs to balance between accurate positioning and obtaining context information, which reduces the accuracy of segmentation. In 2017, He et al. [16] proposed the Mask R-CNN instance segmentation model, which is a milestone in the R-CNN series of models. However, the number of candidate regions of ROI Align in Mask R-CNN is large and has not been filtered by NMS (Non-Maximum Suppression), and the number of segmentation maps that need to be learned is very large, which makes it difficult to learn the segmentation branch. In 2019, Song et al. [17] designed an improved segmentation method based on U-Net, using deep learning technology for deep-sea mineral image segmentation for the first time and achieved certain results. In 2021, Wang et al. [18] added pyramid up-sampling module and residual module into the U-Net model, further improved the accuracy of deep-sea mineral image segmentation, but there is still room for improvement. Ahmed et al. [19] proposed a model called Spatially Constrained Mixture Model, which developed its own maximization step to be used within this framework. Tang et al. [20] proposed DFFNet (Dual Feature Fusion Network) for semantic segmentation to leverage multi-level features and multi-scale context information in an efficient yet effective manner.

GAN was first proposed by Goodfellow et al. [4] in 2014. It is currently one of the most popular technologies in the field of machine learning. GAN occupies a significant position in the areas of image generation, speech conversion, and text generation. However, when the original GAN network faces the presence of random noise input, the output result becomes difficult to control. Therefore, it is natural to think of adding some constraints to the original input. In 2014, Mirza et al. [21] proposed Conditional Generative Adversarial Networks (CGAN). CGAN makes the input and output controllable by adding additional conditional information. However, the edges of the image generated by CGAN are blurred, and the resolution of the generated image is too low. In 2016, Luc et al. [22] first used GAN for semantic segmentation, which improved the high-order consistency between the label and the predicted segmentation map. But the improvement in accuracy was not enough. In 2017, Isola et al. [23] proposed Pix2Pix based on CGAN to achieve image translation. The biggest contribution of this method was to provide a unified framework to solve the problem of image translation. But the performance of this method in high-resolution images is not satisfactory. In 2017, Zhu et al. [24] proposed Cycle GAN, which has achieved the translation from unpaired image to image and only needs to provide images of different domains. But Cycle GAN changes the background while changing the object and lacks diversity. In 2018, an important update for Pix2Pix named Pix2PixHD (Pixel to Pixel High Definition) was proposed by Wang et al. [25], which can realize high-resolution image generation and semantic editing of pictures. In view of the high resolution of the original

deep-sea mineral images, the Pix2PixHD model will be used for deep-sea mineral segmentation. In 2021, Jiang et al. [26] proposed Asrnet (adversarial segmentation and registration networks) for the simultaneous estimation of the blood vessel segmentation and the registration of multispectral images via an adversarial learning process. Fang et al. [27] proposed a classification algorithm optimization based on Triple-GAN which uses Random Forests to classify real samples and use Least Squares Generative Adversarial Networks (LSGAN) ideological structure loss function to avoid gradients disappear.

Generally speaking, traditional methods and deep learning-based segmentation methods have their own advantages and disadvantages in different domains. Considering that deep learning methods have good applications in specific fields, it is vitally important to explore deep-learning methods suitable for deep-sea mineral image segmentation.

3 Image Segmentation Models Based on GANs

3.1 Main Principles

The GAN network mainly consists of two parts: Generator (G) and Discriminator (D). The general steps of training GAN are as follows: fixed generator, update discriminator. Randomly take some samples from real pictures. Then use the vector of some random samples to get some fake samples through the generator. The true samples are labeled 1, and the fake samples are labeled 0. After constructing the supervised data, the discriminator can be trained. The goal of training the generator is to get high scores for real pictures and low scores for fake pictures. Fixed the discriminator, update generator. Using the vector of random samples, some fake samples are obtained through the generator, then scored by the discriminator. The goal of training the discriminator is to get high scores on fake pictures. The generator and discriminator can be trained by the iteration of the above steps. When using GAN for segmentation, the traditional segmentation network can be used as a generator, then add a discriminator after the generator, the general architecture is shown in Fig. 1.

