Aquaculture International (2024) 32:5625–5638
https://doi.org/10.1007/s10499-024-01439-x

UISS-Net: Underwater Image Semantic Segmentation Network for improving boundary segmentation accuracy of underwater images

ZhiQian He[1] · LiJie Cao[1] · JiaLu Luo[1] · XiaoQing Xu[1] · JiaYi Tang[1] · JianHao Xu[1] · GengYan Xu[1] · ZiWen Chen[2]

Received: 22 December 2023 / Accepted: 16 February 2024 / Published online: 28 February 2024

Abstract
Image semantic segmentation is widely used in aquatic product measurement, aquatic biological cell segmentation, and aquatic biological classifications. However, underwater image segmentation has low accuracy and poor robustness because of turbid underwater environments and insufficient light. Therefore, this paper proposes an Underwater Image Semantic Segmentation Network (UISS-Net) for underwater scenes. Firstly, the backbone network uses an auxiliary feature extraction network to improve the extraction of semantic features for the backbone network. Secondly, the channel attention mechanism enhances the vital attention information during feature fusion. Then, multi-stage feature input up-sampling is used to recover better semantic features in the network during up-sampling. Finally, the cross-entropy loss function and dice loss function are used to focus on the boundary semantic information of the target. The experimental results show that the network effectively improves the boundary of the target object after segmentation, avoids aliasing with other classes of pixels, improves the segmentation accuracy of the target boundary, and retains more feature information. The mean intersection over union (mIoU) and mPA of UISS-Net in the semantic Segmentation of Underwater IMagery (SUIM) dataset achieve 72.09% and 80.37%, respectively, 9.68% and 7.63% higher than the baseline model. In the Deep Fish dataset, UISS-Net achieved 95.05% mIoU, 12.3% higher than the traditional model.

Keywords Underwater semantic segmentation · Auxiliary feature extraction network · Inverted multi-scale feature fusion · Combined loss function

Handling editor: Gavin Burnell

✉ LiJie Cao
caolijie@dlou.edu.cn

[1] School of Information Engineering, Dalian Ocean University, Dalian 116023, China

[2] Huazhong University of Science and Technology, Wuhan 430000, China

Introduction

In recent years, deep learning has been continuously developed. Computer vision has been widely used in aquaculture, such as aquaculture seedling counting (Xue et al. 2023), aquatic product measurement, and fish behavior detection (Hong et al. 2014). Semantic segmentation is one of the crucial directions in deep learning research. It has been widely used in underwater vision tasks, for example, the length measurement of a fish body is achieved by image segmentation to get the outline of the fish and then extract the top center line of the fish body (Zhao et al. 2022). To obtain a more accurate boundary estimation of the fish, segmentation of the fish in the image was performed by Mask R-CNN (Garcia et al. 2020). The length of the shrimp was assessed using an image segmentation-based technique (Harbitz 2007). However, the aquatic product measurement estimation methods used above are based on the original image segmentation network to obtain the outline of the aquatic product segmentation, and the length estimation is performed pixel by pixel. The models used are not proposed for underwater scenarios, and the boundaries of aquatic products obtained from image segmentation are blurred, resulting in significant numerical errors when measurements are performed later.

As image segmentation is widely used in various scenarios, many scholars have proposed image semantic segmentation models for underwater scenarios. For example, an approach based on image contours and joint loss function gives high accuracy in underwater public datasets (Chicchon et al. 2023). A U-Net-based semantic segmentation method for underwater rubbish images is proposed to provide adequate localization information for underwater robots (Wei et al. 2022). Based on the DeepLabv3 + model, the Unsupervised Colour Correction Method (UCM) is incorporated to improve the images' quality and obtain higher boundary features (Liu and Fang 2020). Facing the lack of underwater semantic segmentation style datasets, a semantic segmentation of underwater images (SUIM) dataset is proposed, and a SUIM-Net model is presented (Jahidul, et al. 2020). Hambarde et al. (Hambarde et al. 2021) proposed an end-to-end underwater generative adversarial network (UW-GAN) for depth estimation of single underwater images. Dudhane et al. (Dudhane et al. 2020) proposed a novel end-to-end deep network for underwater image restoration using a channeled color feature extraction module, a dense residual feature extraction module, and a custom loss function. Liu et al. (Liu et al. 2022) proposed an underwater enhancement method based on object-guided dual-adversarial contrast learning to solve the problem that the enhanced image may not be conducive to detection effectiveness. Patil et al. (Patil et al. 2019) proposed a new unpaired motion saliency learning method for foreground segmentation in underwater videos by frame-by-frame motion saliency estimation using several initial video frames and corresponding frames. However, although it focuses on object contour accuracy in the literature mentioned above (Chicchon et al. 2023), it is semantic segmentation of high-resolution images. Literature (Wei et al. 2022) proposes a semantic segmentation method for underwater rubbish images based on U-Net, which focuses more on improving the segmentation accuracy of the network on the target. Literature (Liu and Fang 2020) improves segmentation boundary features by unsupervised color correction methods, but the number of parameters of this network is too large. In the literature (Hambarde et al. 2021; Dudhane et al. 2020; Liu et al. 2022; Patil et al. 2019), the detection accuracy is improved by the underwater image enhancement or restoration method, which does not improve the object contour boundary of underwater semantic segmentation. The above method does not solve the problem of boundary blurring in the image segmentation of aquatic products, which leads to the poor measurement

accuracy of aquatic products. Therefore, when facing the semantic segmentation task of aquatic product images, a network which can achieve high accuracy with clear and precise boundaries is needed.

