frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 23 May 2024
DOI 10.3389/fmars.2024.1411717

OPEN ACCESS

EDITED BY
Haiyong Zheng,
Ocean University of China, China

REVIEWED BY
Xuebo Zhang,
Northwest Normal University, China
Zhao Shengrong,
Qilu University of Technology, China
Xuekai Wei,
Chongqing University, China

*CORRESPONDENCE
Hua Li
lihua@hainanu.edu.cn

RECEIVED 03 April 2024
ACCEPTED 01 May 2024
PUBLISHED 23 May 2024

CITATION
Wang C, Duan W, Luan C, Liang J, Shen L and Li H (2024) USNet: underwater image superpixel segmentation via multi-scale water-net.
Front. Mar. Sci. 11:1411717.
doi: 10.3389/fmars.2024.1411717

USNet: underwater image superpixel segmentation via multi-scale water-net

Chuhong Wang[1,2], Wenli Duan[3], Chengche Luan[3], Junyan Liang[3], Lengyu Shen[3] and Hua Li[3]*

[1]School of Information and Communication Engineering, Hainan University, Hainkou, China, [2]School of Electronic Information Engineering, Guangdong Ocean University, Zhanjiang, China, [3]School of Computer Science and Technology, Hainan University, Haikou, China

Underwater images commonly suffer from a variety of quality degradations, such as color casts, low contrast, blurring details, and limited visibility. Existing superpixel segmentation algorithms face challenges in achieving superior performance when directly applied to underwater images with quality degradation. In this paper, to alleviate the limitations of superpixel segmentation when applied to underwater scenes, we propose the first underwater superpixel segmentation network (USNet), specifically designed according to the intrinsic characteristics of underwater images. Considering the quality degradation, we propose a multi-scale water-net module (MWM) aimed at enhancing the quality of underwater images before superpixel segmentation. The degradation-aware attention (DA) mechanism is then created and incorporated into MWM to solve light scattering and absorption, which can decrease object visibility and cause blurred edges. By effectively directing the network to prioritize locations that exhibit a considerable decrease in quality, this method enhances the visibility of those specific areas. Additionally, we extract the deep spatial features using the coordinate attention method. Finally, these features are fused with the shallow spatial information using the dynamic spatiality embedding module to embed comprehensive spatial features. Training and testing were conducted on the SUIM dataset, the underwater change detection dataset, and UIEB dataset. Experimental results show that our method achieves the best scores in terms of achievable segmentation accuracy, undersegmentation error, and boundary recall evaluation metrics compared to other methods. Both quantitative and qualitative evaluations demonstrate that our method can handle complicated underwater scenes and outperform existing state-of-the-art segmentation methods.

KEYWORDS

superpixel segmentation, underwater images, image enhancement, spatial information fusion, neural network

1 Introduction

Over the past decades, underwater image processing has garnered considerable attention, since it plays a vital role in all kinds of underwater practical applications (Li et al., 2019; Peng et al., 2023; Zhou et al., 2023), including marine biology and archaeology (Arnaubec et al., 2023; Calantropio and Chiabrando, 2024), marine ecology (Strachan, 1993; Catalan et al., 2023), underwater internet of things (Qiu et al., 2019), and underwater acoustic field (Yang, 2023; Zhang et al., 2024). While underwater image processing is crucial in these fields, it faces significant challenges due to the inherent quality degradation in underwater environments. These degradations, including color casts, low contrast, and blurred details, greatly hinder the performance of image processing algorithms tailored to natural, terrestrial conditions. Seeking to effectively alleviate the problem of quality degradation and develop efficient algorithms for underwater image segmentation processing is a major challenge. If it can be solved, it will greatly enhance the potential and practicality of underwater image applications. Therefore, developing robust adaptive algorithms that can cope with the adverse effects of color shift, contrast reduction, and detail loss is critical to unlocking the full spectrum of underwater image data for computer vision tasks and practical applications.

In recent years, computer vision technology has advanced rapidly. In the prediction of 3D visual saliency, the Multi-input Multi-output Generative Adversarial Network (Song et al., 2023) proposed leverages 2D image saliency and 3D object categorization to enhance the accuracy of saliency prediction, offering new insights into human visual perception in 3D environments. In the video description, the Reconstruction Network (Zhang et al., 2019b) has been proposed to enhance the natural language description of video content by employing an encoder–decoder–reconstructor architecture that leverages bidirectional flows between visual information and textual representation, significantly boosting the performance of video captioning tasks. However, superpixels are compact groups of pixels that share similar low-level visual properties such as color, texture, and contrast. They are commonly used in computer vision and image processing tasks (Kumar, 2023; Barcelos et al., 2024) as an intermediate representation of an image, which is more perceptually meaningful than individual pixels. Superpixel segmentation is a computer vision technique that involves grouping pixels with color, texture, and other low-level properties into regions or clusters that perceptually belong together while drastically lowering the number of primitives for downstream tasks, such as saliency (Cong et al., 2017a, b, 2019), object tracking (Kim et al., 2019), image enhancement (Fan et al., 2017; Subudhi et al., 2021), image reconstruction (Fan et al., 2018b; Li et al., 2020), and optical flow (Sultana et al., 2022).

For superpixel segmentation of underwater images, existing superpixel segmentation algorithms are challenging to achieve superior performance due to the quality degradation. To be more specific, distinguishing between object and background colors poses a significant challenge for algorithms due to the presence of color casts and low contrast in underwater scenes. These issues further complicate the accurate adherence of object boundaries by the algorithm. To alleviate this issue, we design a multi-scale water-net module (MWM), which is used to enhance the quality of underwater images before superpixel segmentation. Compared to the water-net (Li et al., 2019), we introduce a U-shape architecture model instead of convolutional neural networks (CNNs) to obtain both high-resolution coarse-grained features and low-resolution fine-grained features. These features can facilitate the generation of more accurate enhanced results. Moreover, one of the main challenges in underwater environments is the loss of fine details and boundaries caused by light scattering and absorption in water, resulting in low visibility and blurry edges of objects. We have also designed a novel degradation-aware attention (DA) incorporated into MWM, which effectively guides the network to prioritize regions with notable quality degradation, thereby enhancing the visibility of those areas.

As the simple comparison shown in Figure 1, we can see that our proposed USNet adheres to the object boundaries more accurately and can better distinguish between object and background than the state-of-the-art superpixel segmentation method (Wang et al., 2021).

The main contributions of the paper can be highlighted as follows:

1. We propose an end-to-end superpixel segmentation network (i.e., USNet), which is designed based on the characteristics of

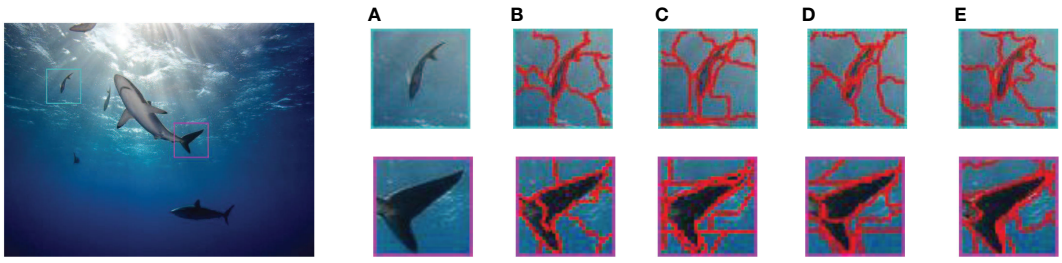

FIGURE 1
A simple comparison of the state-of-the-art superpixel segmentation methods. **(A)** The original image. **(B)** The SLIC method designed for nature images. **(C)** The FCN method designed for nature images. **(D)** The AINet method designed for nature images. **(E)** Our USNet method designed for underwater images.

underwater images. The framework can reduce the negative influence of quality degradation and generate uniform and compact superpixels by considering shallow and deep spatial features in the meantime. To the best of our current knowledge, this is the first attempt to devise a deep superpixel segmentation network for underwater images.

2. We design a multi-scale water-net module (MWM) to enhance the quality of underwater images before superpixel segmentation.

3. We design novel degradation-aware attention (DA) to enforce the network to pay more attention to quality-degraded regions, and the DA is embedded in MWM.

4. Extensive experiments on different datasets demonstrate that our proposed USNet achieves state-of-the-art performance both qualitatively and quantitatively. We also perform elaborate ablation studies to validate the effectiveness of each component in our network.

2 Related work

Superpixel segmentation is a pivotal technique in the realm of computer vision and image processing, aimed at partitioning an image into a set of compact and nearly homogeneous regions known as superpixels. These regions are characterized by their similarity in terms of color, texture, and contrast, which makes them particularly useful for a variety of applications such as object recognition, image segmentation, and feature extraction. The process of superpixel segmentation can be broadly categorized into two main approaches: traditional and supervised methods, each with its own set of algorithms and characteristics.

