

Article

Extraction Method for Factory Aquaculture Based on Multiscale Residual Attention Network

Haiwei Zhang ^{1,2}, Jialan Chu ^{1,*}, Guize Liu ¹, Yanlong Chen ¹ and Kaifei He ² 

¹ National Marine Environmental Monitoring Center, Dalian 116023, China; hwzhang@s.upc.edu.cn (H.Z.); gzliu@nmemc.org.cn (G.L.); ylchen@nmemc.org.cn (Y.C.)

² College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China; kfhe@upc.edu.cn

* Correspondence: jlchu@nmemc.org.cn

Abstract: The rapid development of factory aquaculture not only brings economic benefits to coastal areas but also poses numerous ecological and environmental challenges. Therefore, understanding the distribution of coastal factory aquaculture is of great significance for ecological protection. To tackle the issue of the complex spectral and spatial characteristics in remote-sensing images of different factory aquaculture plants in coastal areas, a multiscale residual attention network (MRAN) model for extracting factory aquaculture information is proposed in this study. MRAN is a modification of the U-Net model. By introducing a residual structure, an attention module, and a multiscale connection MRAN can solve the problem of inadequately detailed information extraction from a complex background. In addition, the coastal areas of Huludao City and Dalian City in Liaoning Province were selected as the research areas, and experiments were conducted using the domestic Gaofen-1 remote-sensing image data. The results indicate that the pixel accuracy (PA), mean PA, and mean intersection over union of the proposed model are 98.31%, 97.85%, and 92.46%, respectively, which are superior to those of other comparison models. Moreover, the proposed model can effectively reduce misidentification and missing identification phenomena caused by complex backgrounds and multiple scales.

Keywords: factory aquaculture; U-Net model; residual structure; attention module; multiscale connection; Gaofen-1 image



Academic Editors: Chao Chen,
Qi Wang, Xiyong Hou and Peng Li

Received: 25 January 2025

Revised: 15 March 2025

Accepted: 17 March 2025

Published: 20 March 2025

Citation: Zhang, H.; Chu, J.; Liu, G.; Chen, Y.; He, K. Extraction Method for Factory Aquaculture Based on Multiscale Residual Attention Network. *Remote Sens.* **2025**, *17*, 1093. <https://doi.org/10.3390/rs17061093>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As shown in Figure 1, factory aquaculture is a type of high-density intensive-farming production mode that combines modern industrial and biological technologies in closed and semiclosed water bodies [1]. The development of the factory aquaculture industry not only provides huge economic benefits but also poses several ecological and environmental challenges. For example, the disordered discharge of aquaculture tailwater pollutes the water quality of offshore seawater and frequently causes marine natural disasters, such as red and green tides [2]; in addition, it damages the sandy shoreline near the culture area, resulting in natural habitat loss, ecosystem degradation, and other problems [3,4]. Therefore, strengthening factory aquaculture area monitoring is critical to ensure the orderly development of the coastal factory aquaculture industry.



Figure 1. Factory aquaculture workshop.

Remote-sensing technology has the characteristics of large scale, low cost, and near real-time sensing and thus can obtain land cover information from multiple dimensions and periods [5]. Therefore, remote sensing has become a key technology for ecological environment and biological monitoring. Several methods for extracting aquaculture area information based on remote-sensing images have been proposed [6–10]. For example, Wen et al. extracted aquaculture ponds in the coastal zone of the Beibu Gulf of Guangxi in 2019 based on sentinel temporal remote-sensing data using multithreshold segmentation and object-oriented classification methods [11]. Chu et al. proposed a support vector machine algorithm that combines spectral and textural features to extract floating-raft aquaculture information of the Lianyungang sea area of Jiangsu province [12]. Shi et al. introduced a fast, pixel-wise labeling method called a scanning convolutional network for mudflat aquaculture area detection using infrared remote-sensing images [13]. However, to the best of our knowledge, there is no relevant research on the remote-sensing extraction of distribution information on factory aquaculture.

Recently, the theory and application of deep learning has made rapid progress. Some researchers have employed deep learning semantic segmentation technologies for building recognition in high-resolution remote-sensing images. For example, Long et al. improved the U-Net network structure using a multiscale atrous-convolution module and realized the segmentation of pixel-level urban buildings in high-resolution remote-sensing images [14]. He et al. proposed an E-Unet deep learning network based on atrous convolution to improve extraction accuracy for buildings [15]. These studies provide ideas for the remote-sensing information extraction of factory aquaculture plants.

Compared with the extraction of other buildings, challenges faced in the extraction of factory aquaculture plants primarily include the following. (1) Objects of the same category may show different scale differences in remote-sensing images, such as an aquaculture factory with a length of more than ten meters and those with a length of several hundred meters. (2) The phenomenon of “homogeneous foreign bodies” makes different objects exhibit similar appearance features, such as building roofs and aquaculture buildings coated with mortar. (3) The background is relatively more complex compared to other types of extraction, such as urban building or cultivated land extraction, where the background is simpler and fewer categories are involved. For factory aquaculture plants, the background contains many unrelated categories, such as seawater, bare land, and aquaculture ponds. Field photos and remote-sensing images illustrating these three situations are shown in Figure 2. To address these issues and bridge the current research gap in factory aquaculture image extraction, herein, we modify the U-Net model by introducing a residual structure,

an attention module, and a multiscale connection. We propose an extraction method based on a multiscale residual attention network (MRAN) to enhance the accuracy of factory aquaculture area identification.