To address the problems of aquatic product boundary blurring and segmentation accuracy in the semantic segmentation of underwater images, we propose an Underwater Image Semantic Segmentation Network model(UISS-Net). Firstly, the backbone network uses an auxiliary feature extraction network to improve the network feature extraction capability. Secondly, up-sampling feature fusion is performed using the multi-scale feature fusion network (MSFFN) proposed in this paper to enhance the input of shallow semantic information. Then, a channel attention mechanism is added to the feature fusion. Finally, a combination of the cross-entropy loss function and dice loss function is used as the loss function in this paper. Validation is performed on publicly available SUIM and Deep Fish datasets (Laradji et al. 2020) to evaluate the model performance. The experimental results show that the model proposed in this paper has higher detection accuracy and boundary accuracy when facing underwater images.

The subsequent work is organized as follows: the "Materials and methods" section presents the datasets used and the structure of the proposed UISS-Net model. The "Analysis of experimental results" section presents the data results of the model on the SUIM and Deep Fish public datasets. The "Conclusion" section draws the conclusion.

Materials and methods

Materials

Image datasets

In this paper, the performance of the model is validated by using the dataset SUIM proposed by Islam et al. (2020) for semantic segmentation of underwater images and the dataset Deep Fish proposed by Saleh et al. (2020) for underwater scenes. The SUIM dataset contains 1525 natural underwater images with their real semantic labels and a test set of 110 images collected during ocean exploration and human–robot cooperation experiments. The dataset is pixel-level annotated for eight object classes: background (waterbody) (BW), human divers (HD), aquatic plants and sea-grass (PF), wrecks or ruins (WR), robots (RO), reefs and invertebrates (RI), fish and vertebrates (FV), and sea-floor and rocks (SR). The Deep Fish dataset consists of approximately 40,000 underwater images collected from 20 Australian tropical marine environment habitats. The dataset initially contained only classification labels, and 300 fish semantic segmentation labels were added later. In this paper, the training and test sets are divided according to the ratio of 9:1. The images of the two datasets are shown in Fig. 1.

Environment setup

The experimental graphics card is NVIDIA GeForce RTX 4060Ti, PyTorch version 1.7, with 100 training epochs, and the optimizer was the Adam optimizer. Pre-training was performed on the ImageNet 1 K dataset using the transfer learning method, where the pre-training weights did not freeze the initial weights during the first five epochs of training. The number of batches is set to 6, the initial learning rate is set to 0.0004, the weights decay to 0.0001, the

Fig. 1 Dataset images. The first two rows of data are from SUIM dataset images and the last two rows of data are from Deep Fish dataset images

momentum is set to 0.5, and the learning rate uses a "COS" strategy that gradually decays with the number of iterations.

Experimental evaluation criteria

The semantic segmentation evaluation metrics are all calculated based on the confusion matrix as shown in Table 1.

In order to comprehensively evaluate the network architecture performance, we use mean intersection over union (mIoU), mean pixel accuracy (mPA), and accuracy as evaluation metrics. The equations for accuracy and mIoU are as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{FP+FP+FN+TN}} \tag{1}$$

$$\text{mIoU} = \frac{1}{N} \times \sum \left(\frac{\text{TP}}{\text{TP+FP+FN}} \right) \tag{2}$$

Table 1 Classification confusion matrix results

Current state	Predicted results	
	Positive	Negative
Positive	TP (true positive)	FN (false negative)
Negative	FP (false positive)	TN (true negative)

where N denotes the number of categories, TP denotes actual cases (number of pixels correctly predicted by the model as positive cases), FP denotes false-positive cases (number of pixels incorrectly predicted by the model as positive cases), and FN denotes false-negative cases (number of pixels incorrectly predicted by the model as negative cases).

The mPA is the proportion of the number of pixels correctly classified for each class calculated separately, then averaged cumulatively. Assuming that P is the accuracy of each category pixel, the mPA is given as follows:

$$\mathrm{mPA} = \frac{\mathrm{Sum}(P)}{N} \tag{3}$$

Methods

In semantic segmentation, the Backbone Network is the basis of the semantic segmentation model, which is responsible for extracting features (e.g., edges, texture, and shape) from the input image. A strong backbone network improves the model ability to generalize to different datasets, making the model adaptable to different semantic segmentation tasks. Cross-entropy loss measures the gap between the model prediction and the true labels, prompting the model to classify each pixel point more accurately. Dice loss emphasizes the overlap between predicted and real regions rather than simple classification correctness, which helps the model capture the full region of interest better while ensuring accuracy. Therefore, good backbone feature extraction network and loss function can greatly mention the boundary contour accuracy of semantic segmentation.

