QTU-Net: Quaternion Transformer-Based U-Net for Water Body Extraction of RGB Satellite Image

Mingzhi Wang, Chunshan Li, *Member, IEEE*, Xiaofei Yang, *Member, IEEE*, Dianhui Chu, Zhiquan Zhou, and Raymond Y. K. Lau, *Senior Member, IEEE*

***Abstract*— Deep learning models have achieved great success in water body extraction (WBE) from remote sensing images. However, the existing deep learning-based extraction methods exhibit limitations in their ability to fully explore the intricate interconnections inherent in RGB color satellite imagery and to enhance semantic representation across diverse regions. Furthermore, these methods often struggle with challenges posed by the uneven distribution of water bodies at different scales within the image, as well as substantial color disparities between water and land areas. In this article, we tackle WBE task from quaternion domain and introduce a novel approach called quaternion transformer-based U-Net (QTU-Net) to address these challenges. Our method specifically leverages quaternion convolution operations to capture the holistic relationships among RGB channels, thereby enhancing the semantic representation of WBE. Additionally, we propose a quaternion initialization module (QIM) to determine optimal RGB weights and facilitate the generation of quaternion data. To further improve the accuracy of water body delineation, we incorporate an innovative multiscale similarity aggregation attention (MSAA) component that enhances local similarity capture across various scales. Finally, we evaluate the proposed QTU-Net based on three publicly available benchmark datasets. The experimental results demonstrate that the proposed QTU-Net outperforms state-of-the-art baseline methods.**

***Index Terms*— Convolutional neural network (CNN), quaternion convolution, transformer network, U-Net, water body extraction (WBE).**

I. Introduction

WATER is the foundation that supports the survival of various biological activities and is the basis for the development of social civilization [1]. Water bodies, such as oceans, rivers, lakes, reservoirs, and ponds, play a critical role in the Earth's ecosystem and are essential for a wide range of human activities, including agriculture, industry, transportation, and recreation [2], [3]. Moreover, water bodies significantly impact on land use and urban planning [4], and the effective management of water resources is crucial for flood control, biodiversity conservation, and disaster mitigation. Therefore, the accurate identification and delineation of the locations and shapes of water bodies represent a profoundly significant area of research. With the rapid advancement of remote sensing technology, the interpretation of remote sensing imagery has emerged as an effective method for acquiring morphological maps of water bodies.

Water body extraction (WBE) from remote sensing images mainly involves separating the water body pixels from the images. Over the course of several decades, numerous methods for WBE have been devised employing various satellite platforms [5], [6], [7]. These existing methods for WBE are designed using synthetic aperture radar (SAR), multispectral, and optical remote sensing images. The RGB remote sensing images possess abundant textural and chromatic characteristics that closely align with the human visual perception, thereby offering an intuitive representation of the morphology of aquatic environments. This facilitates both manual and automated interpretation processes. Consequently, such imagery has been extensively adopted as the primary data type for WBE endeavors. With the rapid advancement of artificial intelligence technology, there has been a significant shift in the primary techniques used for WBE, moving away from conventional manual interpretation toward the utilization of deep learning-based approaches. Despite some deep learning approaches showing notable efficacy in WBE endeavors, the existing methods fall short in adequately accounting for the distinctive traits of water bodies in RGB satellite images, including their unique morphology, spatial distribution, and color attributes. As evident in Fig. 1, water bodies in satellite imagery typically demonstrate a concentrated spatial distribution and consistent color characteristics, facilitating clear differentiation from other land features. Moreover, a pronounced contrast at the pixel level exists between water bodies and their surroundings, enabling their distinct identification. Furthermore, in urban areas and other complex environments, satellite images

Manuscript received 20 March 2024; revised 10 June 2024; accepted 4 July 2024. Date of publication 11 July 2024; date of current version 12 August 2024. This work was supported in part by the Major Scientific and Technological Innovation Project of Shandong Province of China under Grant 2021ZLGX05 and Grant 2020CXGC010705, in part by the National Natural Science Foundation of China under Grant 62301174, and in part by Guangzhou Basic and Applied Basic Research Topic (Young Doctor "Sailing" Project) under Grant 2024A04J2081. *(Corresponding authors: Chunshan Li; Xiaofei Yang.)*

Mingzhi Wang, Chunshan Li, Dianhui Chu, and Zhiquan Zhou are with Harbin Institute of Technology, Weihai 264209, China (e-mail: wangmingzhi618@gmail.com; lics@hit.edu.cn; chudh@hit.edu.cn; zzq@hitwh.edu.cn).

Xiaofei Yang is with the School of Electronic and Communication Engineering, Guangzhou University, Guangzhou 510182, China (e-mail: xiaofeiyang@gzhu.edu.cn).

Raymond Y. K. Lau is with the Department of Information Systems, City University of Hong Kong, Hong Kong, China (e-mail: raylau@cityu.edu.hk).

Data is available on-line at: https://github.com/HitMingzhiWang/QTU-Net.