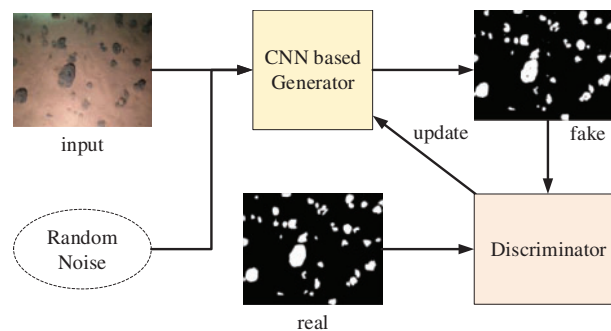


Figure 1: The principle of image segmentation using GANs

3.2 Generator and Discriminator

The network architecture of GAN can be considered an end-to-end model. The first few layers can be used as generators, and the next few layers can be used as a middle layer of the discriminator to output a high-dimensional feature map, which can be regarded as a generated picture.

3.2.1 CGAN

The generator of CGAN is generally as follows: input a normal distribution vector and condition, generate a sample through a generator constructed by a deep neural network. Compared with the original GAN, there is only one more condition input. The discriminator of CGAN is generally as follows: input a sample and condition, which are respectively extracted through the network model for feature extraction, and output a score through the discriminator constructed by the deep neural network. There are two goals of the discriminator. One is that the sample should be as realistic as possible, and the other is that the sample should match the condition. The desired output can be controlled by CGAN.

3.2.2 Cycle GAN

Both GAN and CGAN require real samples that satisfy the condition. Sometimes real samples are not easy to get, so at this time, supervised CGAN needs to be implemented. And then a classic unsupervised GAN network named Cycle GAN was proposed. The Cycle GAN method uses two generators and two discriminators to solve this problem. As shown in Fig. 2, the current task is to convert real pictures into segmented pictures, denoted as Domain X and Domain Y respectively. The steps are as follows: RealX generates FakeY through G_{XY} ; the generated FakeY picture generates RecX through G_{YX} . Using D_x to compare the generated RecX picture with the RealX picture, the closer the two are, the better. Through a number of Y pictures, train Y pictures' discriminator D_y to generate Y segmented pictures, and score them by the discriminator D_y . The higher the score, the better the performance. Through the two generators, the picture of X is converted to Y, and then back to X, the picture of Y is converted to X, and then back to Y, so as to ensure that the two pictures are related.

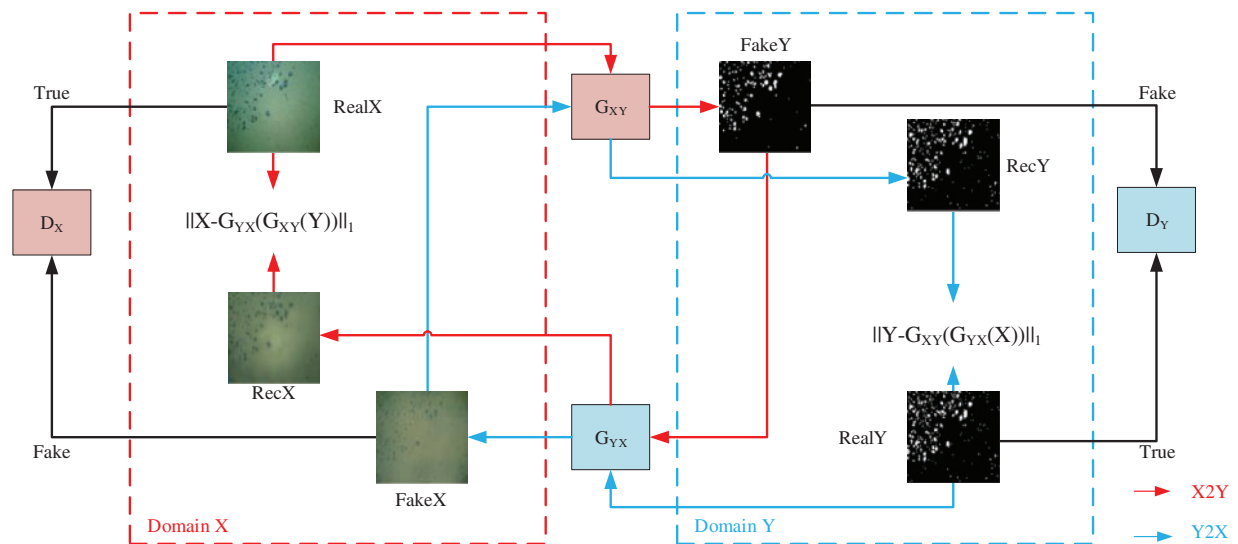


Figure 2: The architecture of Cycle GAN for image segmentation

3.2.3 Pix2pix and Pix2PixHD

Image-to-image translation is to convert the image representation of an object into the other image representation of the object, in other words, it is to find a function to map the A domain image to the

B domain image, which can be applied to many practical problems, such as style transfer, attributes migrate, image super-resolution and so on.