Traditional methods rely on handcrafted features extracted to partition or measure the similarity between pixels to group them into clusters. The normalized cut (Ncut) (Shi and Malik, 2000) algorithm is a graph-based superpixel segmentation method that creates a pixel graph using color and spatial proximity to determine edge weights. However, parameter tuning can be time consuming, and it may not perform well on images with significant variations in texture or lighting. Simple linear iterative clustering (SLIC) (Achanta et al., 2012) is a superpixel segmentation method that employs a regular grid of candidate centers to group pixels based on their color similarity and spatial distance to the nearest center. Although SLIC is computationally efficient and generates high-quality superpixels, it may encounter difficulties in accurately segmenting complex structures and sharp contrast boundaries. Bayesian adaptive superpixel segmentation (BASS) (Uziel et al., 2019) is a method that uses Bayesian inference to estimate the image structure and adjust the number and shape of superpixels adaptively. Its objective is to strike a balance between over-segmentation and under-segmentation.

For a considerable duration, superpixel segmentation has not advanced toward an end-to-end trainable algorithm due to the non-differentiability of the nearest neighbor operation required for computing pixel superpixel associations. Superpixel Sampling Networks (SSNs) (Jampani et al., 2018) addressed this issue by calculating soft pixel–superpixel associations instead of hard associations. Moreover, Superpixel Segmentation with Fully Convolutional Networks (FCN) (Yang et al., 2020) integrates feature extraction and superpixel segmentation into a single step, making it faster and more readily integrable with existing Convolutional Neural Network (CNN) frameworks for downstream tasks. Furthermore, AINET (Wang et al., 2021) proposes the Association Implantation (AI) module, the AI module directly predicts the relationship between pixels and superpixels, instead of predicting the pixel–pixel relationship like FCN.

3 The proposed method

Our proposed USNet introduces several innovative features that differentiate it from traditional superpixel segmentation methods, especially in the context of underwater image processing. USNet is an end-to-end trainable framework designed to address the unique challenges of underwater imagery. Traditional superpixel segmentation techniques often fail to consider specific degradations inherent in underwater environments, such as color casts, low contrast, and blurred details, which can severely affect the performance of segmentation algorithms. In contrast, USNet introduces specialized methods to alleviate these problems, as shown in Figure 2. It adopts the MWM module and introduces a DA mechanism to adapt to degraded areas in the image, aiming to improve the quality of underwater images before the segmentation process begins. In addition, the DSFE module better captures spatial details and uses a CA mechanism to extract deep spatial features, focusing on the maintenance of spatial consistency. These features are then input into DSEM, which normalizes and fuses shallow and deep spatial features, using a dynamic fusion mechanism to adaptively adjust the weight of feature representations. This comprehensive approach ensures that the network can effectively identify and exploit spatial information to generate compact and uniform superpixels, even in complex underwater visual environments. Considering the uniqueness of image enhancement and superpixel segmentation tasks, USNet uses different loss functions for the two tasks and optimizes them separately to avoid mutual interference between the two tasks.

3.1 Multi-scale water-net module

Multi-scale water-net module is the key component to enhance the quality of underwater images, which is inspired by water net (Li et al., 2019), because of its impressive performance and simple but efficient architecture. In the module, we try to alleviate the limitations of superpixel segmentation applied to underwater images, which are the negative influences of various quality degradation existing in underwater scenes. We will introduce various negative influences of quality degradation that exist in underwater images and how we reduce them through multi-scale water-net module below. The details of the multi-scale water-net module are shown in Figure 2. Through a U-shaped structure

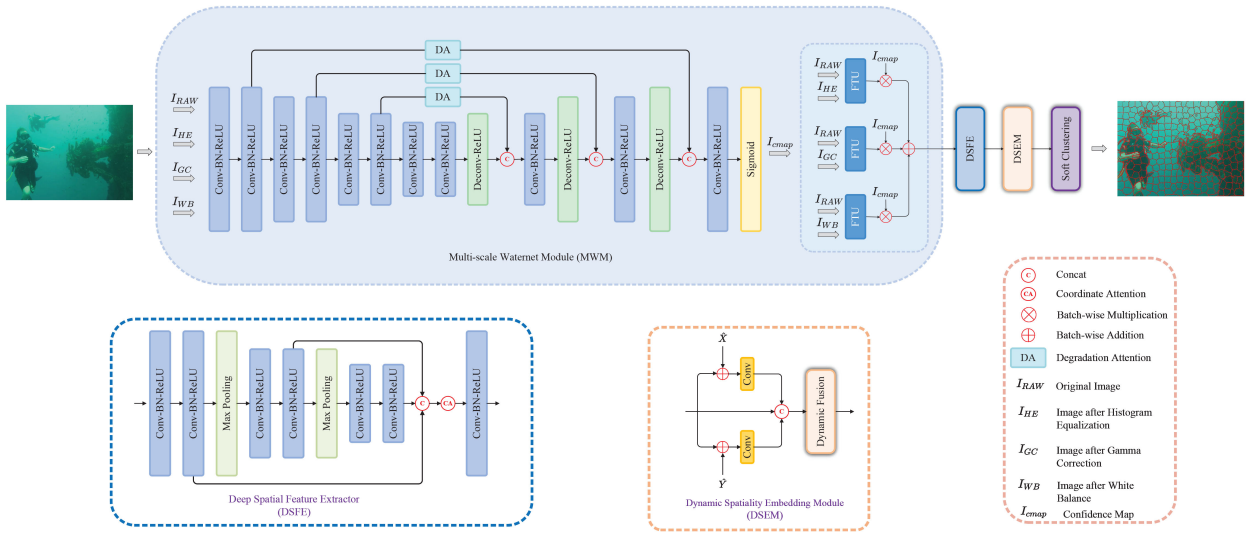

FIGURE 2
The overview architecture of the proposed USNet. It includes the MWM module for preprocessing and enhancing underwater images, the DSFE module for extracting deep spatial features of images, and the DSEM module to fuse and optimize these features to achieve super-pixel segmentation based on underwater image characteristics.

network, high-resolution coarse-grained features and low-resolution fine-grained features of the image are extracted simultaneously. This design enables the network to understand the image content more comprehensively and thus recover details and boundary information more effectively during the image enhancement process. In MWM, a DA mechanism is also incorporated, which guides the network to pay more attention to areas with significant quality degradation in the image. Note that Feature Transformation Unit is used to refine the inputs; more details can be found in Li et al. (2019).

Underwater images are often characterized by low contrast, dark regions, and color casts due to the optical properties of water. To address these issues, histogram equalization (HE) is employed to enhance the contrast by redistributing the brightness levels across the entire dynamic range. This method effectively lightens the dark areas and makes the details more distinguishable, which is crucial for tasks such as object detection and scene analysis. By increasing the contrast, HE ensures that the full range of tones from dark to bright is represented, which leads to a more visually appealing and informative image. Underwater lighting conditions can cause the camera sensor to capture images with a non-linear response, resulting in a loss of detail in the mid to dark tones. Gamma correction (GC) solves this problem by applying a non-linear transformation to the pixel values, which helps restore the perceived brightness and improves the overall visual quality of the image, making it easier to distinguish different elements in the scene. The absorption of red light by water results in a blue or green tint, which can distort the true colors of the underwater image. White-balancing (WB) algorithms estimate the color cast and adjust the color components to neutralize it, aiming to recreate the colors as they would appear under daylight conditions. By correcting the color cast, WB enhances the image's visual fidelity, making it more suitable for further analysis and interpretation. Therefore, we preprocess underwater images through HE, GC, and WB algorithms to improve contrast and detail visibility, illuminate dark areas, and correct color casts, respectively.

Then, the generated images will be concatenated with the original image in the channel dimension as input. Taking into account the straightforward architecture of the CNN employed in water net, which restricts their perceptual capabilities to a limited region, we design a U-shape model to enlarge the receptive field for generating more accurate enhanced result. The model adopts an encoder–decoder architecture. The basic block in encoder consists of "Conv-BN-ReLU", while the basic blocks in decoder consist of "Deconv-ReLU" and "Conv-BN-ReLU". Moreover, we strategically channeled the extracted features through the DA mechanism, nestled between the encoder and decoder. This astute implementation effectively compels the network to allocate heightened attention toward regions afflicted by quality degradation. Furthermore, to counteract the peril of gradient vanishing (He et al., 2016), we judiciously incorporated skip connections, which seamlessly bridge information across different network layers. Finally, after the sigmoid activation function layer, the confidence map will be generated to select the most significant features of inputs to achieve the enhanced result by fusing with the output of Feature Transformation Units.