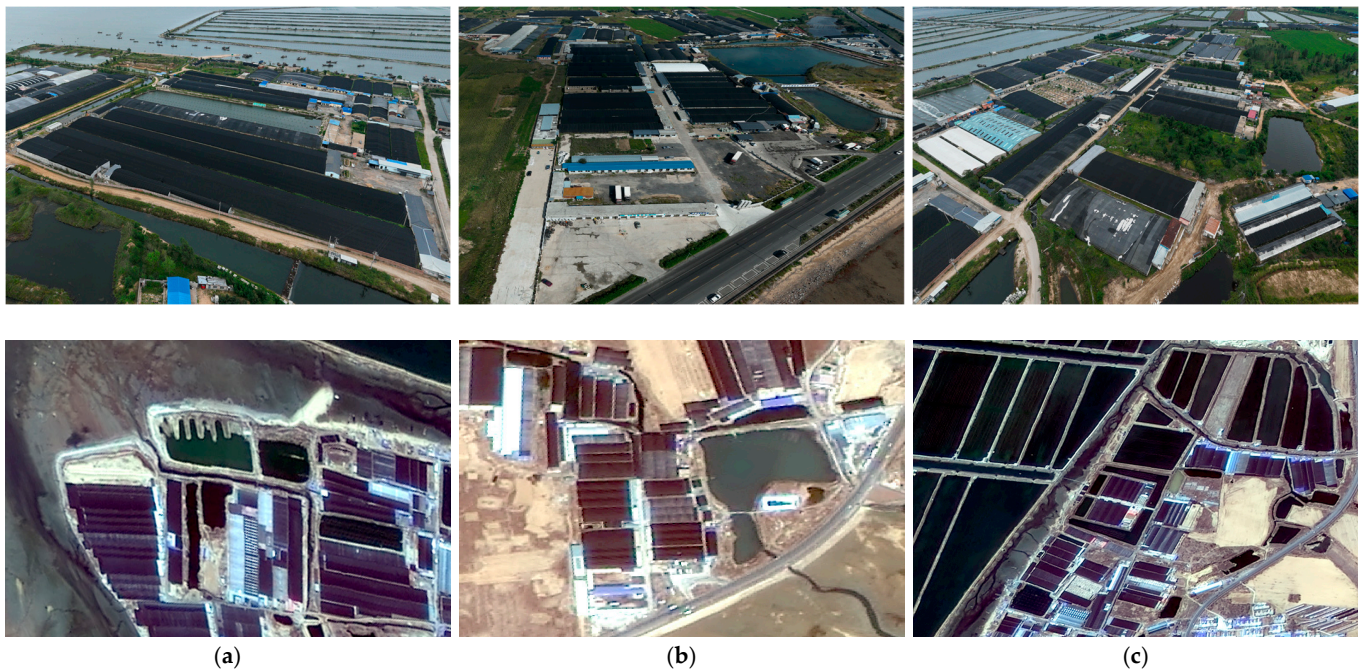


Figure 2. (a–c) Field photos and remote-sensing images illustrating key challenges in factory aquaculture plant image extraction.

2. Materials

As shown in Figure 3, this study took the coastal areas of Huludao City and Dalian City in Liaoning Province as the study areas. The concentrated distribution of factory aquaculture in these areas typically indicates nearshore factory aquaculture, which is suitable for testing model performance. The Gaofen-1 remote-sensing images of the study areas were used for model training and testing. The image acquisition period was April 2023, and the spatial resolution after image fusion was 2 m. By preprocessing the Gaofen-1 remote-sensing images, a dataset required for experiments using manual visual interpretation was constructed. An experimental image and the ground-truth diagram are shown in Figure 4. In the ground-truth diagram, red represents factory aquaculture covered with thermal insulation cotton, denoted as factory aquaculture 1; white indicates factory aquaculture plant coated with mortar, denoted as factory aquaculture 2; and black represents the background. After cropping, 4520 images (192×192 pixels) were obtained, which were divided at a ratio of 8:1:1 for training, verification, and testing, respectively; i.e., 3616 training images, 452 verification images, and 452 testing images.