Pix2Pix is one of the most classic image-to-image translation algorithms based on CGAN, the input image can be used as a condition in image translation to learn from input to output. The mapping between the domains is to get the specified output image. Pix2PixHD is an improved version of Pix2Pix, which can generate high-resolution images and semantic editing pictures. For GAN, the key point is to understand the generator, discriminator and loss function. As shown in Fig. 3, the generators of the Pix2PixHD network are multi-scale. The loss function is composed of GAN loss, feature matching loss, and content loss. To start with, the multi-scale generator of Pix2PixHD mainly includes two parts, G1 and G2, which are similar in structure. G1 represents the global generation network, the input and output size is 1024×512 , G2 represents the local enhancement network, the input and output size is 2048×1024 . In general, it can be seen as embedding another generator G1 in a conventional generator G2. The final output feature of G1 and the intermediate output feature of G2 are fused as the input of the second half of G2. At the same time, in the training process, G1 with a smaller resolution is first trained, and then G1 and G2 are trained together. Secondly, the multi-scale discriminator design of Pix2PixHD first builds a three-layer image pyramid (such as 1024×512 , 512×256) based on the real image and the synthetic image, and trains a discriminator for each layer of the image to distinguish two discriminators with same network structures but different input picture sizes. In this part of the design, it is hoped that the discriminator with a small size will facilitate the synthesis of the overall image, and the discriminator with large size will facilitate the synthesis of the image details.

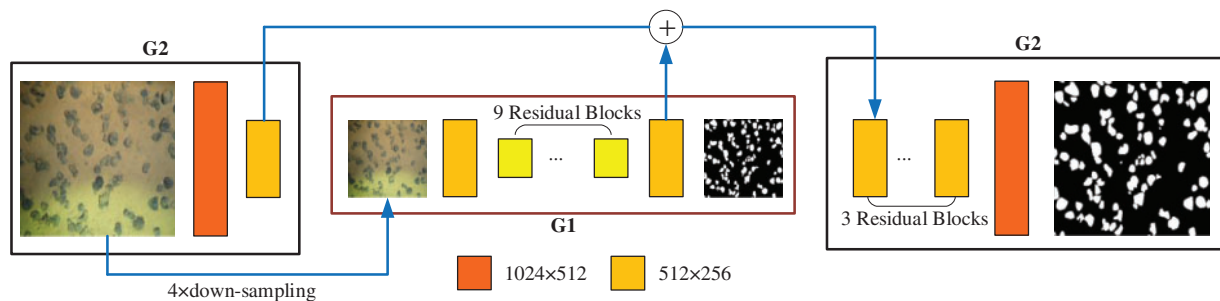


Figure 3: The generator of Pix2PixHD method

3.3 Loss Functions

3.3.1 GAN and CGAN

It is well known that adjusting the neuron weight parameters of the network in a neural network is achieved by defining the loss function and using gradient descent and backpropagation algorithms to continuously optimize the value of the Loss function. The loss function of the original GAN is defined as Eq. (1):

$$L_{GAN} = \min_G \max_D [E_{x \sim p_{data}(x)} \log D(x) + E_{z \sim p_z(z)} \log(1 - D(G(z)))] \quad (1)$$

where D represents the discriminator, G represents the generator, z represents the input random data, z obeys the $Z \sim p_z(z)$ distribution, x represents the input real image data, obeys the p_{data} distribution, and E represents the mathematical expectation.

For CGAN, only the additional information y that needs to be combined with x and z as the input of G and D , therefore the loss function of CGAN is obtained as Eq. (2):

$$L_{CGAN} = \min_G \max_D [E_{x \sim p_{data}(x)} \log D(x|y) + E_{z \sim p_z(z)} \log(1 - D(G(z|y)))] \quad (2)$$

The discriminator D needs to maximize the objective function such that $D(x)$ (or $D(x|y)$) is close to 1 and $D(z)$ (or $D(z|y)$) is close to 0, and the generator G needs to minimize the objective function such that $D(G(z))$ (or $D(G(z|y))$) is close to 1.