3.2 Degradation-aware attention

In view of several regions with severe quality degradation reducing the overall performance of subsequent tasks to a large extent, we devise degradation-aware attention (DA) to enforce the network to pay more attention to quality-degraded regions. Figure 3 shows the effectiveness of our degradation-aware attention. More specifically, our proposed degradation-aware attention comprises a

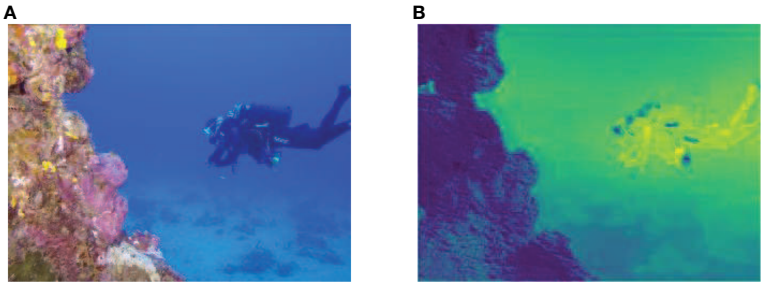

FIGURE 3
The effectiveness of the degradation-aware attention. **(A)** The original image. **(B)** The visualization of deep features after applying degradation-aware attention. We can see that the proposed degradation-aware attention can highlight the severe quality degradation regions while focusing on the main body of the image.

cascade of convolutional block attention module (CBAM) (Woo et al., 2018) and pixel attention (PA) (Qin et al., 2020) in terms of the architecture.

As shown in the schematic illustration of the proposed degradation-aware attention in Figure 4, the deep feature $\mathcal{F} \in \mathbb{R}^{N\times H\times W}$ extracted by "Conv-BN-ReLU" block is fed to this module. Aggregating the most crucial information in the channel and spatial dimensions using CBAM to guarantee optimal results in subsequent operations. Then, the essential deep feature $\mathcal{F}'' \in \mathbb{R}^{N\times H\times W}$ is obtained as follows:

$$\mathcal{F}' = M_c(\mathcal{F}) \otimes \mathcal{F} \tag{1}$$

$$\mathcal{F}'' = M_s(\mathcal{F}') \otimes \mathcal{F}' \tag{2}$$

In Equations 1 and 2, $\otimes$ denotes element-wise multiplication. $M_c \in \mathbb{R}^{N\times 1\times 1}$ denotes a channel attention map, while $M_s \in \mathbb{R}^{1\times H\times W}$ denotes a spatial attention map; they can be formulated as follows:

$$M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AP}(F)) + \mathrm{MLP}(\mathrm{MP}(F))) \tag{3}$$

$$M_s(F) = \sigma(Conv(Concat(\mathrm{AP}(F), \mathrm{MP}(F)))) \tag{4}$$

In Equations 3 and 4, $\sigma(\cdot)$ denotes the Sigmoid activation function. *AP* denotes average-pooling, *MP* denotes max-pooling, while *MLP* denotes multi-layer perception. *Conv* and *Concat* represent the convolution operation and concatenate operation on channel dimension, respectively.

Then, we utilize PA to enforce the network to pay more attention to regions with severe quality degradation, such as thick-hazed or blurring regions, which can be formulated as Equations 5 and 6, where δ represents the ReLU activation function.

$$PA = \sigma(\mathrm{Conv}(\delta(\mathrm{Conv}(\mathcal{F}'')))) \tag{5}$$

$$\mathcal{U} = \mathcal{F}'' \otimes PA \tag{6}$$

3.3 Deep spatial feature extractor

After enhancing the quality of underwater images, we extract the deep features of images for subsequent superpixel segmentation. Considering the importance of spatial information for generating compact and uniform superpixels, the original RGB image is converted to the CIELAB color space following Jampani et al. (2018). Next, we extract shallow spatial features and concatenate them with the image on the channel dimension, resulting in a new feature map that is used for subsequent deep feature extraction. The CA mechanism (Hou et al., 2021) is utilized to extract the deep spatial features.

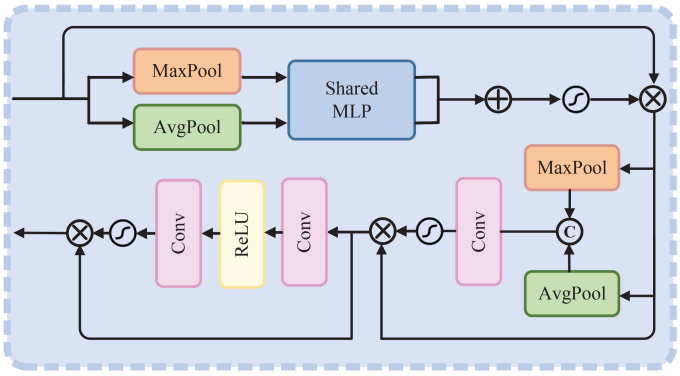

FIGURE 4
The schematic illustration of degradation-aware attention. Note that the channel number of input is the same as output.

However, the shallow spatial features may interfere with the CA to extract the deep spatial features; CA only extracts deep spatial features for the image itself rather than its shallow spatial features. For this reason, the shallow spatial features have been removed from the input item. After that, a CNN is designed to extract the deep features of the input image; the basic block "Conv-BN-ReLU" consists of a convolution layer with 3 × 3 kernel size and batch-normalization layer and a ReLU activation function layer. Next, the deep features after two basic blocks are downsampled through the max-pooling layer.

Specifically, let $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{C\times H\times W}$ be the output of upsamping, where C, H, and W are the number of channels, height, and width of the image, respectively. We first obtain direction-aware feature maps z^h and z^w by aggregating features along vertical and horizontal directions as Equations 7 and 8, respectively.

$$z^h(h) = \frac{1}{W}\sum_{0\le i<W} X(h,i), 0 \le h < H \tag{7}$$

$$z^w(w) = \frac{1}{H}\sum_{0\le j<H} X(j,w), 0 \le w < W \tag{8}$$

Then, the spatial features in vertical and horizontal directions (X and Y) are encoded by convolution operation; the intermediate feature map $f \in \mathbb{R}^{C/r\times(H+W)}$ can be achieved as follows:

$$f = \sigma(Conv_{1\times1}(Concat(z^h, z^w))) \tag{9}$$

In Equation 9, σ is the Sigmoid activation function layer, and $Conv_{1\times1}$ is a convolutional layer with 1 × 1 kernel size. r is the reduction ratio to reduce the complexity of the model.

To ensure that the channel number of feature maps is equal to input X, we first split f into vertical feature map $f^h \in \mathbb{R}^{C/r\times H}$ and horizontal feature map $f^w \in \mathbb{R}^{C/r\times W}$, then convert the channel number of f^h and f^w to input X as follows:

$$g^h = \sigma(Conv_h(f^h)) \tag{10}$$

$$g^w = \sigma(Conv_w(f^w)) \tag{11}$$

In Equations 10 and 11, $Conv_h$ and $Conv_w$ are the convolutional layer with 1 × 1 kernel size. Finally, the deep spatial features can be obtained with the guidance of coordinate attention weight, which can be formulated as Equation 12:

$$\widehat{F}(i,j) = X(i,j) \otimes g^h(i) \otimes g^w(j) \tag{12}$$

where $\hat{F}$ represents the deep spatial features; more details about coordinate attention can be seen in (Hou et al., 2021).

3.4 Dynamic spatiality embedding module

In our previous work (Li et al., 2023), we design the DSEM to achieve comprehensive spatial features to handle various complicated underwater scenes and generate compact and regular superpixels. In this paper, we continue to adopt this method. The schematic illustration of DSEM is shown in Figure 2.

More specifically, first, the shallow spatial features are normalized to prevent an over-consideration of spatial information, since the value of spatial features for a high-resolution image may be too large, which will pollute the image feature representation. Then, $\hat{X}$ and $\hat{Y}$ are added to the deep spatial features and respectively fed to two convolution layers with 1×1 kernel size to fuse the shallow and deep spatial features, and embed the comprehensive spatial features to the network. After that, the features with comprehensive spatial information in vertical and horizontal directions will be concatenated on channel dimension, then the weighting of spatial features will be adaptively adjusted through the dynamic fusion (DF) mechanism to obtain a more effective representation of spatial features. The DF mechanism is based on the channel attention mechanism, the schematic illustration of which has been shown in Figure 5.

3.5 Loss functions

Due to the low correlation between the task of underwater image enhancement and superpixel segmentation, if the loss functions of two tasks are combined for joint training, the loss function of underwater image enhancement will interfere with superpixel segmentation, resulting in erroneous segmentation results. Thus, the task of underwater image enhancement and superpixel segmentation are back-propagated separately.

3.5.1 Underwater image enhancement

Following the previous work (Li et al., 2019), a linear combination of ℓ_1 loss L_{ℓ_1} and the perceptual loss L_{per} is utilized to ensure the high quantitative scores and facilitate the model to produce visually pleasing and realistic results.

The ℓ_1 loss is used to measure the global similarity between the enhanced results $\hat{I}$ and corresponding ground truth I. The calculation formula is as Equation 13, where H and W represent the height and width of the image, respectively.