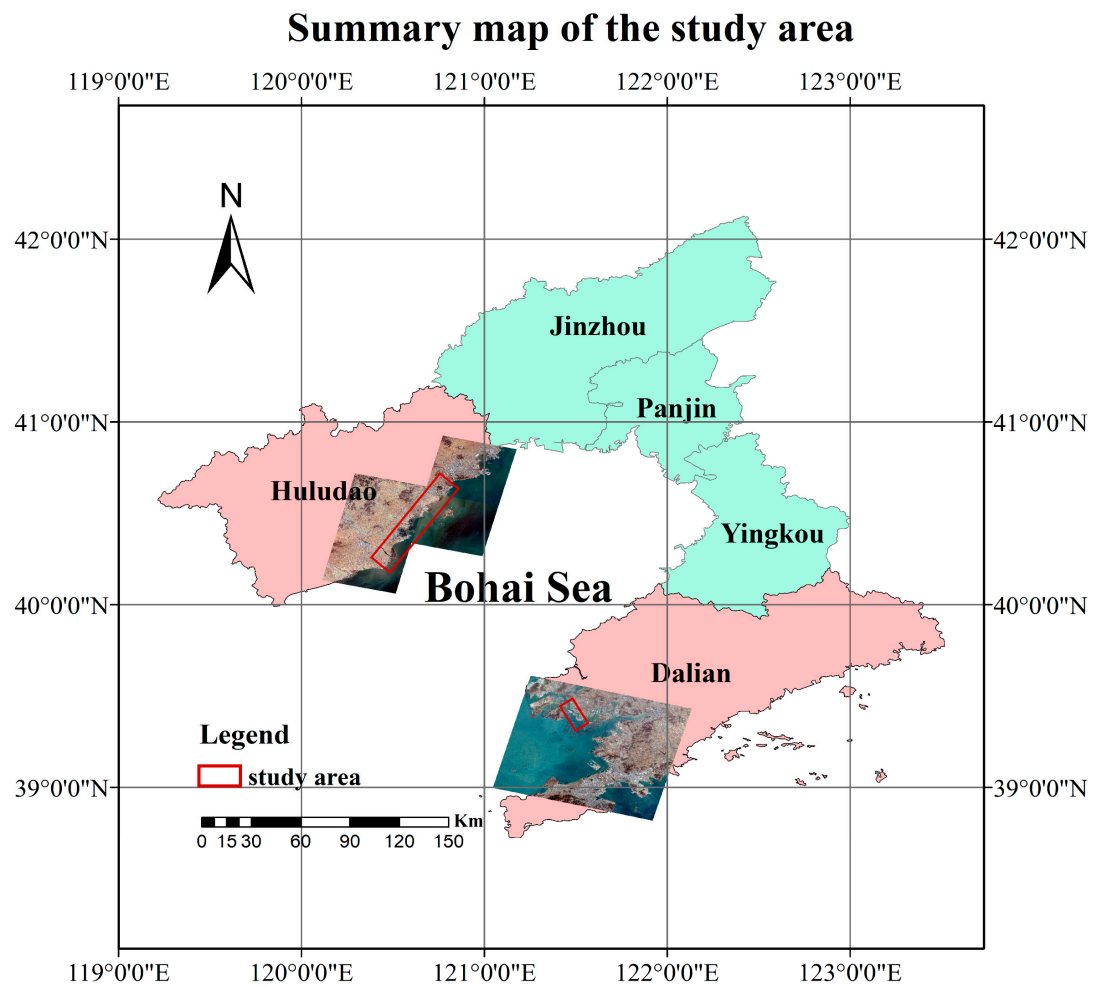


Figure 3. Summary map of study area.



Figure 4. Remote sensing image and ground-truth diagram.

3. Research Method

3.1. Technical Process

The overall technical route of this study is shown in Figure 5. First, the factory aquaculture dataset was constructed via manual annotation. Next, this dataset was fed into the multiscale residual attention network (MRAN) model proposed in this study, and the model parameters were constantly updated through backpropagation for model optimization. Finally, model performance was validated using the test data, and various results were compared and analyzed with those of other classical segmentation models.

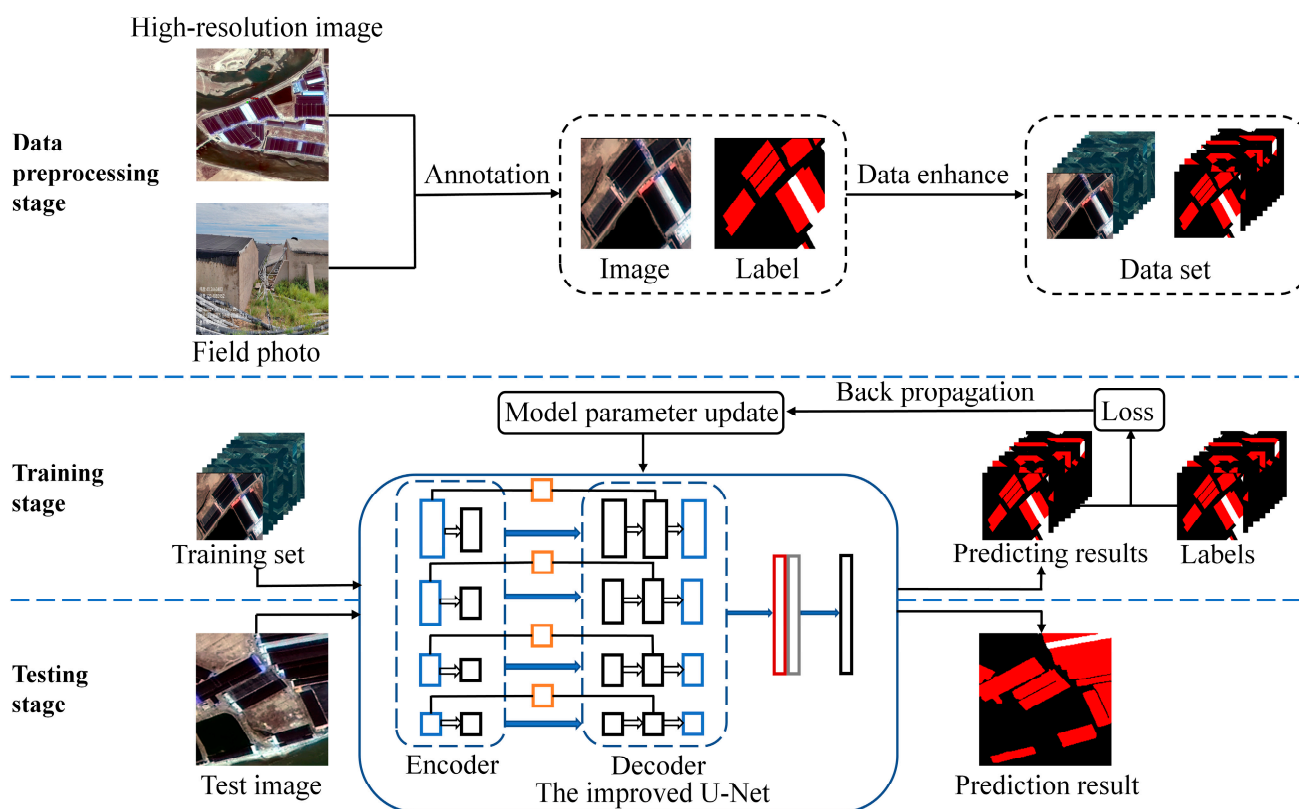


Figure 5. Technical process diagram.