3.3.2 Cycle GAN

The loss function of Cycle GAN is the same as GAN, and the anti-loss is composed of two parts:(shown in Eqs. (3) and (4))

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)} [\log D_Y(y)] + E_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (3)$$

$$L_{GAN}(F, D_X, Y, X) = E_{x \sim p_{data}(x)} [\log D_X(x)] + E_{y \sim p_{data}(y)} [\log(1 - D_X(G(y)))] \quad (4)$$

Theoretically, adversarial training can learn to map output G and F , which produce the same distribution as target domains Y and X respectively. However, with sufficient capacity, the network can map the same set of input images to any random arrangement of images in the target domain. Therefore, the adversarial loss alone cannot guarantee that a single input can be mapped. An additional loss is needed to ensure that G and F can satisfy their respective discriminators and be applied to other pictures.

$$L_{cyc}(G, F) = E_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (5)$$

Therefore, the total loss is defined by Eq. (6). (λ is a hyper parameter that can be adjusted as appropriate.):

$$L_{CycleGAN}(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (6)$$

3.3.3 Pix2pix and Pix2PixHD

The Pix2Pix loss function defined by Eq. (7) is the same as the CGAN, but the L_1 loss is added to the generator network.

$$L_{L_1}(G) = E_{x,y,z} [\|y - G(x, z)\|_1] \quad (7)$$

Compared with Pix2Pix, Pix2PixHD has two improvements: the generation of high-resolution images, semantic editing of pictures. In order to generate high-resolution pictures, Pix2PixHD uses a coarse-to-fine generator, multi-scale discriminator, better loss design shown in Eq. (8), and use the instance boundary map for training. (λ is a hyper parameter that can be adjusted as appropriate.)

$$L_{Pix2pixHD} = L_{CycleGAN}(G, F, D_X, D_Y) + \lambda L_{L_1}(G) \quad (8)$$

4 Experiments and Analysis

4.1 Datasets and Preprocessing

This experiment uses 106 images of different regions, brightness, abundance, depth, and shooting angles provided by the OFOS system, of which 90 images are used as the training set and 16 images as the test set. The corresponding ground truth is created by LabelMe software, which was used to label

the nodule minerals in the images, and the dataset contains only one category of mineral targets, the abbreviation Mn for the word mineral was used as the label name. And then the software will generate JSON files for each image, and these JSON files can be transformed into binary mask label maps, which is the ground truth.

4.2 Segmentation Metrics

In order to evaluate the performance of different deep learning based image segmentation methods on the test set of deep-sea nodule mineral images, accuracy, precision, recall, and IoU (Intersection over Union) score are used as evaluation metrics, accuracy is the ratio of the number of correctly predicted mineral samples to the total number of samples, precision is the ratio of the number of predicted mineral samples to the number of actual mineral samples, recall is the probability that samples that actually belong to minerals are correctly predicted as minerals. The IoU score is a term to describe the extent of overlapping of the number of predicted mineral samples and the number of actual mineral samples. The formulas for accuracy, precision, recall, and IoU are shown in Eqs. (9)–(12), where TP denotes correctly predicted true samples, TN denotes correctly predicted false samples, FP denotes incorrectly predicted true samples, and FN denotes incorrectly predicted false samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (12)$$

4.3 Results and Analysis

According to the comparative experimental analysis of GAN networks and other deep learning networks on the same data set, it can be seen from Figs. 4 and 5 that the Pix2PixHD network can segment more particles in the image, and the edges are clearer. Especially in the place marked by the blue area in Fig. 5 where different segmentation algorithms are used for the same picture, the Pix2PixHD method can not only identify the target mineral more completely but also be less affected by the occlusion of the target by marine organisms or other factors. The results reflect the adaptability of the Pix2PixHD network to hyper parameters. The calculation result is based on the average of the test set scores of the 16 pictures taken. Under the premise of certain other conditions, the evaluation indicators of different algorithms show that the accuracy, precision, recall and IoU scores of the test set using the Pix2PixHD model have reached 97.81, 88.83, 98.55, and 77.98, except that the accuracy rate is slightly below the Cycle GAN network, and IoU score is slightly below to Pix2Pix network, other indicators perform better. The main reason that affects the precision and IoU score of Pix2PixHD in Tab. 1 is the misjudgment. Some areas with traces of minerals are mistaken for the target, and the reason for affecting the quality of the mask is the features of edge information extraction are not thorough enough. It is obvious that based on the above results, the Pix2Pix and Pix2PixHD methods perform better than the U-Net and Improved U-Net methods, and it is effective and feasible in the segmentation of nodule mineral image datasets.