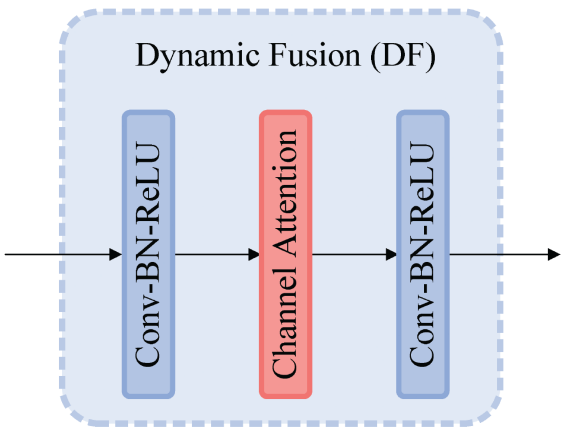

FIGURE 5
The schematic illustration of Dynamic Spatiality Embedding Module.

$$L_{\ell_1} = \sum_{x=1}^{H}\sum_{y=1}^{W}\left|\hat{I}(x,y) - I(x,y)\right| \tag{13}$$

To alleviate the artifact induced by pixel-wise loss function (e.g., ℓ_1 loss) (Li et al., 2019), perceptual loss is introduced to facilitate the enhanced results more visually pleasing and realistic. In addition, perceptual loss also can constrain the consistency between the enhanced results and ground truth to prevent over-enhancement (Ni et al., 2020), which can be formulated as Equation 14:

$$L_{per} = \sum_{x=1}^{H}\sum_{y=1}^{W}\left\|\phi_n(\hat{I})(x,y) - \phi_n(I)(x,y)\right\| \tag{14}$$

where $\phi_j(\cdot)$ represents the jth convolutional layer of a VGG-19 network pretrained on ImageNet dataset (Deng et al., 2009). The final loss of underwater image enhancement $L_{enhance}$ can be obtained as Equation 15, where λ_1 is set to 0.05 following the setting (Li et al., 2019).

$$L_{enhance} = L_{\ell_1} + \lambda_1 L_{per} \tag{15}$$

3.5.2 Superpixel segmentation

To supervise the training of the network for superpixel segmentation, we combine two loss functions to make it more sufficient.

The semantic loss L_{sem} helps to adhere to semantic boundaries, optimize the segmentation process, and improve the accuracy of downstream tasks that rely on semantic segmentation, as follows:

$$L_{sem} = L(R, R^*) \tag{16}$$

In Equation 16, $L(\cdot,\cdot)$ stands for cross-entropy loss, R represents the one-hot semantic label of ground truth, and R^* represents the restructed semantic label.

The compactness loss $L_{compact}$ can ensure that the superpixels are spatially coherent, which is defined as the following ℓ_2 norm:

$$L_{compact} = \left\|I_{xy} - \hat{I}_{xy}\right\|_2 \tag{17}$$

In Equation 17, I_{xy} denotes the spatial information of original image, and $\hat{I}_{xy}$ denotes the spatial information after reconstructing. The overall loss of superpixel segmentation can be formulated as Equation 18, where λ_2 is set to 0.4 according to extensive experience.

$$L_{segment} = L_{\text{sem}} + \lambda_2 L_{\text{compact}} \tag{18}$$

4 Experiments and results

4.1 Experimental setup

4.1.1 Datasets

In our experiment, we use SUIM dataset (Islam et al., 2020), underwater change detection dataset (Radolko et al., 2016), UIEB dataset (Li et al., 2019) for training and testing. SUIM is a large-scale and popular underwater image dataset with semantic annotations, which contains over 1,500 images and includes eight object categories. Underwater change detection dataset contains videos of five scenes, including caustics, fish swarm, two fishes, marine snow, and small aquaculture. Each video contains 1,100 frames and provides semantic annotations of the last 100 frames for evaluation. UIEB is a real-world underwater image enhancement dataset, which contains 950 real underwater images, of which 890 images are provided with corresponding ground truth. This dataset is very popular in the field of underwater image enhancement.

Considering that if the underwater image enhancement module is not fully trained, the network may produce unsatisfactory superpixel segmentation results. First, the Multi-scale water-net module is pretrained using 890 semantically annotated images from UIEB for 20K iterations. Next, in terms of supervised learning for both tasks, 1,040 images from the SUIM training set with size 640×480 were used for training. However, these images lack the ground truth of underwater image enhancement and cannot simultaneously provide ground truth for both superpixel segmentation and underwater image enhancement tasks. Therefore, we utilize the state-of-the-art underwater image enhancement method Ucolor (Li et al., 2021) to generate the enhancement results of the training data and carefully select 1,040 ground truth images with good enhancement effects. Finally, 110 images of size 640 × 480 from the SUIM test set and 100 images of size 1,920 × 1,080 from the caustic scene are used for testing.

4.1.2 Evaluation metrics

We employed three commonly used evaluation metrics, namely, achievable segmentation accuracy (ASA), undersegmentation error (UE), and boundary recall (BR), to assess the performance of our model in our experiments. ASA measures the similarity between the ground truth segmentation and the superpixel segmentation. It measures the percentage of pixels that are correctly assigned to the corresponding superpixel in the ground truth segmentation. BR measures the accuracy of the superpixel boundaries by calculating the fraction of correctly overlapped pixels of superpixel segments from the ground truth boundaries. UE measures the extent to which a superpixel algorithm fails to segment an image accurately. It calculates the fraction of pixels that are not assigned to a superpixel in the ground truth segmentation but are assigned to a superpixel in the superpixel segmentation. These metrics are commonly used in the field (Jampani et al., 2018; Yang et al., 2020; Wang et al., 2021; Li et al., 2023) and serve as reliable indicators of the accuracy and effectiveness of superpixel segmentation algorithms (detailed definitions in Stutz et al., 2018).

4.1.3 Implementation details

During the training stage, the original images are randomly cropped to size 200 × 200 as input and horizontal and vertical flipping is performed for data augmentation. Since underwater image enhancement and superpixel segmentation are back-propagated separately, two optimizers based on Adam with

default parameters (Kingma and Ba, 2015) (β_1 = 0.9 indicating that the current gradient information is slightly more significant than the past gradients in updating the parameters, β_2 = 0.999 indicating that the optimizer is quite sensitive to recent changes in the gradient's scale) are used to respectively optimize the modules of two tasks. The learning rate of the optimizer for underwater image enhancement is set to 1e−3 and decreases by 0.1 every 5K iterations, while the initial learning rate of the optimizer for superpixel segmentation is set to 2e−4 and decreases by half every 2k iterations, then the learning rate is fixed to 1e−5 after 10K iterations. A batch-mode learning method with a batch size of 8 is applied. In addition, superpixels are enforced to be spatially connected to follow (Jampani et al., 2018; Yang et al., 2020) for fair comparison. All experiments are implemented by PyTorch framework on a PC with NVIDIA RTX 2080 Ti GPU.

4.2 Comparison with state-of-the-art methods

In our evaluation, we compare our methods against other state-of-the-art methods, including SSN (Jampani et al., 2018), FCN (Yang et al., 2020), AINET (Wang et al., 2021), SLIC (Achanta et al., 2012), SNIC (Achanta and Süsstrunk, 2017), and our methods on SUIM and caustics scene in underwater change detection dataset, respectively. For a fair comparison, we adopt the parameter settings used in the original works and implement all methods using either the code provided by (Soomro and Wang, 2017) or the code from the original authors.

4.2.1 Quantitative comparison

The quantitative comparison results of our proposed method and other state-of-the-art methods test on SUIM and caustics scene in underwater change detection dataset are shown in Figures 6, 7, while other state-of-the-art methods keep their original training dataset BSDS500. In Figure 6, we can see that our method achieves the top score on all mentioned metrics of the SUIM and caustics datasets. Taking 300 superpixels in SUIM dataset for example, our method's minimum percentage gain (computed with the highest score of the compared methods) of UE is 10.6%, while that of BR is 5.7%. Furthermore, in Figure 7, we enhance the input images of compared methods through MWM to demonstrate the effectiveness of other components in USNet. As can be seen, our method still outperforms other algorithms on the SUIM dataset. Taking 300 superpixels in SUIM dataset for example, our method's minimum percentage gain (computed with the highest score of the compared methods) of UE is 7.7%, while that of BR is 4.8%.