3.2. Overall Network Structure

The proposed MRAN model was based on the U-Net model [16], incorporating a residual structure, an attention module, and a multiscale connection; the network structure is shown in Figure 6. First, a common convolutional block of U-Net was replaced with a residual block to enhance feature extraction ability and network stability, which improves segmentation accuracy. Second, in the encoding stage, the results of the convolution of residual blocks were sent into the attention module to learn the importance of different channels and spatial positions, adjust the weight values of various information adaptively, and enhance useful information while suppressing irrelevant information and improving segmentation accuracy. Finally, via a multiscale connection, the shallow-level features of the encoding stage, deep-level features of the decoding stage, and results of the convolutional attention calculation were cascaded to integrate the features of different levels and improve the learning ability of the model for multiscale objects.

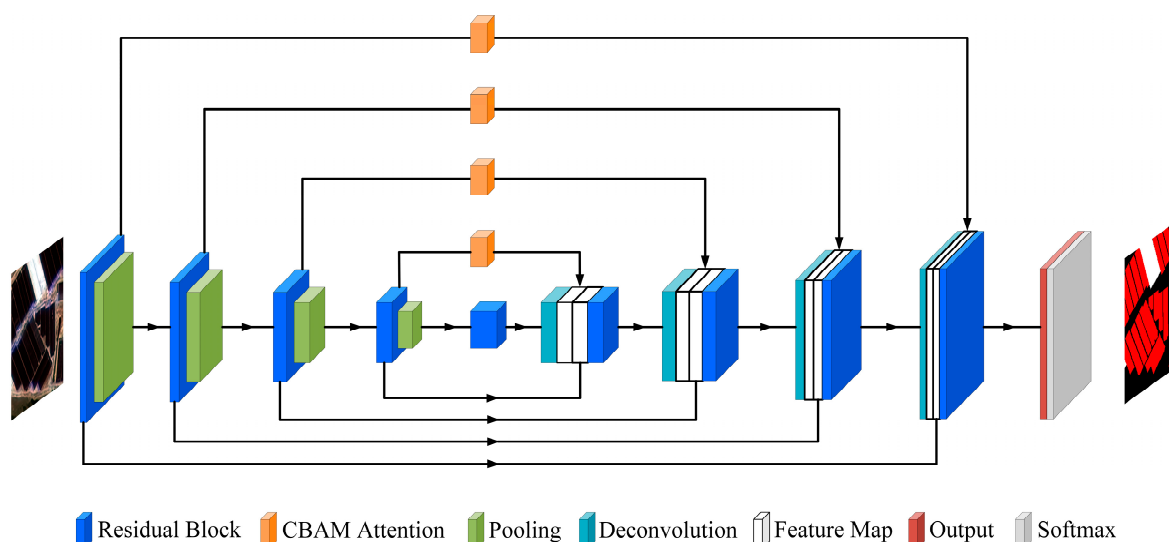


Figure 6. Network structure diagram.

3.3. Residual Structure

In 2016, He et al. [17] proposed a deep residual network. By introducing residual blocks, the network can fit a residual map during learning, rather than directly fitting the desired underlying map. The residual block can learn the residual difference between the input and output such that the model can not only learn new features but also retain and use the previous layer features, enhancing the model feature extraction ability and segmentation accuracy. The calculation formula for a residual block is as follows:

$$y = \mathcal{F}(x, \{W_i\}) + W_s x, \quad (1)$$

where x and y denote the input and output vectors of a residual block, respectively, the function $\mathcal{F}(x, \{W_i\})$ represents learned residual mapping, and W_s is a 1×1 convolution layer used to match the input and output dimensions.

Herein, every two common convolutional layers in the U-Net model were replaced by one residual block, and a total of nine residual blocks were obtained. Consequently, the encoder and decoder stages have five and four residual blocks, respectively. A schematic of a residual block is shown in Figure 7. Each residual block contained two convolution layers, and after each convolution layer, there was a batch normalization layer and a ReLU activation function. Because the number of channels in the input feature graph changed after two convolution layers, a 1×1 convolution was employed to adjust the number of channels in the input feature graph, enabling identity mapping and allowing it to quickly connect with the feature graph after the two convolution layers.

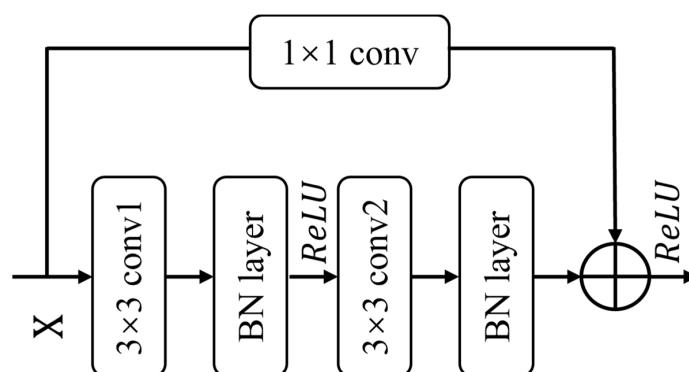


Figure 7. Residual block diagram.