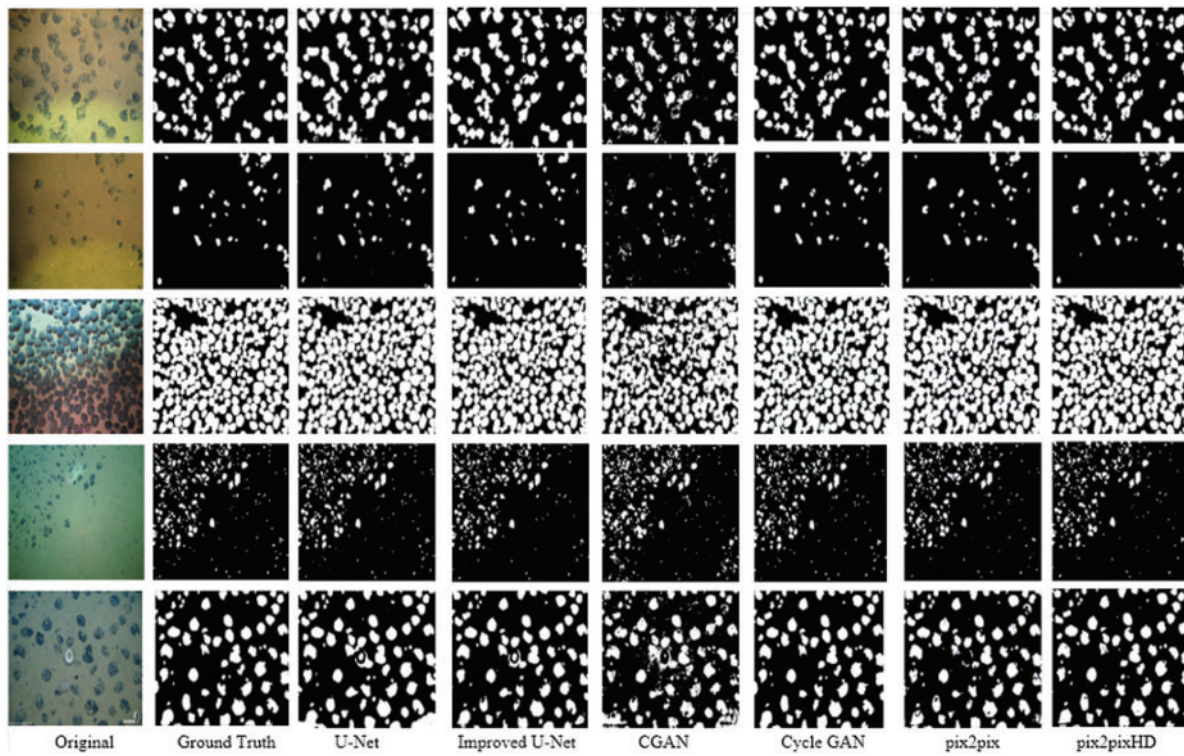


Figure 4: The segmentation results obtained by different segmentation methods

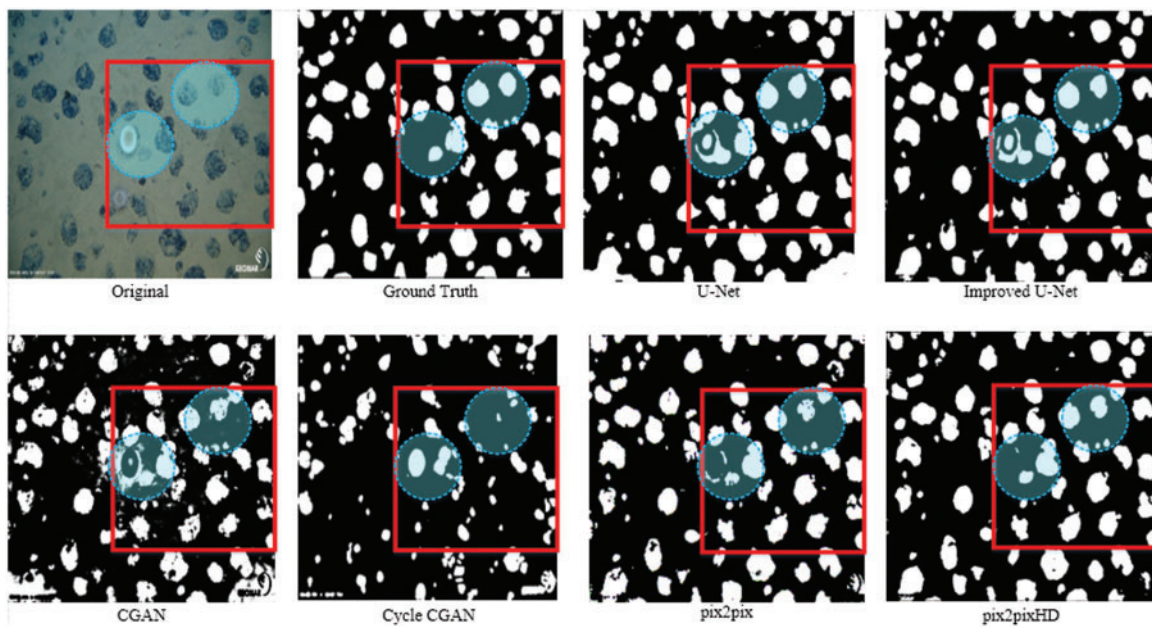
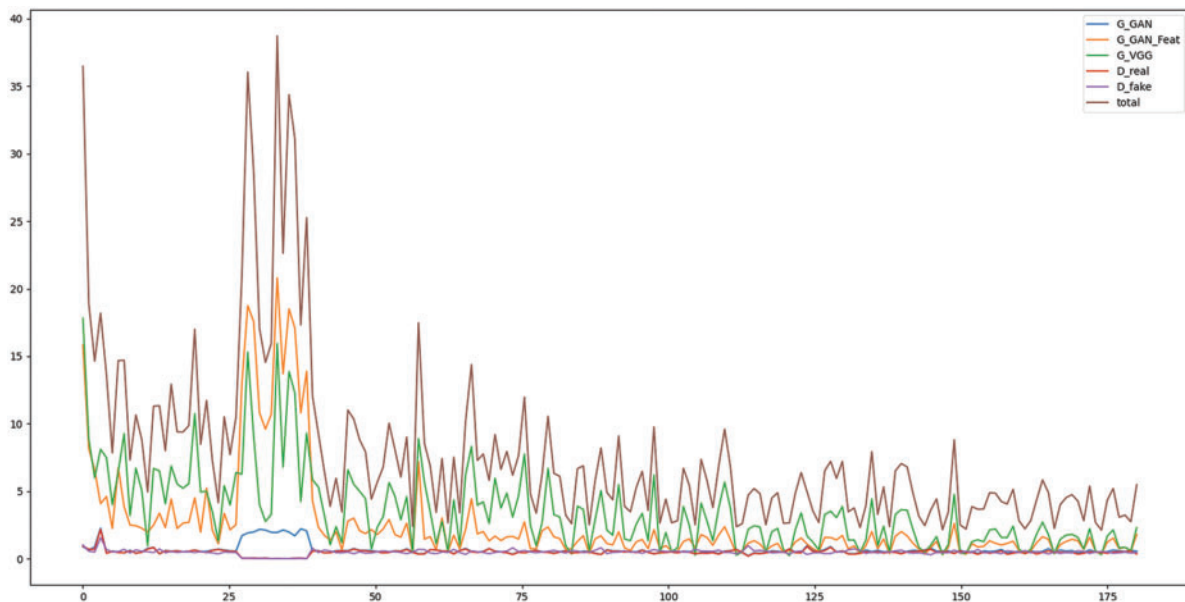


Figure 5: Partial detail maps of mineral image segmentation

Table 1: Comparison of segmentation performances between GANs and other methods

Methods	Accuracy	Precision	Recall	IoU
U-Net [28]	96.97	84.17	79.02	71.12
Improved U-Net [28]	96.80	82.94	79.97	71.39
CGAN	95.92	79.28	80.10	67.51
Cycle GAN	91.10	94.82	77.00	70.71
Pix2Pix	94.79	88.49	82.23	78.49
Pix2PixHD	97.81	88.83	98.55	77.98