Concerning the other state-of-the-art methods, due to the low quality of the images in the underwater image dataset, it presents a challenge to achieve satisfactory performance when trained with the SUIM dataset. Therefore, for comparative experiments, we use SUIM datasets as training datasets for other state-of-the-art methods. Similarly, Figures 8, 9 show the quantitative comparison results, but other state-of-the-art methods were trained using the underwater dataset (SUIM). We can observe that the segmentation performance of other state-of-the-art methods using SUIM as the training set suffers from varying degrees of negative impact, due to quality degradation. In contrast, our methods still continued to achieve superior segmentation performance. Taking 500

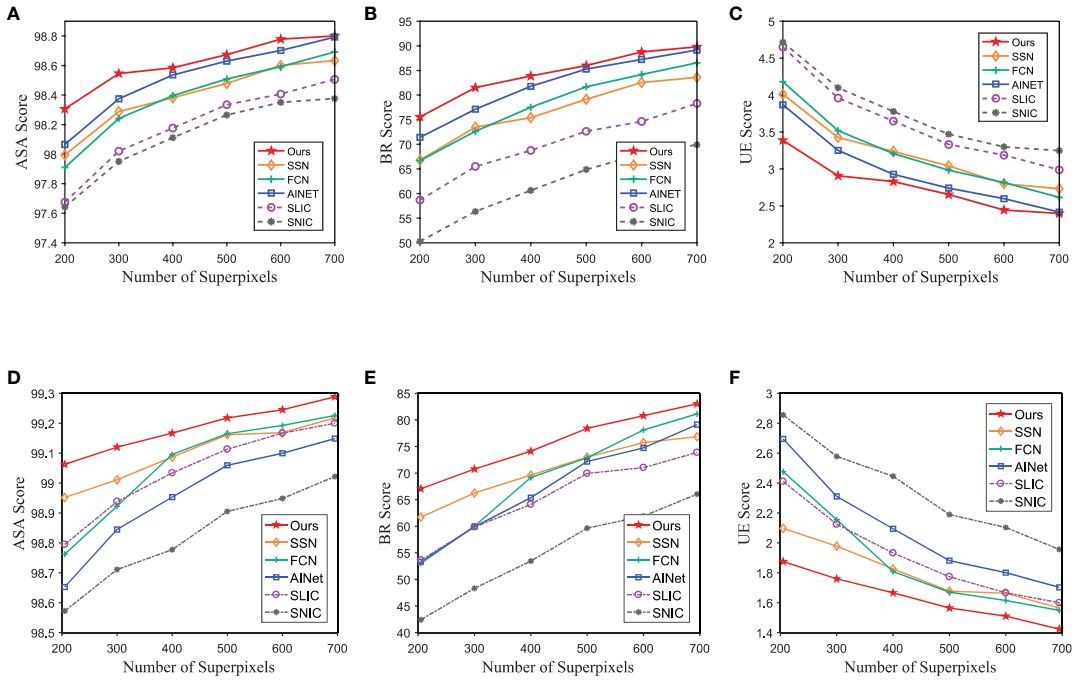

FIGURE 6
Quantitative comparison of the proposed method and other state-of-the-art methods using BSDS500 as the training set. Panels **(A–C)** are the performance on the test set of SUIM dataset. Panels **(D–F)** are the performance on the caustics scene in underwater change detection dataset.

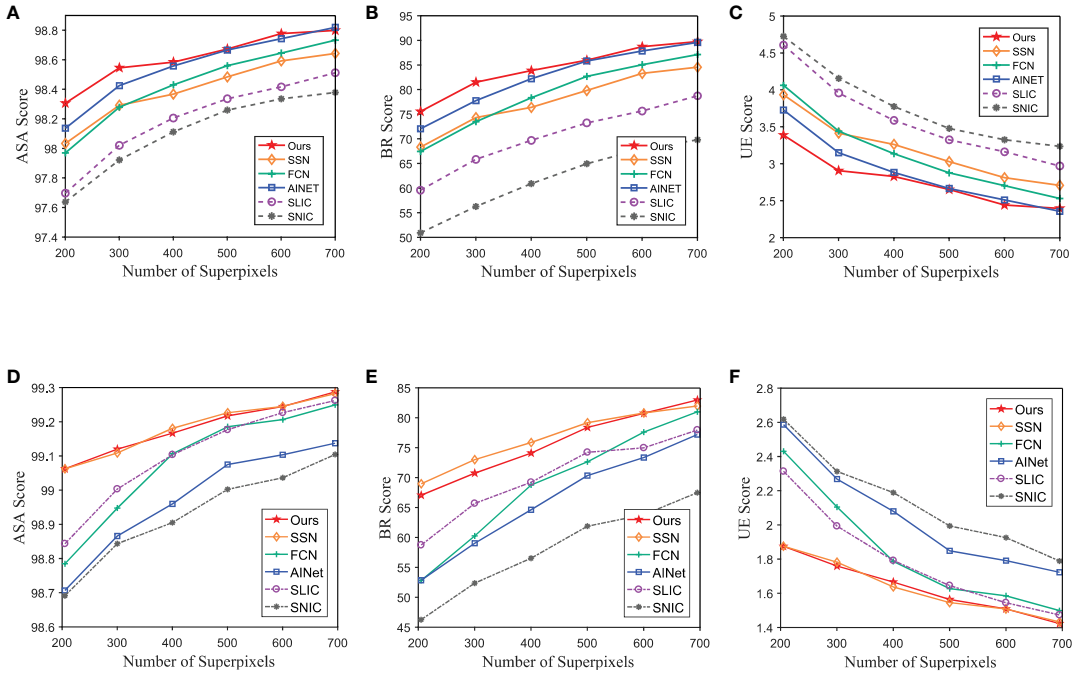

FIGURE 7
Quantitative comparison of the proposed method and other state-of-the-art methods using BSDS500 as the training set, the input images of compared methods are enhanced by MWM. Panels **(A–C)** are the performance on the test set of SUIM dataset. Panels **(D–F)** are the performance on the caustics scene in underwater change detection dataset.

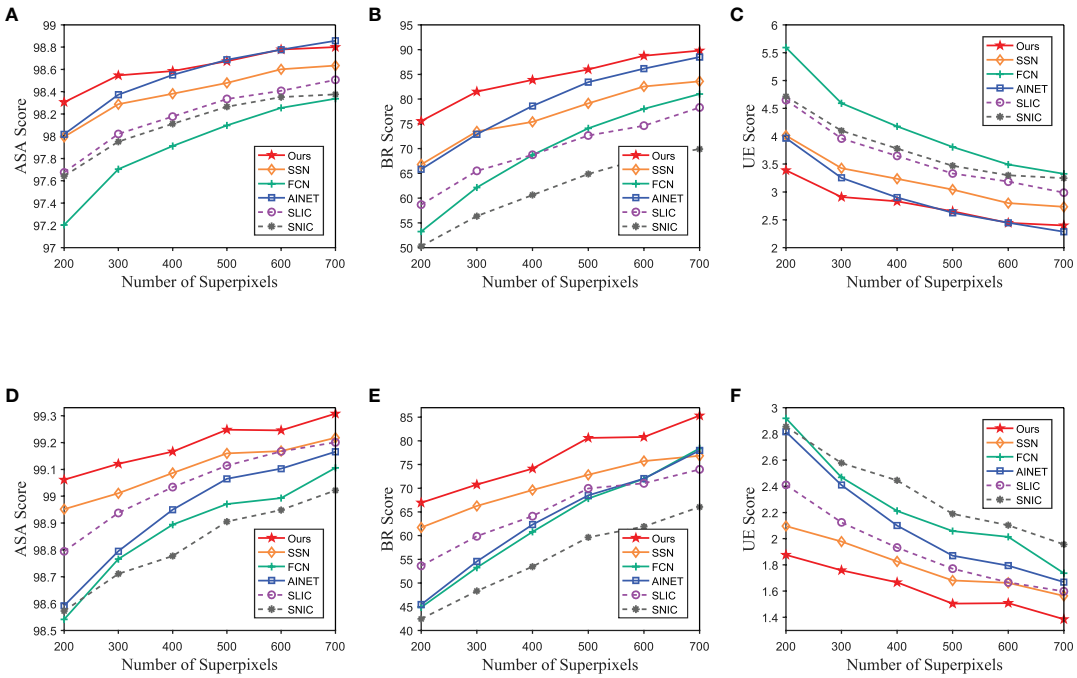

FIGURE 8
Quantitative comparison of the proposed method and other state-of-the-art methods using SUIM as the training set. Panels **(A–C)** are the performance on the test set of SUIM dataset. Panels **(D–F)** are the performance on the caustics scene in underwater change detection dataset.