3.4. Attention Module

When humans process an image, they do not directly process the entire image but selectively focus on important information after viewing the entire picture. Inspired by this, an attention mechanism was applied to a semantic segmentation model. Woo et al. proposed a convolutional block attention module (CBAM) [18] that integrates two attention mechanisms: channel, and space attention. Channel attention can learn the importance of each channel feature and thus increase or decrease channel weights. Meanwhile, spatial attention learns features in the spatial domain of a feature map and then enhances or suppresses regional features. Therefore, herein, the convolution results of each residual block were sent to the CBAM module in the downsampling stage for subsequent multiscale connection.

3.5. Multiscale Connection

In deep neural networks, the feature maps of various layers exhibit different resolutions. A shallow feature map has a higher resolution and lower abstraction levels than a deep feature map. U-Net connects a feature map in the encoder with that in the decoder with the same resolution via a skip connection, providing an idea for a multiscale connection. Therefore, a connection path between different levels was added based on U-Net, connecting not only the feature maps of the same resolution in the encoder and decoder but also feature maps calculated via the CBAM module. The results show that this type of multiscale connection can help the model capture different degrees of contextual information by combining feature maps of various scales for richer feature representation.

4. Experiment and Analysis

4.1. Experimental Environment

The experimental platform used in this study was a Windows 10 system, the deep learning software framework was Pytorch 1.12.1, and the Python version was 3.10. The hardware configuration used in the experiment was as follows: Intel(R) Xeon(R) CPU E5-2620 0@2.00GHz processor, NVIDIA GeForce RTX 3080 graphics card.

During model training, the Adam optimizer was used to update the adaptive learning rate of various parameters. The initial learning rate was set to 0.001, a learning rate exponential decay strategy was adopted, and the proportion of each decay was 0.95. The batch size of each input into the training network was set to 8. Cross entropy was employed as the loss function. The initial training rounds of each model were set to 100, and network training was stopped in advance when the loss value of the verification set did not decrease within 10 epochs.

4.2. Evaluation Metrics

To quantify the model's detection performance, pixel accuracy (PA), mean PA (MPA), and mean intersection over union (mIoU) were used as indicators to compare the generated and interpreted images pixel by pixel. PA provides a global performance evaluation and offers a quick overview of the model's classification accuracy. However, PA does not account for category, which can lead to misleading conclusions. In cases where background pixels far outnumber target pixels, PA may still be high even if the model performs poorly in segmenting the target. MPA addresses this limitation by considering classification accuracy across all categories, making it better suited for datasets with imbalanced categories. Meanwhile, mIoU evaluates the degree of overlap between predicted and real regions, providing a more comprehensive measure of segmentation performance. mIoU is particularly effective for complex scenes and small-target segmentation tasks.

In addition, precision (P), recall (R), and F1-score were utilized to evaluate segmentation performed using the proposed model on various ground objects to determine the classification effect of the proposed model on various target ground objects. The P, R and F1-score metrics are widely used to evaluate the performance of classification models, particularly when dealing with unbalanced datasets or when balancing prediction accuracy and completeness is necessary. A higher P indicates greater reliability in positive sample predictions, making it useful for assessing misidentifications of the model. A higher R reflects the model's ability to identify positive samples, making it suitable for evaluating missed detections. The F1-score, a harmonic average of P and R, provides a balanced assessment of the model's overall performance.

The calculation formulas for the indicators are as follows:

$$PA = \frac{\sum_{i=1}^C TP_i + \sum_{i=1}^C TN_i}{\sum_{i=1}^C (TP_i + TN_i + FP_i + FN_i)} \quad (2)$$

$$MPA = \frac{1}{C} \sum_{i=1}^C PA_i \quad (3)$$

$$mIoU = \frac{1}{C} \sum_{i=1}^C \left(\frac{TP_i}{TP_i + FP_i + FN_i} \right) \quad (4)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (6)$$

$$F1 - score = 2 \times \frac{P_i \times R_i}{P_i + R_i}, \quad (7)$$

where C is the total number of categories; TP and TN represent the number of pixels correctly classified as positive and negative; and FP and FN denote the number of pixels incorrectly classified as positive and negative, respectively.

4.3. Results and Analysis

In semantic segmentation, networks such as U-Net, U-Net++, SegNet, PSPNet, and DeepLab v3+ have been widely used. U-Net integrates low-level details with high-level semantic information through skip connections between the encoder and decoder, allowing it to perform well even with limited annotated data, making it suitable for small-sample datasets. U-Net++ enhances feature interaction by incorporating dense connections in the encoder, improving accuracy and robustness in complex scenes. SegNet preserves more boundary information by retaining the maximum-pooled indices in the encoder and passing them to the decoder as location information. PSPNet uses a pyramid pooling module to extract advanced features at different scales and generate subregions of different sizes to achieve effective expression of features at different scales. DeepLab v3+ uses the Xception structure as the backbone network and applies depthwise separable convolutions to reduce model parameters, thereby enhancing the model's segmentation accuracy.