In order to verify the effectiveness of the Pix2PixHD network more clearly and intuitively, a series of comparative experiments were used to segment the deep-sea mineral data set. Fig. 6 shows the trend of the loss value of the Pix2PixHD network during the training process. The total loss value of the Pix2PixHD method is composed of G_GAN (Generator Least-Squares GAN Loss), G_GAN_Feat (Generator Feature Matching Loss), G_VGG (Generator VGG Perceptual Loss), D_real (Discriminator loss for real X domain samples) and D_fake (Discriminator loss for generated Y domain samples). It can be clearly seen that the loss of the loss value has a certain degree of volatility, but in general, the algorithm is constantly approaching the expected value. When the epoch is increased, the decrease in the loss value does not change significantly, but the training time is longer.

**Figure 6:** The trend of Pix2PixHD training loss

5 Conclusion

In this study, GAN networks were used to identify nodule minerals in the deep ocean. Because nodular minerals have irregular morphology and similar colors to sediments, traditional image segmentation methods are difficult to accurately segment minerals. In view of the good performance

of the GAN series network on natural data sets in different fields, it is natural to think that the GAN network can also be applied to the segmentation of deep-sea mineral images. Experimental results show that the use of the Pix2PixHD network can accurately detect and segment deep-sea minerals, especially in the improvement of accuracy and recall rate has a good performance. In terms of innovation, this research effectively expands the application scope of generative adversarial networks, making it possible for generative adversarial networks to effectively process underwater images. Using GAN networks can reduce the influence of nodule granularity, color, morphology, and illumination on the segmentation results. This method is suitable for deep-sea mineral images under different sea areas and densities, and the predicted mask results can be used to evaluate the abundance, grain size, and coverage of nodule minerals. However, in nodular minerals, sometimes complicated situations such as adhesion of nodular minerals and sedimentation of blocked minerals may occur, which brings difficulties to the segmentation of nodular minerals. In addition, biological segmentation in deep-sea images will be considered to distinguish individual organisms around different nodule minerals, all of which require further exploration.

Acknowledgement: Thanks to other teachers and students in the Media Computing Laboratory of the Minzu University of China and anonymous reviewers for their valuable comments and contributions to this research.

Funding Statement: This work was supported in part by national science foundation project of P.R. China under Grant No.52071349, No.U1906234 Partially Supported by the Open Project Program of Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resource MESTA-2020-B001, the cross discipline research project of Minzu University of China (2020MDJC08), the Graduate Research and Practice Projects of Minzu University of China.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] N. Toro, R. I. Jeldres, J. A. Órdenes, P. Robles and A. Navarra, "Manganese nodules in Chile, an alternative for the production of Co and Mn in the future—A review," *Minerals*, vol. 10, no. 8, pp. 674–693, 2020.
- [2] C. L. Van Dover. "Tighten regulations on deep-sea mining," *Nature*, vol. 470, no. 7332, pp. 31–33, 2011.
- [3] M. Yu, X. G. Deng, H. Q. Yao and Y. G. Yang, "The progress in the investigation and study of global deep-sea polymetallic nodules," *Geology in China*, vol. 45, no. 1, pp. 29–38, 2018.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, D. Warde-Farley, S. Ozair *et al.* "Generative adversarial nets," in *Proc. of the 25th Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, vol. 27, pp. 1–9, 2014.
- [5] H. Cho, K. H. Kim, S. K. Son and J. Hyun. "Fine-scale microbial communities associated with manganese nodules in deep-sea sediment of the Korea deep ocean study area in the northeast equatorial Pacific," *Ocean Science Journal*, vol. 53, no. 2, pp. 337–353, 2018.
- [6] T. Schoening, D. O. B. Jones and Greinert J. "Compact-morphology-based poly-metallic nodule delineation," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [7] H. D. Mao, Y. L. Liu, H. Z. Yan, C. Qian and J. Xue, "Image processing of manganese nodules based on background gray value calculation," *Computers Materials & Continua*, vol. 65, no. 1, pp. 511–527, 2020.
- [8] J. Ji, Y. Li, and Y. Li. "Current Trends and Prospects of Underwater Image Processing." in *Artificial Intelligence and Robotics*, Springer, Cham, pp. 223–228, 2018.