The overall architecture of the UISS-Net network in this paper references the U-shaped framework of U-Net (Ronneberger et al. 2015). The overall network architecture is shown in Fig. 2.

In Fig. 2, the encoder part performs feature extraction using an auxiliary network to obtain richer semantic information. The decoder part performs attentional feature fusion on the obtained feature map.

Encoder

In the face of tasks such as target detection or semantic segmentation in underwater scenes, low light and turbid water in underwater images are the main reasons for poor extracting features, and the difference between high-level and low-level semantic information is significant. Therefore, the encoder part of the UISS-Net network is mainly composed of the main backbone network and the auxiliary network.

The main backbone network uses ResNet50 as the base model. Also, pre-training weights of ResNet50 on the ImageNet 1 K dataset are used to give the network a faster convergence rate. The auxiliary network uses the lightweight GhostConv (Han et al. 2020) to generate more feature maps through cheap operations. GhostConv first performs a 1×1 convolution aggregating informative features between channels and then uses grouped convolutions to generate new feature maps. In order to reduce the amount of network computation, GhostConv divides the traditional convolution into two steps. Firstly, the traditional convolution generates feature maps with smaller channels to reduce the computation. Then, based on the obtained feature maps, it reduces the computation through the cheap operation, generates new feature maps, and splices the two groups of feature maps together to obtain the final. The traditional convolution operation is a combination of convolution-batch normalization BN-nonlinear activation. In contrast, linear transformation or cheap

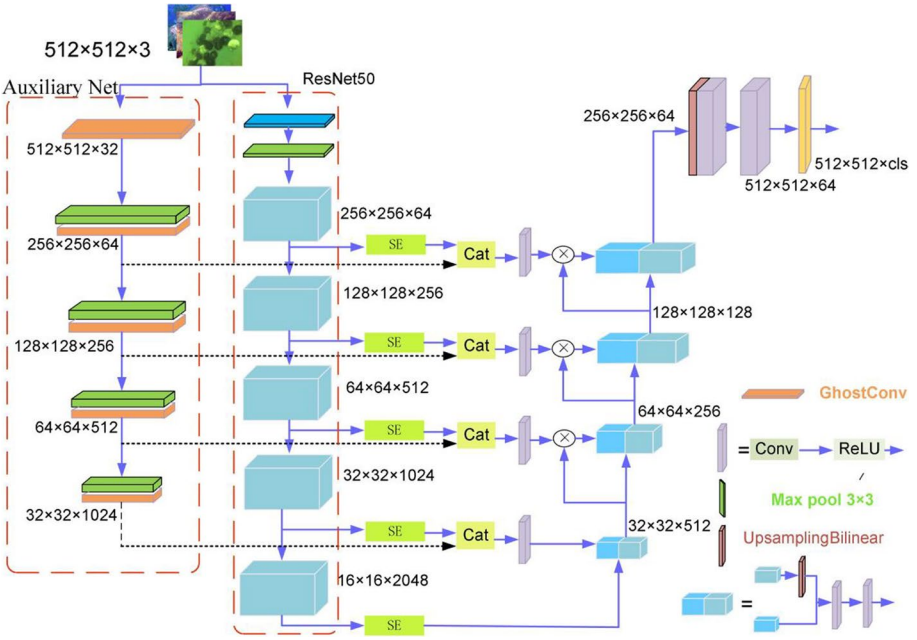

Fig. 2 UISS-Net network architecture

operation refers to ordinary convolution without batch normalization and nonlinear activation. The structure of the ghost module is shown in Fig. 3.

The use of auxiliary networks can enhance the robustness of the model and make it better adaptable to the input data of the underwater scene. At the same time, when the main backbone network lacks feature extraction capability in the feature extraction process, the auxiliary backbone network can make up for the missing part of the main backbone network and has the effect of multi-scale feature extraction. The auxiliary backbone network can learn a different representation of the features in the primary backbone network, resulting in a better generalization of the model. The overall structure of UISS-Net is shown in Table 2.