To evaluate the proposed model's performance, it was compared with these classical semantic segmentation models mentioned above. The extraction results of various models on the test set are shown in Figure 8. Overall, the six models can identify both types of factory aquaculture, but there are large differences in details. Features in rows 1 and 2 are very similar to those in factory aquaculture, posing a great challenge to segmentation. In row 1, U-Net and SegNet misidentify some broken and similar ground object information as factory aquaculture. In row 2, U-Net++ encounters obvious misidentification issues. In rows 3 and 4, PSPNet and DeepLab v3+ cannot separate different aquaculture types

individually, and adhesion forms between densely distributed areas of factory aquaculture, which can only be divided into large areas. This is because PSPNet and DeepLab v3+ use a pyramid pool module, which can expand the sensory field and capture contextual information but produce a “grid effect”, resulting in adhesion between images with small gaps. In column 5, although the SegNet model can segment factory aquaculture in general, “pepper and salt noise” is very obvious because of its pixel-by-pixel classification method, and there are numerous broken patches in the segmentation results, which is unsuitable for factory aquaculture image extraction. The proposed model considerably mitigates these issues. By introducing a residual structure and an attention module, the proposed model captures fine details better, reduces interference from surrounding background information, and enhances segmentation robustness. In addition, the multiscale connections enable the model to capture different levels of contextual information resulting in enhanced feature representation.

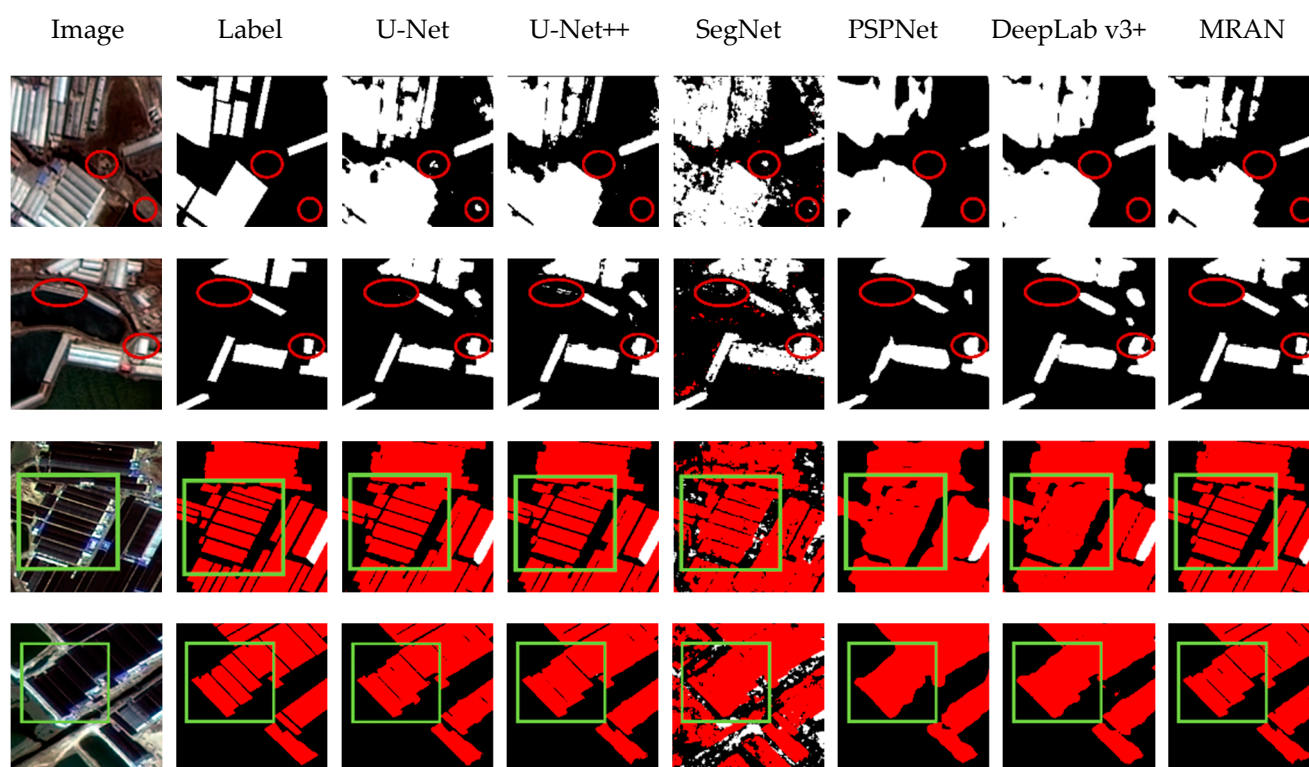


Figure 8. Visualization of test set prediction results.

Table 1 shows the quantitative evaluation results of different models on the test set. The PA, MPA, and mIoU of the proposed model are the highest (98.31%, 97.85%, and 92.46%, respectively), consistent with the segmentation visualization results. Compared with U-Net, U-Net++, SegNet, PSPNet, and DeepLab v3+, the PA of the proposed model increased by 0.92%, 0.63%, 18.31%, 2.43%, and 3.05%; MPA increases by 0.75%, 0.71%, 12.43%, 3.27%, and 1.97%; and mIoU increases by 3.77%, 2.62%, 38.42%, 8.46%, and 10.32%, respectively. Moreover, the proposed model greatly improves the mIoU results, indicating its high accuracy in the identification and positioning of various categories of objects in the segmentation task and ability to effectively reduce misidentification and missing detection phenomena, thereby enhancing the overall segmentation quality.

Table 1. Accuracy evaluation of different models on the test set.