- [9] T. Schoening, B. Steinbrink, D. Brün, T. Kuhn and T. W. Nattkemper, "Ultra-fast segmentation and quantification of poly-metallic nodule coverage in high-resolution digital images," in *Proc. of the 42th Underwater Mining Institute (UMI)*, Rio de Janeiro and Porto de Galinhas, Brazil, pp. 1–10, 2013.
- [10] T. Schoening, D. Langenkämper, B. Steinbrink, D. Brün and T. W. Nattkemper, "Rapid image processing and classification in underwater exploration using advanced high-performance computing," in *2015 MTS/IEEE OCEANS*, Washington, DC, USA, pp. 1–4, 2015.
- [11] T. Schoening, "Automated detection in benthic images for megafauna classification and marine resource exploration: Supervised and unsupervised methods for classification and regression tasks in benthic images with efficient integration of expert knowledge," *Ph.D. dissertation*, Bielefeld University, Germany, 2015.
- [12] T. Schoening, T. Kuhn, D. O. B. Jones, E. Simon-Lledo and T. W. Nattkemper, "Fully automated image segmentation for benthic resource assessment of poly-metallic nodules," *Methods in Oceanography*, vol. 15, pp. 78–89, 2016.
- [13] D. Ciresan, A. Giusti, L. Gambardella and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. of the 25th Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, pp. 2843–2851, 2012.
- [14] O. Ronneberger, P. Fischer and T. Brox. "U-Net: Convolutional networks for biomedical image segmentation," in *18th Int. Conf. on Medical Image Computing and Computer Assisted Intervention*, Munich, Germany, pp. 1–8, 2015.
- [15] J. Long, E. Shelhamer and T. Darrell. "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 3431–3440, 2015.
- [16] K. M. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," in *Proc. of the IEEE Int. Conf. on Computer Vision (CVPR)*, Honolulu, HI, USA, pp. 2961–2969, 2017.
- [17] W. Song, N. Zheng, X. C. Liu, L. R. Qiu and R. Zheng, "An improved U-net convolutional networks for seabed mineral image segmentation," *IEEE Access*, vol. 7, pp. 82744–82752, 2019.
- [18] H. L. Wang, L. H. Dong, W. Song, X. Zhao, J. X. Xia and T. M. Liu, "Improved u-net-based novel segmentation algorithm for underwater mineral image," *Intelligent Automation & Soft Computing*, vol. 32, no. 3, pp. 1573–1586, 2022.
- [19] M. M. Ahmed, S. A. Shehri, J. U. Arshed, M. U. Hassan, M. Hussain *et al.*, "A weighted spatially constrained finite mixture model for image segmentation," *Computers, Materials & Continua*, vol. 67, no. 1, pp. 171–185, 2021.
- [20] X. Tang, W. Tu, K. Li and J. Cheng, "DFFNet: An IoT-perceptive dual feature fusion network for general real-time semantic segmentation," *Information Sciences*, vol. 565, pp. 326–343, 2021.
- [21] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, pp. 1–7, 2014.
- [22] P. Luc, C. Couprie, S. Chintala and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408*, pp. 1–12, 2016.
- [23] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1125–1134, 2017.
- [24] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE Int. Conf. on Computer Vision (CVPR)*, Honolulu, HI, USA, pp. 2223–2232, 2017.
- [25] T. C. Wang, M. Y. Liu J. Y. Zhu, A. Tao, J. Kautz *et al.*, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 8798–8807, 2018.

- [26] Y. Jiang, Y. Zheng, X. Sui, W. Jiao, Y. He *et al.*, “Asrnet: Adversarial segmentation and registration networks for multispectral fundus images,” *Computer Systems Science and Engineering*, vol. 36, no. 3, pp. 537–549, 2021.
- [27] K. Fang and J. Q. OuYang, “Classification algorithm optimization based on triple-GAN,” *Journal on Artificial Intelligence*, vol. 2, no. 1, pp. 1–15, 2020.
- [28] L. H. Dong, H. L. Wang, W. Song, J. X. Xia and T. M. Liu, “Deep-sea nodule mineral image segmentation algorithm based on mask R-CNN,” in *ACM Turing Award Celebration Conf.-China (ACM TURC)*, Hefei, China, pp. 278–284, 2021.