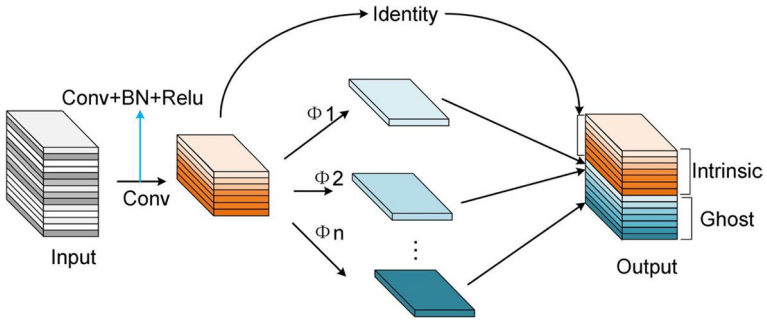

Fig. 3 Ghost module structure

Table 2 UISS-Net overall structure

Layer name	Output size	Main network	Auxiliary networks
Layer1 Conv1	256×256	7×7, 64, stride 2 3×3 max pool, stride2	GhostConv, 64 3×3 max pool, stride 2
Layer2 Conv2_x	128×128	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	GhostConv, 256 3×3 max pool, stride 2
Layer3 Conv3_x	64×64	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	GhostConv, 512 3×3 max pool, stride 2
Layer4 Conv4_x	32×32	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	GhostConv, 1024 3×3 max pool, stride 2
Layer5 Conv5_x	16×16	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	

Decoder

The UISS-Net network structure is a symmetrical U-shaped structure. Five bilinear interpolation samples were performed at up-sampling, respectively. As the high-level feature map has richer contextual information, the low-level feature map has richer spatial detail information. For both high-level and low-level features, simple feature fusion methods will ignore the feature map information diversity, leading to lower segmentation accuracy. Therefore, this paper proposes a feature fusion method for MSFFN derived from FPN (Lin et al. 2017) and PANet feature fusion architecture, as shown in Fig. 4.

The feature fusion method used in this paper is shown as d in Fig. 4. At the time of up-sampling, the features of this layer are fused with the encoder output features of the next layer by linear interpolation. The combination of pre-feature fusion through up-sampling expands the sensory field so that the model can better understand the input feature map context information.

In order to obtain better semantic information about the features, this paper incorporates the SE channel attention mechanism in the feature fusion. The weight of each channel of the feature map is obtained when performing feature fusion; then, this weight is used to assign a

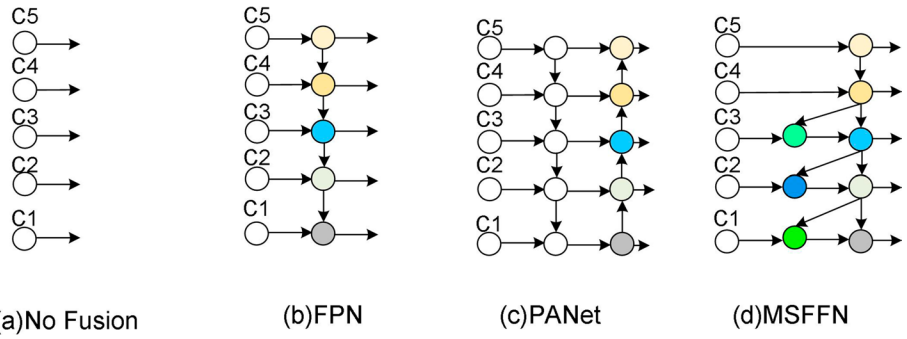

Fig. 4 Convergence of network architecture features

new weight value to each feature, thus allowing the neural network to focus on specific feature channels. The channels of the feature map that are useful for feature fusion are boosted, and the feature channels that are not very useful for the current task are suppressed.

Loss function

The loss function of UISS-Net uses the binary cross-entropy loss function and dice loss function (Milletari, et al. 2016). Equation 4 is the cross-entropy loss (CE_Loss) function, and Eq. 5 is the dice edge segmentation loss function. The cross-entropy function performs well in dealing with multi-classification problems and can effectively measure the difference between the model output and the actual labels. Furthermore, the cross-entropy function can help the model converge faster and achieve better performance with better robustness. However, cross-entropy loss requires high model stability and is sensitive to the sample equilibrium distribution. The dice loss function performs well in scenarios with severe imbalances of positive and negative samples and focuses more on mining the foreground region during training. Therefore, we use a combination of the cross-entropy loss function and the dice loss function to improve the convergence speed of the model when it is used to improve the model for dataset unevenness.

$$L_{CE} = -\sum_{i=0}^{N} y_i \ln\left(\sigma\left(x_i\right)\right) \tag{4}$$

$$\begin{aligned} &\text{Dice coefficent} = \frac{2TP}{2TP+FP+FN} \\ &L_{\text{dice loss}} = 1 - \text{Dice coefficent} \end{aligned} \tag{5}$$

The inputs to the cross-entropy loss function used in this paper are the model training outputs and the segmentation labels, as shown in Eq. 6.

$$L_1 = L_{\text{CE}}(\text{GT},\text{output}(x)) \tag{6}$$

where L_1 is the primary network loss function, GT is the accurate semantic tag, output(x) is the obtained segmentation result, and L_{CE} is the computation process of cross-entropy loss.