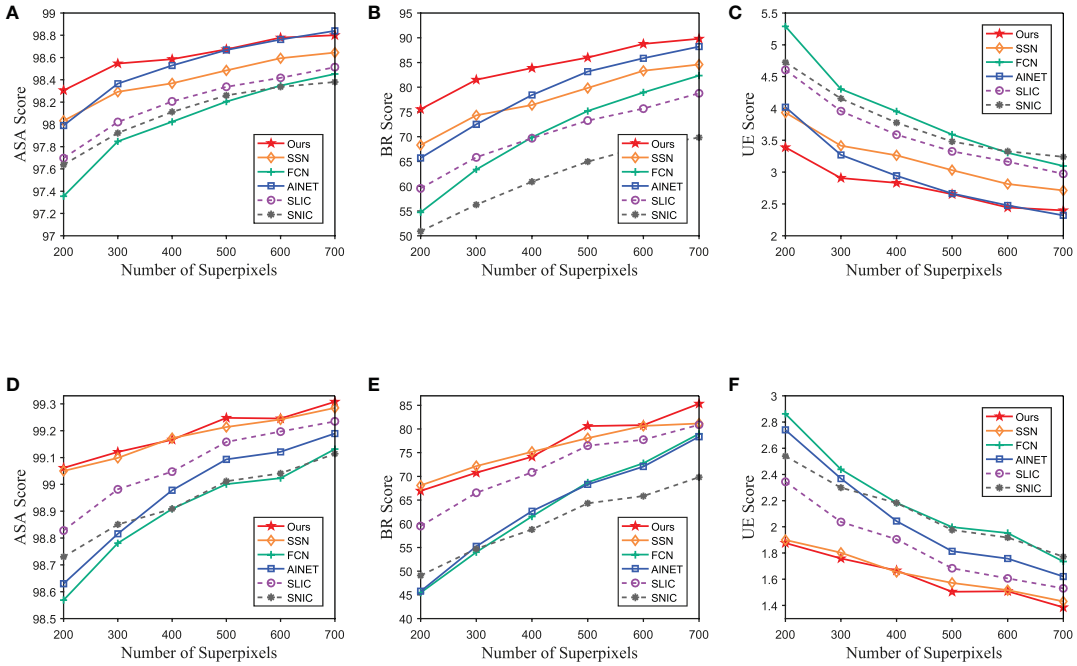

FIGURE 9
Quantitative comparison of the proposed method and other state-of-the-art methods using SUIM as the training set; the input images of compared methods are enhanced by MWM. Panels **(A–C)** are the performance on the test set of SUIM dataset. Panels **(D–F)** are the performance on the caustics scene in underwater change detection dataset.

superpixels in the caustic scene for example, our method's minimum percentage gain (computed with the highest score of the compared methods) of UE is 10.5%, while that of BR is 10.6%. The results prove that our method can handle complicated underwater scenes and outperform existing state-of-the-art segmentation methods. In addition, this observation highlights the convenience and efficiency of the MWM approach. As a result, MWM approach has the potential to be broadly applicable to other fields in the future, such as image segmentation and object detection.

4.2.2 Qualitative comparison

As the qualitative comparison results shown in Figures 10, 11, we present in detail the visual effects obtained through our

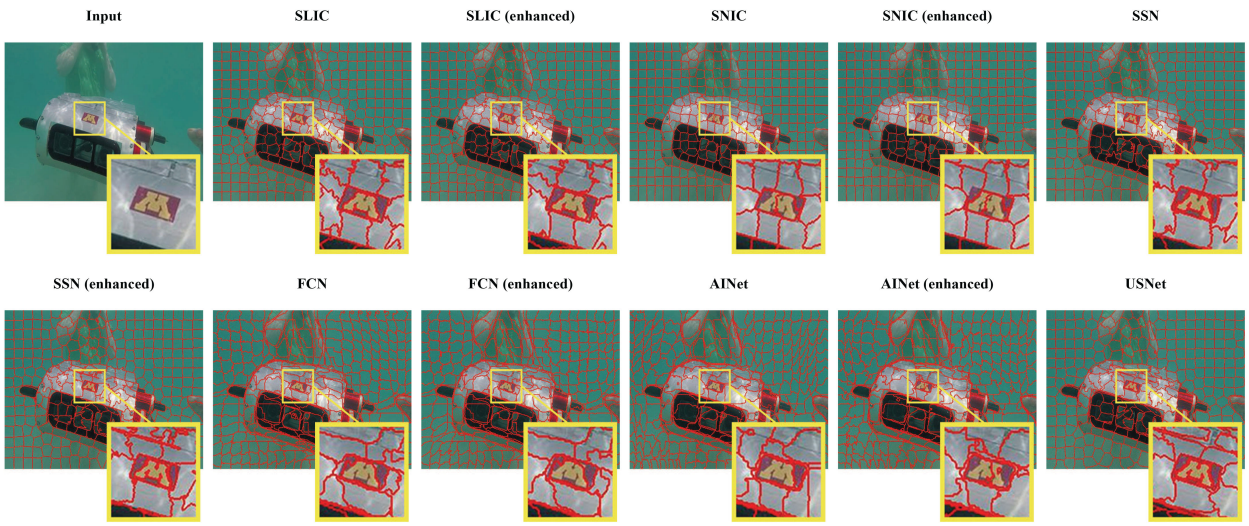

FIGURE 10
Qualitative comparison of the proposed method and other state-of-the-art methods on SUIM.

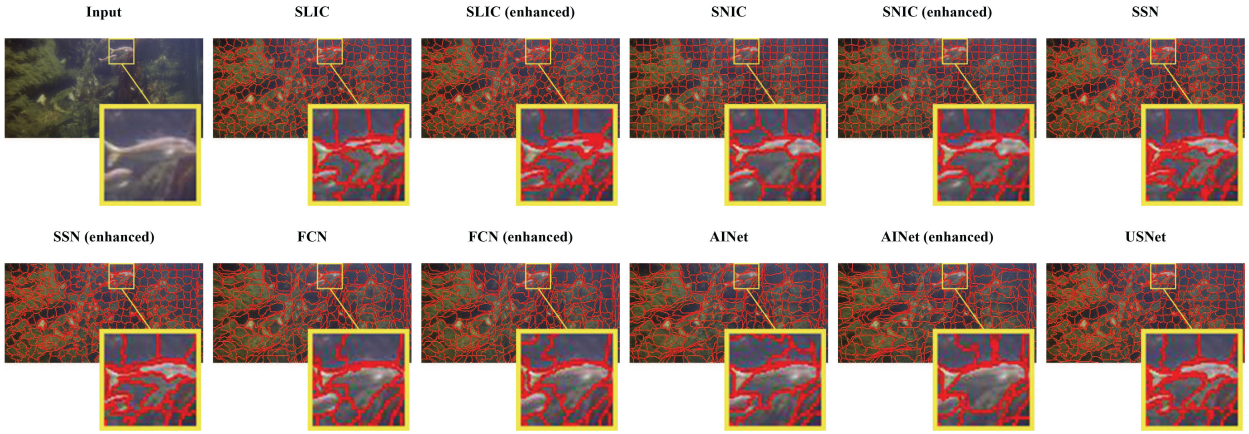

FIGURE 11
Qualitative comparison of the proposed method and other state-of-the-art methods on caustics.

approach and other state-of-the-art methods on the SUIM and caustics datasets, with and without the MWM augmentation for image segmentation. It is apparent that our method adheres more accurately and comprehensively to the boundaries when compared to other methods, which often fail to capture such intricate details. Regarding the SUIM dataset specifically, we can find that SNIC, FCN, and AINET are unable to fully segment the letter M, while SLIC and SSN can segment the important edges of the M but are unable to fit them together seamlessly. Only our method is capable of both fully segmenting and fitting the edges of the M clearly and completely. In addition, concerning the caustics dataset, where part of the segmentation focus appears blurred and colors are weak, our algorithm excels at distinguishing the target object from the background. It can smoothly segment fish and can even accurately capture the complete boundaries of low-contrast tail and fins of fish, which other algorithms cannot achieve.

To summarize, as shown in the figures, while augmenting the input for other methods can enhance their visual segmentation performance, the proposed USNet still achieves superior visual performance on the SUIM and caustics datasets. Moreover, we can observe that utilizing MWM to improve other segmentation algorithms enhances the segmentation results, indicating the strong generalization performance of MWM and its potential for application in other domains in the future.

4.3 Ablation experiments

We can observe that the ablation model with MWM can largely improve the score of UE and BR, since MWM can enhance the quality of underwater images as shown in Figure 12. Note that Full Model means our methods USNet including MWM module, DSFE module with CA mechanism, DSEM module, and others. CA+DSEM refers to USNet without an MWM module, MWM+CA means USNet without a DSEM module, MWM refers to USNet without the DSFE module and DSEM module, the baseline is SSN.

Furthermore, the ablation model with CA and improved DSEM also improves the BR score, indicating that comprehensive spatial information can capture the boundaries of complicated underwater scenes. However, the improvement in the UE score of this model is only slight compared to the baseline, since there are various quality degradations in underwater scenes that introduce more superpixel segmentation errors concerning the ground truth.

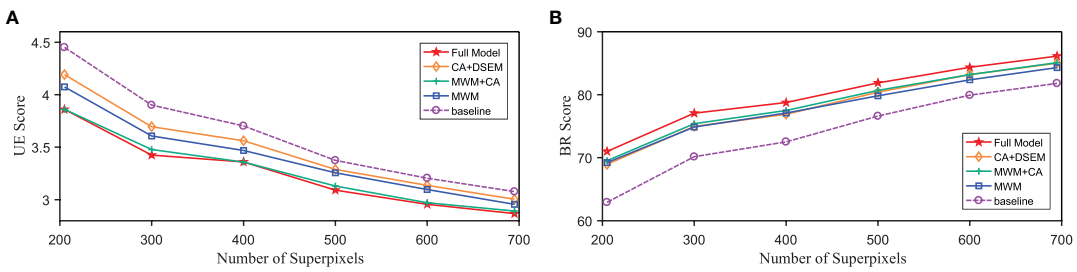

FIGURE 12
Results of ablation studies on SUIM. **(A)** UE metrics. **(B)** BR metrics.