Metric	U-Net	U-Net++	SegNet	PSPNet	DeepLab v3+	MRAN
PA (%)	97.39	97.68	80.00	95.88	95.26	98.31
MPA (%)	97.10	97.14	85.42	94.58	95.88	97.85
mIoU (%)	88.69	89.84	54.04	84.00	82.14	92.46

PA, pixel accuracy; MPA, mean PA; mIoU, mean intersection over union; and MRAN, multiscale residual attention network.

Table 2 summarizes the segmentation accuracy of the proposed model on various ground objects. The P, R, and F1-score of factory aquaculture 2 are lower than those of background and factory aquaculture 1. This is because factory aquaculture 2 includes the entire process of the aging of thermal insulation cotton from gray white to bright white, which is a gradual process, and ground object information is more complex. Moreover, factory aquaculture 2 exhibits certain similarities with the building roof, posing certain identification challenges to the model.

Table 2. Segmentation accuracy of the proposed MRAN on various ground objects.

Metric	Background	Factory Aquaculture 1	Factory Aquaculture 2
P (%)	99.62	94.66	88.63
R (%)	98.38	98.81	96.37
F1-score (%)	98.99	96.69	92.34

P, precision and R, recall.

The efficiency of an algorithm largely determines its practicability. Therefore, the training times of different models were compared in this study. Table 3 lists the number of parameters for each model and the average training time per 10 epochs on the training set. The training time of U-Net per 10 epochs is the least, only 10 min. DeepLab v3+ has the largest number of parameters and requires the longest training time of 43 min per 10 epochs. Comparing the proposed model with U-Net, the proposed model does not require a huge increase in the number of parameters, and its training time per 10 epochs is only increased by 7 min.

Table 3. Number of parameters and training efficiency of different models.

Model	Training Duration (min)	Parameter Quantity
U-Net	10	31.0×10^6
U-Net++	28	47.2×10^6
SegNet	15	29.4×10^6
PSPNet	25	46.6×10^6
DeepLab v3+	43	54.7×10^6
MRAN	17	35.9×10^6

4.4. Ablation Experiment

To verify the effectiveness of each method introduced in the proposed model, ablation experiments were conducted on the test set; the results are summarized in Table 4. U-Net indicates the original U-Net model that remains unmodified, Residual denotes that a common convolutional block is replaced by a residual block, and Multi indicates a multiscale connection. The incorporation of the attention module to U-Net causes 0.50%, 0.26%, and 2.11% performance improvements, while replacing a common convolutional block with a residual block results in performance improvements of 0.87%, 0.40%, and

3.56% for PA, MPA, and mIoU, respectively. Compared with only including the residual structure and attention module, PA, MPA, and mIoU increase by 0.04%, 0.02%, and 0.12%, respectively, after incorporating the multiscale connection. The above analysis reveals that each method included in the proposed model improves performance to a certain extent and makes the segmentation results more accurate.

Table 4. Evaluation indicators of ablation experiment.

PA (%)	MPA (%)	mIoU (%)	U-Net	CBAM	Residual	Multi
97.39	97.10	88.69	✓			
97.89	97.36	90.80	✓	✓		
98.06	97.79	91.50	✓	✓		✓
98.26	97.50	92.25	✓		✓	
98.27	97.83	92.34	✓	✓	✓	
98.31	97.85	92.46	✓	✓	✓	✓

5. Conclusion and Discussion

Understanding the spatial distribution of coastal factory aquaculture is crucial for protecting coastal ecosystems, ensuring seafood supply, and supporting the development of sustainable marine economies. This effort not only advances the concepts of “cultivate sea, herd and fish”, and building a “blue granary”, but also helps effectively optimize the spatial layout of offshore mariculture, promote the transformation and upgrading of marine fisheries, and ensure national food security. In addition, it facilitates the effective planning of aquaculture waters, establishes and improves the monitoring system for marine aquaculture tailwater, and gradually strengthens centralized management of factory aquaculture tailwater, ultimately contributing to the sustainable utilization of marine resources and the healthy development of coastal economies.

Herein, based on a factory aquaculture dataset established using the Gaofen-1 remote-sensing images, the U-Net model is improved by introducing a residual network, CBAM, and a multiscale connection, and the rapid and accurate extraction of factory aquaculture using remote-sensing images is realized. The experimental results show that the PA, MPA, and mIoU of the proposed model are 98.31%, 97.85%, and 92.46%, respectively, which can meet the business requirements of factory aquaculture monitoring. The proposed method not only allows enterprises to track changes in factory aquaculture areas and optimize layouts but also provide government agencies with critical decision-making support for fisheries resource management and environmental protection, which helps prevent illegal or unregulated aquaculture activities that could harm the environment, facilitate timely adjustments to aquaculture planning, and reduce pressure on marine ecosystems.

Nevertheless, the proposed model still faces some shortcomings, such as the misidentification of factory aquaculture plant with gradual changes in thermal insulation cotton from gray white to bright white, and poor segmentation accuracy for the slices of factory aquaculture plant boundary information. In addition, other aquaculture types, such as pond farming and cage farming, were not included. In the future, we will attempt to further improve the accuracy of factory aquaculture image extraction using techniques such as visual transformer [19], reinforcement learning [20], and semi-supervised learning [21]. Furthermore, fusing remote-sensing images from different sources could help develop a unified model capable of handling diverse aquaculture types, providing robust technical support for the comprehensive monitoring of coastal fishery resources.