The segmentation loss function input used in this paper is the model training output with a pre-processed actual boundary image, as in Eq. 7.

$$L_2 = L_{\text{dice loss}}\left(GT_{(\text{seg})}, \text{output}\left(x_{\text{seg}}\right)\right) \tag{7}$$

where L_2 is the edge loss, $\text{GT}_{(\text{seg})}$ is the edge label obtained from the true semantic labels from the real semantic labels, $\text{output}(x_{(\text{seg})})$ is the edge extracted from the edge extracted from the segmentation result, and $L_{\text{dice loss}}$ is the computation procedure of dice loss.

UISS-Net's loss function is shown as follows:

$$L = L_1 + L_2 \tag{8}$$

Analysis of experimental results

In order to validate the performance of different major feature extraction networks in UISS-Net, ablation experiments were conducted in the validation set. In the experiment, the VGG backbone network output dimension is different from the auxiliary network,

Table 3 Ablation experiments with different master feature extraction networks

Model	Input Size	mIoU (%)	mPA (%)	Accuracy (%)
UISS-Net$_{(VGG)}$	512×512	65.51	77.29	82.68
UISS-Net$_{(ResNet18)}$	512×512	63.96	75.51	81.14
UISS-Net$_{(ResNet34)}$	512×512	68.89	77.4	85.9
UISS-Net$_{(ResNet50)}$	512×512	69.68	78.38	86.97

Table 4 Ablation experiments with different loss functions

NO	$L_{1(CE\_Loss)}$	$L_{2(dice\ loss)}$	mIoU (%)	mPA (%)	Accuracy (%)
1	√		70.73	79.56	86.82
2		√	68.12	76.54	85.72
3	√	√	72.09	80.37	86.93

so the matching process is performed on the VGG backbone network output dimension. The experimental results are shown in Table 3.

Although ResNet50 is more complex and requires more computational resources than models such as ResNet18, the training efficiency has been greatly improved with increased computational power and optimization algorithms. Compared to the VGG network, ResNet50 can capture more complex and subtle image features through its deep residual network structure. Therefore, ResNet50 is used in the UISS-Net main feature extraction network.

In order to verify the effectiveness of the improved loss function, ablation experiments are carried out for different loss function combination methods, and the results are shown in Table 4.

The ablation experiments in Table 4 show that combining the cross-entropy loss function with dice loss reduces the model dependence on a single loss function, improves the model robustness, and improves the model generalization ability.

In order to verify the method proposed in this paper, the backbone network of UISS-Net is replaced with ResNet50 as the baseline model. On the SUIM dataset, the auxiliary feature extraction module (AFEM), the multi-scale feature fusion module (MSFFM) proposed in this paper, the SE attention mechanism (SEA), and the boundary loss function (BLF) are added to the baseline model step by step, respectively. The results of the ablation experiments are shown in Table 5.

Table 5 shows that the auxiliary feature extraction network can significantly improve the model's performance. Also, the subsequent addition of the three modules can improve the model's performance. Finally, mIoU, mPA, and accuracy reach 72.09%, 80.37%, and 86.93%, which are 9.68%, 7.63%, and 0.98% higher than those of the baseline model. Figure 5 illustrates the loss of the UISS-Net model on the SUIM dataset.

From Fig. 5, it is shown that the model proposed in this paper has faster convergence than U-Net. Figure 6 shows explicitly the mIoU and mPA for each classification of SUIM data.

In order to further analyze the performance of the proposed method, we compared it with U-Net, FCN, SegNet, Deeplab, and so on. These semantic segmentation models are compared on SUIM and Deep Fish datasets, respectively.

Table 5 UISS-Net ablation experiments on the SUIM dataset

NO	AFEM	SEA	MSFFM	BLF	mIoU (%)	mPA (%)	Accuracy (%)
1					62.41	72.74	85.95
2	√				69.68	78.38	86.97
3	√	√			70.17	79.5	86.95
4	√	√	√		70.73	79.56	86.82
5	√	√	√	√	72.09	80.37	86.93

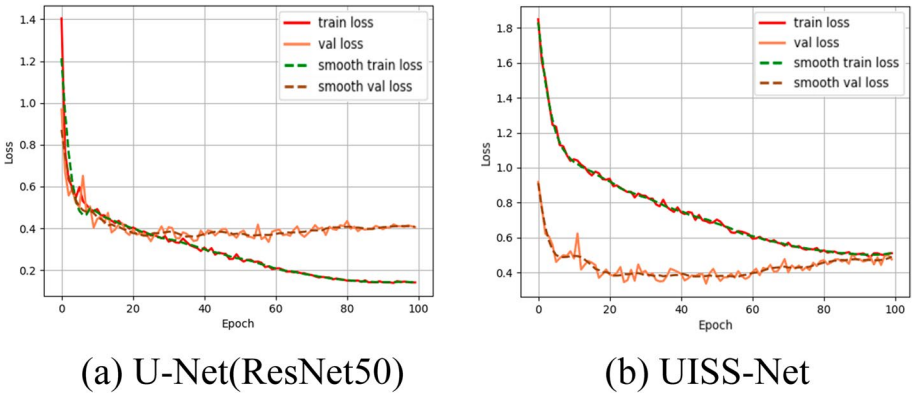

Fig. 5 Comparison of Loss of training results from U-Net (ResNet50) and UISS-Net models