TABLE 1 Results on SUIM.

Model	MAE ↓	S-Measure↑
SSN	0.1896	0.6223
FCN	0.1910	0.6189
SLIC	0.1949	0.6308
SNIC	0.1966	0.6243
AINet	0.1966	0.6243
USNet	**0.1859**	**0.6311**

↑ denotes that higher is better, and ↓ indicates the opposite.
The bold values represent the best results.

4.4 Application on salient object detection

Salient object detection (SOD) has a wide range of applications in fields such as object segmentation (Wang et al., 2015), object detection (Zhang et al., 2019a; Jiao et al., 2021), visual tracking (Li et al., 2015; Wang et al., 2017), and image compression (Guo and Zhang, 2009; Fang et al., 2013). Due to the absence of high-level knowledge, several existing methods still focus on exploiting low-level cues, such as contrast (Perazzi et al., 2012; Cheng et al., 2013; Yang et al., 2013; Cheng et al., 2014) and boundary prior (Wei et al., 2012) to improve the accuracy of s object detection algorithms. However, these methods suffer from fragility and lack a principled optimization framework. To address these issues, Zhu et al. (2014) proposed a novel method for salient object detection by utilizing superpixels instead of hard segmentation. They observed that background regions are more connected to image boundaries than salient object regions and treated the problem as a saliency value optimization problem for all superpixels in an image. The method used superpixels to construct an undirected weighted graph, which represented the relationships between different regions in the image and allowed for the calculation of the saliency of each region graph. The method precisely captured the spatial layout of objects and background regions in natural images while circumventing the challenging issue of algorithm and parameter selection associated with hard segmentation, resulting in improved performance.

To demonstrate the superior performance of our method in downstream tasks, we evaluated its performance against six state-of-the-art methods, including our proposed method, SSN (Jampani et al., 2018), FCN (Yang et al., 2020), AINET (Wang et al., 2021), SNIC (Achanta and Süsstrunk, 2017), and the default SLIC (Achanta et al., 2012) used as the superpixel segmentation method (Zhu et al., 2014). In our experiments, we used the SUIM dataset for evaluation purposes and resized all images to 400 × 400 for the sake of convenience in experimentation. We evaluate our model's performance using two metrics: Mean Absolute Error (MAE) (Perazzi et al., 2012) and Enhanced Alignment Measure (E-measure) (Fan et al., 2018a). MAE calculates the average difference between the binary ground truth and the predicted saliency map, but it only considers pixel-wise errors. On the other hand, E-measure incorporates structural cues to evaluate the model's performance.

Table 1 presents the results of the quantitative evaluation, which demonstrate that our method outperforms other state-of-the-art methods in terms of both MAE and E-measure. Furthermore, Figure 13 provides visual evidence that our saliency map can

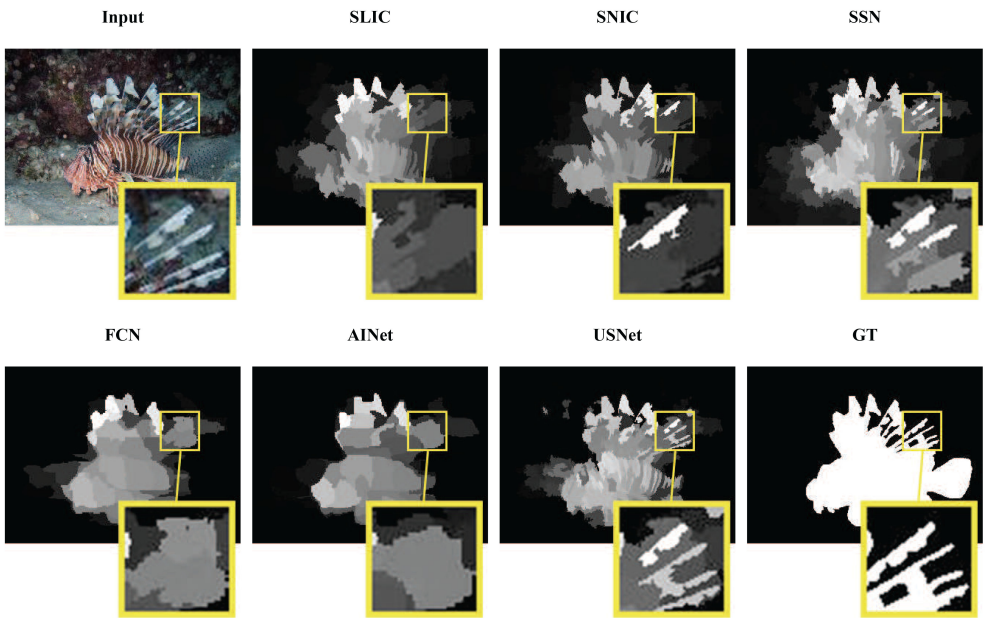

FIGURE 13
Visual comparison of SOD results obtained using various superpixel segmentation methods reveals that our method is capable of capturing more features compared to others.

capture more details compared to other methods. This validation confirms that our method performs well in downstream tasks, both quantitatively and qualitatively.

5 Conclusion

In this paper, we have proposed an end-to-end superpixel segmentation network for underwater images (USNet). Considering a variety of quality degradation appears in underwater scenes, we design a multi-scale water-net module (MWM) to enhance the quality of underwater images before superpixel segmentation to alleviate such issues. Since several regions with severe quality degradation reduce the overall performance of subsequent tasks, we also design a degradation-aware attention (DA) to enforce the network to pay more attention to high-degradation regions. Moreover, we utilize the coordinate attention mechanism to extract the deep spatial features, which are fused with the shallow spatial features to embed comprehensive spatial features through the dynamic spatial embedding module.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

CW: Conceptualization, Writing – original draft, Writing – review & editing, Methodology, Project administration. WD: Conceptualization, Methodology, Writing – review & editing. CL: Conceptualization, Writing – review & editing. JL: Formal Analysis, Writing – review & editing. LS: Writing – review & editing. HL: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grant 62201179; in part by Hainan Provincial Natural Science Foundation of China under Grant 622RC623; in part by the Innovation Platform for "New Star of South China Sea" of Hainan Province under Grant No. NHXXRCXM202306; in part by the specific research fund of The Innovation Platform for Academicians of Hainan Province under Grant No.YSPTZX202410; and in part by the Research Start-up Fund of Hainan University [No. KYQD(ZR)-22015].

Acknowledgments

We thank the editors and reviewers for valuable and thoughtful comments to help improve this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* (Piscataway, NJ: IEEE) 34, 2274–2282. doi: 10.1109/TPAMI.2012.120

Achanta, R., and Süsstrunk, S. (2017). "Superpixels and polygons using simple non-iterative clustering," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4895–4904. doi: 10.1109/CVPR.2017.520

Arnaubec, A., Ferrera, M., Escartín, J., Matabos, M., Gracias, N., and Opderbecke, J. (2023). Underwater 3d reconstruction from video or still imagery: Matisse and 3dmetrics processing and exploitation software. *J. Mar. Sci. Eng.* 11, 985. doi: 10.3390/jmse11050985

Barcelos, I. B., Belém, F. D. C., João, L. D. M., Patrocínio, Z. K. D. Jr., Falcão, A. X., and Guimarães, S. J. F. (2024). A comprehensive review and new taxonomy on superpixel segmentation. *ACM Comput. Surv.* 56, 1–39. doi: 10.1145/3652509

Calantropio, A., and Chiabrando, F. (2024). Underwater cultural heritage documentation using photogrammetry. *J. Mar. Sci. Eng.* 12, 413. doi: 10.3390/jmse12030413

Catalan, I. A., Álvarez-Ellacuría, A., Lisani, J.-L., Sanchez, J., Vizoso, G., Heinrichs-Maquilon, A. E., et al. (2023). Automatic detection and classification of coastal mediterranean fish from underwater images: Good practices for robust training. *Front. Mar. Sci.* 10, 1151758. doi: 10.3389/fmars.2023.1151758

Cheng, M.-M., Mitra, N. J., Huang, X., Torr, P. H., and Hu, S.-M. (2014). Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 569–582. doi: 10.1109/TPAMI.2014.2345401

Cheng, M.-M., Warrell, J., Lin, W.-Y., Zheng, S., Vineet, V., and Crook, N. (2013). "Efficient salient region detection with soft image abstraction," in *Proceedings of the IEEE International Conference on Computer vision*. (Piscataway, NJ: IEEE), 1529–1536.