Author Contributions: Conceptualization, H.Z., J.C., G.L. and Y.C.; methodology, H.Z. and J.C.; software, H.Z.; validation, H.Z., J.C. and G.L.; formal analysis, G.L. and Y.C.; investigation, H.Z. and J.C.; resources, J.C.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review

and editing, K.H.; visualization, H.Z.; supervision, J.C.; project administration, Y.C. and K.H.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62071491 and Grant 41706105.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the first author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBAM	Convolutional block attention module
MPA	Mean pixel accuracy
MRAN	Multiscale residual attention network
PA	Pixel accuracy

References

1. Luo, Q.; Li, J.; Chang, Z.Q.; Chen, Z.; Qiao, L.; Yang, L.G. Changes in zooplankton community structure in industrial shrimp farming ponds. *Prog. Fish. Sci.* **2020**, *41*, 131–139.
2. Qiu, R.S.; Han, L.M.; Yin, W. Green development evaluation and time-space evolution characteristics of mariculture industry in China. *Sci. Geogr. Sin.* **2023**, *43*, 1793–1802.
3. Ahmed, N.; Thompson, S.; Glaser, M. Global Aquaculture Productivity, Environmental Sustainability, and Climate Change Adaptability. *Environ. Manag.* **2019**, *63*, 159–172. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, C.; Meng, Q.H.; Chu, J.L.; Liu, G.Z.; Wang, C.J.; Zhao, Y.Y.; Zhao, J.H. Analysis on the status of mariculture in China and the effectiveness of mariculture management in the Bohai Sea. *Mar. Environ. Sci.* **2021**, *40*, 887–897.
5. Zhang, X.C.; Huang, J.F.; Ning, T. Progress and Prospect of Cultivated Land Extraction from High-Resolution Remote Sensing Images. *Geomat. Inf. Sci. Wuhan Univ.* **2023**, *48*, 1582–1590.
6. Ren, C.Y.; Wang, Z.M.; Zhang, Y.Z.; Zhang, B.; Chen, L.; Xi, Y.B.; Xiao, X.M.; Doughty, R.B.; Liu, M.Y.; Jia, M.M.; et al. Rapid expansion of coastal aquaculture ponds in China from Landsat observations during 1984–2016. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101902. [[CrossRef](#)]
7. Wang, M.; Mao, D.H.; Xiao, X.M.; Song, K.S.; Jia, M.M.; Ren, C.Y.; Wang, Z.M. Interannual changes of coastal aquaculture ponds in China at 10-m spatial resolution during 2016–2021. *Remote Sens. Environ.* **2023**, *284*, 113347. [[CrossRef](#)]
8. Lacaux, J.P.; Tourre, Y.M.; Vignolles, C.; Ndione, J.A.; Lafaye, M. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sens. Environ.* **2007**, *106*, 66–74. [[CrossRef](#)]
9. Rahman, A.F.; Dragoni, D.; Didan, K.; Barreto-Munoz, A.; Hutabarat, J.A. Detecting large scale conversion of mangroves to aquaculture with change point and mixed-pixel analyses of high-fidelity MODIS data. *Remote Sens. Environ.* **2013**, *130*, 96–107. [[CrossRef](#)]
10. Cui, B.G.; Fei, D.; Shao, G.H.; Lu, Y.; Chu, J.L. Extracting Raft Aquaculture Areas from Remote Sensing Images via an Improved U-Net with a PSE Structure. *Remote Sens.* **2019**, *11*, 2053. [[CrossRef](#)]
11. Wen, K.; Yao, H.M.; Huang, Y.; Chen, H.Q.; Liao, P.R. Remote sensing image extraction for coastal aquaculture ponds in the Guangxi Beibu Gulf based on Google Earth Engine. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 280–288.
12. Chu, J.L.; Shao, G.H.; Zhao, J.H.; Gao, N.; Wang, F.; Cui, B.G. Information extraction of floating raft aquaculture based on GF-1. *Sci. Surv. Mapp.* **2020**, *45*, 92–98.
13. Shi, T.Y.; Zou, Z.X.; Shi, Z.W.; Chu, J.L.; Zhao, J.H.; Gao, N.; Zhang, N.; Zhu, X.Z. Mudflat aquaculture labeling for infrared remote sensing images via a scanning convolutional network. *Infrared Phys. Technol.* **2018**, *94*, 16–22. [[CrossRef](#)]
14. Long, L.H.; Zhu, Y.T.; Yan, J.W.; Liu, J.J.; Wang, Z.Y. New building extraction method based on semantic segmentation. *Natl. Remote Sens. Bull.* **2023**, *27*, 2593–2602. [[CrossRef](#)]
15. He, Z.M.; Ding, H.Y.; An, B.Q. E-Unet: A atrous convolution-based neural network for building extraction from high-resolution remote sensing images. *Acta Geod. Cartogr. Sin.* **2022**, *51*, 457–467.
16. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical image Computing and Computing-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

17. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
18. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
19. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.H.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 $\times$ 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual Event, 3–7 May 2021.
20. Ma, C.F.; Xu, Q.S.; Wang, X.F.; Jin, B.; Zhang, X.Y.; Wang, Y.F.; Zhang, Y. Boundary-Aware Supervoxel-Level Iteratively Refined Interactive 3D Image Segmentation with Multi-Agent Reinforcement Learning. *IEEE Trans. Med. Imaging* **2021**, *40*, 2563–2574. [[CrossRef](#)] [[PubMed](#)]
21. Wang, L.F.; Wang, S.S.; Qi, J.; Suzuki, K. A Multi-Task Mean Teacher for Semi-Supervised Facial Affective Behavior Analysis. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 3596–3601.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.