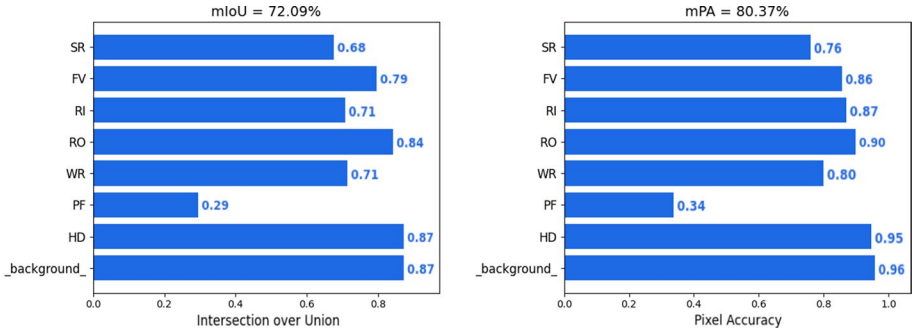

Fig. 6 shows the mIoU and mPA of the UISS-Net model for the SUIM validation dataset

Table 6 gives the mIoU of the SUIM dataset and the intersection over union (IoU) for each category, including the eight categories of water body background (waterbody) (BW), human divers (HD), aquatic plants and sea-grass (PF), wrecks or ruins (WR), robots (RO), reefs and invertebrates (RI), fish and vertebrates (FV), and sea-floor and rocks (SR).

Table 6 shows that the UISS-Net network achieves 72.09% of the mIoU indicator on the SUIM test set, which has a significant advantage over the classical networks such as U-Net,

Table 6 Comparison between the SUIM dataset and current mainstream models

Model	IoU/(%)								mIoU (%)
	BW	HD	PF	WR	RO	RI	FV	SR	
U-Net	79.46	32.25	21.85	33.94	23.65	50.28	38.16	42.16	39.85
U-Net$_{(ResNet)}$	90.14	72.53	2.37	62.65	59.19	69.93	73.13	69.31	62.41
U-Net$_{(VGG)}$	90	79	4	62	51	71	74	68	62.22
SegNet (Badrinarayanan et al. 2017)	80.63	45.67	17.45	32.24	55.72	47.62	43.92	51.51	46.85
SUIMNet (Islam et al. 2020)	80.64	63.45	23.27	41.25	60.89	53.12	46.02	57.12	53.22
PSPNet (Zhao et al. 2017)	82.51	65.04	28.54	46.56	62.88	55.80	46.78	55.98	55.51
Deeplab (Chen et al. 2018)	81.82	50.26	17.05	43.33	63.60	57.18	43.56	55.35	51.52
LEDNet (Wang, et al. 2019)	82.96	58.47	18.02	42.86	50.96	58.13	46.13	54.99	51.36
BiseNetv2 (Changqian et al. 2021)	83.67	59.29	18.27	39.58	56.54	58.16	47.33	56.93	52.47
UISS-Net (our)	87.18	87.03	29.48	71.27	84.11	70.70	79.44	67.54	72.09

Table 7 Comparison with current mainstream models in the Deep Fish dataset

Model	IoU (%)		mIoU/(%)
	Background	Foreground	
SUIM-Net (Islam et al. 2020)	99.03	78.40	88.71
SegNet (Badrinarayanan et al. 2017)	98.89	68.94	83.91
DeepLab-v3 (Chen et al. 2017)	99.11	71.35	85.23
PSPNet (Zhao et al. 2017)	99.15	72.61	85.88
FCN (Shelhamer et al. 2015)	99.21	66.30	82.75
HANet (Choi et al. 2020)	99.25	81.37	90.31
DGCNet (Zhang, et al. 2020)	99.21	81.42	90.32
MFAS-Net (Haider et al. 2022)	99.15	84.86	92.01
DPANet (Zhang et al. 2021)	99.31	82.56	85.88
UISS-Net (our)	99.55	90	95.05

SegNet, and PSPnet. Comparing the backbone network replaced by the U-Net and Deeplab networks, the proposed network has the highest accuracy, and the mIoU indicator is far superior to the current mainstream networks.

Table 7 shows the results of comparing the Deep Fish dataset with the current mainstream semantic segmentation models. Among them, our proposed UISS-Net model with mIoU of 95.05% has better segmentation results. The mIoU is improved by 12.3% compared to the traditional FCN semantic segmentation model.