Cong, R., Lei, J., Fu, H., Huang, Q., Cao, X., and Hou, C. (2017a). Co-saliency detection for rgbd images based on multi-constraint feature matching and cross label propagation. *IEEE Trans. Image Process.* 27, 568–579. doi: 10.1109/TIP.2017.2763819

Cong, R., Lei, J., Fu, H., Lin, W., Huang, Q., Cao, X., et al. (2017b). An iterative co-saliency framework for rgbd images. *IEEE Trans. Cybern.* 49, 233–246. doi: 10.1109/TCYB.2017.2771488

Cong, R., Lei, J., Fu, H., Porikli, F., Huang, Q., and Hou, C. (2019). Video saliency detection via sparsity-based reconstruction and propagation. *IEEE Trans. Image Process.* 28, 4819–4831. doi: 10.1109/TIP.83

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision*

and Pattern Recognition. (Piscataway, NJ: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Fan, D.-P., Gong, C., Cao, Y., Ren, B., Cheng, M.-M., and Borji, A. (2018a). Enhanced-alignment measure for binary foreground map evaluation. *arXiv preprint arXiv:1805.10421*. doi: 10.24963/ijcai.2018

Fan, F., Ma, Y., Li, C., Mei, X., Huang, J., and Ma, J. (2017). Hyperspectral image denoising with superpixel segmentation and low-rank representation. *Inf. Sci.* 397, 48–68. doi: 10.1016/j.ins.2017.02.044

Fan, L., Chen, L., Zhang, C., Tian, W., and Cao, D. (2018b). Collaborative three-dimensional completion of color and depth in a specified area with superpixels. *IEEE Trans. Ind. Electron.* 66, 6260–6269. doi: 10.1109/TIE.2018.2873474

Fang, Y., Lin, W., Chen, Z., Tsai, C.-M., and Lin, C.-W. (2013). A video saliency detection model in compressed domain. *IEEE Trans. circuits Syst. video Technol.* 24, 27–38. doi: 10.1109/TCSVT.2013.2273613

Guo, C., and Zhang, L. (2009). A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. image Process.* 19, 185–198. doi: 10.1109/TIP.2009.2030969

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ: IEEE), 770–778.

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Piscataway, NJ: IEEE), 13713–13722.

Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., et al. (2020). "Semantic segmentation of underwater imagery: Dataset and benchmark," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. (Piscataway, NJ: IEEE), 1769–1776. doi: 10.1109/IROS45743.2020.9340821

Jampani, V., Sun, D., Liu, M., Yang, M., and Kautz, J. (2018). Superpixel sampling networks. In *Proceedings of the European Conference on Computer Visio, ECCV 2018* (Berlin: Springer), 363–380.

Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L., et al. (2021). New generation deep learning for video object detection: A survey. *IEEE Trans. Neural Networks Learn. Syst.* 33, 3195–3215. doi: 10.1109/TNNLS.2021.3053249

Kim, C., Song, D., Kim, C.-S., and Park, S.-K. (2019). Object tracking under large motion: Combining coarse-to-fine search with superpixels. *Inf. Sci.* 480, 194–210. doi: 10.1016/j.ins.2018.12.042

Kingma, D. P., and Ba, J. (2015). "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. (Piscataway, NJ: IEEE).

Kumar, B. V. (2023). An extensive survey on superpixel segmentation: A research perspective. *Arch. Comput. Methods Eng.* 30, 3749–3767. doi: 10.1007/s11831-023-09919-8

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.83

Li, X., Han, Z., Wang, L., and Lu, H. (2015). Visual tracking via random walks on graph model. *IEEE Trans. Cybern.* 46, 2144–2155. doi: 10.1109/TCYB.2015.2466437

Li, H., Liang, J., Wu, R., Cong, R., Wu, W., and Kwong, S. T. W. (2023). Stereo superpixel segmentation via decoupled dynamic spatial-embedding fusion network. *IEEE Trans. Multimed.* 26, 367–378. doi: 10.1109/TMM.2023.3265843

Li, Y., Liu, Y., Zhu, J., Ma, S., Niu, Z., and Guo, R. (2020). Spatiotemporal road scene reconstruction using superpixel-based markov random field. *Inf. Sci.* 507, 124–142. doi: 10.1016/j.ins.2019.08.038

Ni, Z., Yang, W., Wang, S., Ma, L., and Kwong, S. (2020). Towards unsupervised deep image enhancement with generative adversarial network. *IEEE Trans. Image Process.* 29, 9140–9151. doi: 10.1109/TIP.83

Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process*. 32, 3066–3079. doi: 10.1109/TIP.2023.3276332

Perazzi, F., Krähenbühl, P., Pritch, Y., and Hornung, A. (2012). "Saliency filters: Contrast based filtering for salient region detection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2012.6247743

Qin, X., Wang, Z., Bai, Y., Xie, X., and Jia, H. (2020). "Ffa-net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, (Menlo Park: AAAI) Vol. 34. 11908–11915.

Qiu, T., Zhao, Z., Zhang, T., Chen, C., and Chen, C. P. (2019). Underwater internet of things in smart ocean: System architecture and open issues. *IEEE Trans. Ind. Inf.* 16, 4297–4307. doi: 10.1109/TII.9424

Radolko, M., Farhadifard, F., and von Lukas, U. F. (2016). "Dataset on underwater change detection," in *OCEANS 2016 MTS/IEEE Monterey*. 1–8. doi: 10.1109/OCEANS.2016.7761129

Shi, J., and Malik, J. M. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 888–905. doi: 10.1109/34.868688

Song, R., Zhang, W., Zhao, Y., Liu, Y., and Rosin, P. L. (2023). 3d visual saliency: an independent perceptual measure or a derivative of 2d image saliency? *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 13083–13099. doi: 10.1109/TPAMI.2023.3287356

Soomro, N. Q., and Wang, M. (2017). Superpixel segmentation: A benchmark. *Signal Process. Image Commun.* 56, 28–39. doi: 10.1016/j.image.2017.04.007

Strachan, N. (1993). Recognition of fish species by colour and shape. *Image Vision comput.* 11, 2–10. doi: 10.1016/0262-8856(93)90027-E

Stutz, D., Hermans, A., and Leibe, B. (2018). Superpixels: An evaluation of the state-of-the-art. *Comput. Vision Image Understand.* 166, 1–27. doi: 10.1016/j.cviu.2017.03.007

Subudhi, S., Patro, R. N., Biswal, P. K., and Dell'Acqua, F. (2021). A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 14, 5015–5035. doi: 10.1109/JSTARS.2021.3076005

Sultana, N., Mridula, D. T., Sheikh, Z., Iffath, F., and Shopon, M. (2022). "Dense optical flow and residual network-based human activity recognition," in *New Approaches for Multidimensional Signal Processing: Proceedings of International Workshop, NAMSP 2021* (Singapore: Springer Singapore), 163–173.

Uziel, R., Ronen, M., and Freifeld, O. (2019). "Bayesian adaptive superpixel segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Piscataway, NJ: IEEE). doi: 10.1109/ICCV43118.2019

Wang, L., Lu, H., and Yang, M.-H. (2017). Constrained superpixel tracking. *IEEE Trans. cybern.* 48, 1030–1041. doi: 10.1109/TCYB.2017.2675910

Wang, W., Shen, J., and Porikli, F. (2015). "Saliency-aware geodesic video object segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3395–3402.

Wang, Y., Wei, Y., Qian, X., Zhu, L., and Yang, Y. (2021). "Ainet: Association implantation for superpixel segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7078–7087.

Wei, Y., Wen, F., Zhu, W., and Sun, J. (2012). "Geodesic saliency using background priors," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7–13, 2012. 29–42 (Berlin, Heidelberg: Springer), Proceedings, Part III 12.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. (Berlin: Springer), 3–19.

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimed. Tools Appl.* 83, 31957–31973. doi: 10.1007/s11042-023-16757-0

Yang, F., Sun, Q., Jin, H., and Zhou, Z. (2020). "Superpixel segmentation with fully convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Piscataway, NJ: IEEE). doi: 10.1109/CVPR42600.2020

Yang, C., Zhang, L., Lu, H., Ruan, X., and Yang, M.-H. (2013). "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3166–3173.

Zhang, M., Li, J., Wei, J., Piao, Y., and Lu, H. (2019a). Memory-oriented decoder for light field salient object detection. *Adv. Neural Inf. Process. Syst.* 32, 896–906.

Zhang, W., Wang, B., Ma, L., and Liu, W. (2019b). Reconstruct and represent video contents for captioning via reinforcement learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 3088–3101. doi: 10.1109/TPAMI.34

Zhang, X., Yang, P., Wang, Y., Shen, W., Yang, J., Ye, K., et al. (2024). Lbf-based cs algorithm for multireceiver sas. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. doi: 10.1109/LGRS.2024.3379423

Zhou, J., Pang, L., Zhang, D., and Zhang, W. (2023). Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE J. Ocean. Eng.* 48, 474–488. doi: 10.1109/JOE.2022.3223733

Zhu, W., Liang, S., Wei, Y., and Sun, J. (2014). "Saliency optimization from robust background detection," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. (Piscataway, NJ: IEEE), 2814–2821. doi: 10.1109/CVPR.2014.360