We use the validation set of the SUIM dataset to perform predictions in the U-Net and UISS-Net models, respectively, whose results are shown in Fig. 7. From Fig. 7, it can be seen that UISS-Net can segment the underwater scene with high detection and recognition accuracy, and at the same time, the boundary precision and accuracy of the segmented underwater objects are higher than those of the traditional algorithms.

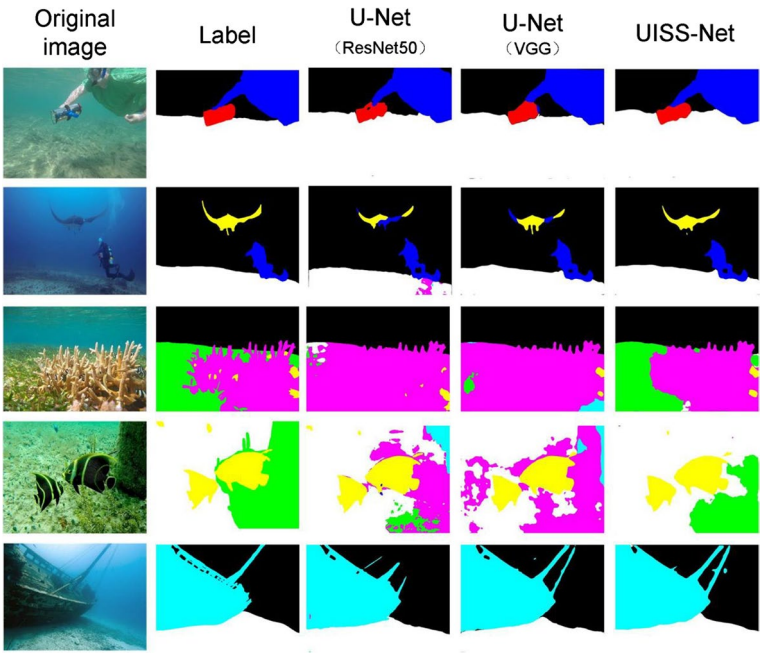

Fig. 7 Prediction results of different models on the SUIM validation set

Conclusion

This paper proposes an Underwater Image Semantic Segmentation Network (UISS-Net) for underwater scenes. Firstly, the use of an auxiliary feature extraction network is proposed to solve the problem of feature extraction difficulty caused by turbidity and insufficient light in underwater scenes. Second, an inverted multi-scale feature fusion approach is proposed to solve the problem of the difference in semantic information between the higher and lower layers of the network. Then, a channel attention mechanism is added to the feature fusion to prevent the loss of important feature information. Finally, a combined loss function is used to improve the model's accuracy with the problem of sample imbalance. The proposed network in this paper has an efficient backbone feature extraction module for the UISS-Net model compared to the methods in the literature, thus improving the semantic information extraction capability of the model. The proposed MSFFM feature fusion approach combines features from different levels, thus enhancing the model's ability to understand the data and make decisions. Combining the loss functions improves the model generalization ability to reduce the risk of overfitting.

The proposed UISS-Net model achieved 72.09% and 80.37% for mIoU and mPA in the SUIM dataset and 95.05% for mIoU in the Deep Fish dataset. Additional modules are embedded due to the backbone network and feature fusion of UISS-Net. Therefore, the number of model parameters is large and less efficient to process for mobile deployed devices (e.g., underwater robots). However, the proposed network solves the problems of poor performance of existing semantic segmentation models in underwater scenes and

rough segmentation boundaries, which is of great significance to the study of semantic segmentation of underwater images.

Author contribution ZhiQian He: conceptualization, methodology, software, visualization, writing the original draft. LiJie Cao: conceptualization, supervision, funding acquisition, review. JiaLu Luo: literature search, verification articles. XiaoQing Xu: literature search. JiaYi Tang: literature search. JianHao Xu: literature search. GengYan Xu: literature search. ZiWen Chen: review.

Funding Liaoning Provincial Education Department Scientific Research Funding Project, Grant/Award Number: LJKZ0731.

Data availability The UISS-Net model is available on GitHub at https://github.com/Hezhiqian97/UISS-Net.

Declarations

Consent for publication All authors have seen the manuscript and approved it to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Competing interests The authors declare no competing interests.

References

Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Changqian Yu, Gao C, Wang J et al (2021) BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation, International Journal of Computer Vision. Voi 129:3051–3068. https://doi.org/10.1007/s11263-021-01515-2

Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. arXiv.1–14. https://doi.org/10.48550/arXiv.1706.05587

Chen L-C, Papandreou G, Kokkinos I et al (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chicchon M, Bedon H, Del-Blanco CR, Sipiran I (2023) Semantic segmentation of fish and underwater environments using deep convolutional neural networks and learned active contours. IEEE Access 11:33652–33665. https://doi.org/10.1109/ACCESS.2023.3262649

Choi S, Kim JT, Choo J (2020) Cars cant fly up in the sky: improving urban-scene segmentation via height-driven attention networks. Proceedings of the Computer Vision and Pattern Recognition, Seattle, Online, USA, 2020 June 16-18. IEEE, New York, pp 9373–9383

Dudhane A, Hambarde P, Patil P, Murala S (2020) Deep underwater image restoration and beyond. IEEE Signal Process Lett 27:675–679. https://doi.org/10.1109/LSP.2020.2988590

Garcia R, Prados R, Quintana J, Tempelaar A, Gracias N, Rosen S, Vagstol H, Lovall K (2020) Automatic segmentation of fish using deep learning with application to fish size measurement. ICES J Mar Sci 77:1354–1366. https://doi.org/10.1093/icesjms/fsz186

Haider A, Arsalan M, Choi J, Sultan H, Park KR (2022) Robust segmentation of underwater fish based on multi-level feature accumulation. Front Mar Sci 9:1010565. https://doi.org/10.3389/fmars.2022.1010565

Hambarde P, Murala S, Dhall A (2021) UW-GAN: single-image depth estimation and image enhancement for underwater images. IEEE Trans Instrum Meas 70(5018412):1–12. https://doi.org/10.1109/TIM.2021.3120130

Han K, Wang Y, Tian Q et al (2020) GhostNet: more features from cheap operations. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, pp 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165

Harbitz A (2007) Estimation of shrimp (Pandalus borealis) carapace length by image analysis. ICES J Mar Sci 64(5):939–944. https://doi.org/10.1093/icesjms/fsm047
Hong H, Yang X, You Z, Cheng F (2014) Visual quality detection of aquatic products using machine vision. Aquacult Eng 63:62–71. https://doi.org/10.1016/j.aquaeng.2014.10.003
Islam MJ, Edge C et al (2020) Semantic segmentation of underwater imagery: dataset and benchmark. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, pp 1769–1776. https://doi.org/10.1109/IROS45743.2020.9340821
Laradji H, Konovalov DA, Bradley M et al (2020) A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. Sci Rep 10:14671. https://doi.org/10.1038/s41598-020-71639-x
Lin T-Y, Dollar P, Girshick R et al (2017) Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp 936–944. https://doi.org/10.1109/CVPR.2017.106
Liu F, Fang M (2020) Semantic segmentation of underwater images based on improved Deeplab. J Mar Sci Eng 8(3):188. https://doi.org/10.3390/jmse8030188
Liu R, Jiang Z, Yang S, Fan X (2022) Twin adversarial contrastive learning for underwater image enhancement and beyond. IEEE Trans Image Process 31:4922–4936. https://doi.org/10.1109/TIP.2022.3190209
Milletari, Fausto, Navab et al (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. Fourth International Conference on 3D Vision (3DV), Stanford, pp 565–571. https://doi.org/10.1109/3DV.2016.79
Patil PW, Thawakar O, Dudhane A, Murala S (2019) Motion saliency based generative adversarial network for underwater moving object segmentation. 2019 IEEE International Conference on Image Processing (ICIP), Taipei, pp 1565–1569. https://doi.org/10.1109/ICIP.2019.8803091
Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015. Lecture Notes in Computer Science. Springer, Cham, vol 9351. https://doi.org/10.1007/978-3-319-24574-4_28
Saleh A, Laradji IH et al (2020) A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. Sci Rep 10(14671). https://doi.org/10.1038/s41598-020-71639-x
Shelhamer E, Long J, Darrell T et al (2015) Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, pp 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
Wang Y et al (2019) Lednet: a lightweight encoder-decoder network for real-time semantic segmentation. 2019 IEEE International Conference on Image Processing (ICIP), Taipei, pp 1860–1864. https://doi.org/10.1109/ICIP.2019.8803154
Wei L, Kong S, Wu Y, Yu J (2022) Image semantic segmentation of underwater garbage with modified U-Net architecture model. Sensors 22(17):6546. https://doi.org/10.3390/s22176546
Xue B, Green R, Zhang M (2023) Artificial intelligence in New Zealand: applications and innovation. J Royal Society of New Zealand 53(1):1–5. https://doi.org/10.1080/03036758.2023.2170165
Zhang L et al (2020) Dual graph convolutional network for semantic segmentation. arXiv. IEEE, New York. https://doi.org/10.48550/arXiv.1909.06121
Zhang W, Wu C, Bao Z (2021) Dpanet: dual pooling-aggregated attention network for fish segmentation. Iet Comput Voi 16:67–82. https://doi.org/10.1049/cvi2.12065
Zhao H, Shi J, Qi X et al (2017) Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp 6230–6239. https://doi.org/10.1109/CVPR.2017.660
Zhao Y-P, Sun Z-Y, Hai D, Bi C-W, Meng J, Cheng Y (2022) A novel centerline extraction method for overlapping fish body length measurement in aquaculture images. Aquacult Eng 99:102302. https://doi.org/10.1016/j.aquaeng.2022.102302

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com