Digital Object Identifier 10.1109/TGRS.2024.3426475

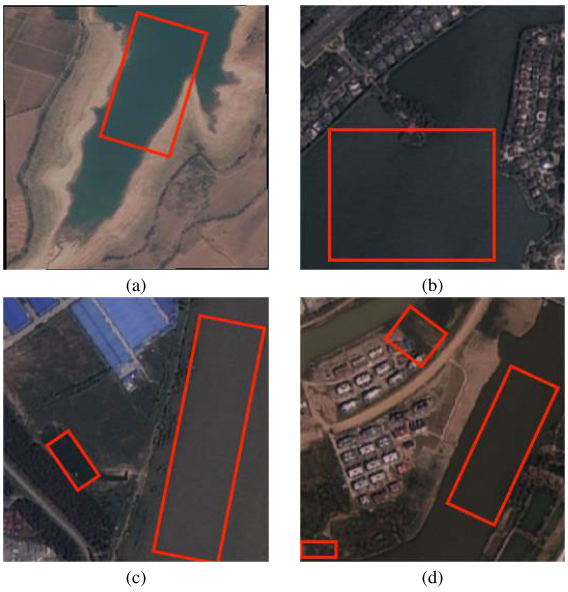

Fig. 1. RGB satellite images of water bodies. The red box represents a large area of water that is clustered together and has similar colors. (a) and (b) Exhibit expansive single water body aggregations, whereas (c) and (d) present a diverse array of water bodies at different scales, highlighting a substantial variation in their sizes.

often capture water bodies of varying sizes, posing additional challenges for accurate segmentation. The prevailing deep learning-based methods for water bodies extraction primarily rely on discerning color disparities between water areas and adjacent surfaces. These approaches do not optimally capture the color characteristics inherent in satellite imagery and maintain a high degree of independence among high-dimensional channels. Consequently, it hinders effective interaction among the RGB components, impeding the revelation of complex relationships crucial for precise segmentation.

To address these challenges, in this article, we rethink the WBE task from quaternion domain and introduce a novel segmentation model called quaternion transformer-based U-Net (QTU-Net). The QTU-Net could take advantage of U-Net architecture and quaternion to explore the intricate interconnections within RGB color satellite images, the uneven distribution of water bodies, and the substantial color disparities to enhance semantic representation and improve the performance. Specifically, we propose a quaternion initialization module (QIM) to differentiate the weights of RGB components in different object semantic features and map the three RGB channel components into the quaternion domain. Furthermore, we propose a multiscale similarity aggregation attention (MSAA) module to capture spatial contextual information. Unlike the previous attention mechanisms, the MSAA is an innovative approach that models pixel disparities between water body feature and other feature, effectively extracting water bodies of diverse scales from the image. The QTU-Net is constructed with a symmetrical U-shaped architecture, incorporating the QIM and a sequence of MSAA modules. It is noting that the QTU-Net is a proficient method operating within the quaternion domain, which is different from other WBE methods.

In essence, the primary contributions of this article can be summarized as follows.

1) We revisit WBE in the quaternion domain and propose a novel method called QTU-Net specifically tailored for remote sensing images. QTU-Net adeptly captures complex interconnections, uneven distribution, and color variations associated with water bodies.
2) An innovative QIM, utilizing efficient channel attention (ECA), is devised to extract quaternion data and discern unique water body attributes from RGB satellite images.
3) We introduce an MSAA module, leveraging convolution operations to enhance contextual understanding and accurately delineate water body boundaries.
4) Through extensive experiments on three real datasets, we validate the effectiveness of QTU-Net, demonstrating its superior performance in WBE.

The rest of this article is organized as follows. In Section II, we review the previous work on WBE. In Section III, we elaborate on the structure of QTU-Net and detail the components of QIM and MSAA. In Section IV, we describe the experimental steps and conduct a detailed analysis of the experimental results. Finally, we present our conclusions in Section V.

II. Related Work

A. Conventional Methods for WBE

The existing methods for WBE can broadly be classified into three main categories: threshold-based approaches, spectral index techniques, and machine learning methods. In the case of the threshold methods, Cao et al. [8] employed the bimodal distribution of SAR image intensity maps to establish the threshold for water bodies segmentation. Klemenjak et al. [9] employ mathematical morphology to autonomously extract river structures from SAR data. In the research of water bodies extraction from multispectral remote sensing images, the utilization of the Normalized Difference Water Index (NDWI) model [10] is a widely adopted method. The NDWI utilizes the contrast between the green light band and the near-infrared band to accentuate water-related information while suppressing details related to soil, vegetation, and other surface features. The NDWI can be calculated as follows:

$$\text{NDWI} = \frac{(\text{GREEN} - \text{NIR})}{(\text{GREEN} + \text{NIR})} \tag{1}$$

where GREEN is a band that encompasses reflected green light and NIR represents reflected near-infrared radiation. Inspired by NDWI, several novel index models with enhanced extraction accuracy have emerged, such as WNDWI [11], MST-NDWI [12], and MNDWI [13]. Nonetheless, spectral index techniques relying on NDWI frequently necessitate manual refinements to water body maps. Furthermore, NDWI's susceptibility to atmospheric conditions and its diminished performance in intricate settings, such as densely populated urban areas, render it a comparatively less reliable method for WBE.

In the realm of machine learning methods, both clustering algorithms and support vector machines (SVMs) [14] have

emerged as powerful techniques for WBE. Cordeiro et al. [15] introduced a nonparametric unsupervised automatic algorithm that leverages multidimensional clustering and a high-performance subsampling technique to identify inland water pixels from satellite imagery, particularly useful for large-scale scenes. Yousefi et al. [16] proposed an approach that integrates principal component analysis (PCA) with the K-means clustering algorithm to refine clustering outcomes, enhancing the accuracy of water body delineation. Furthermore, Sarp and Ozcelik [17] underscored the efficacy of SVMs in WBE tasks. It is worth noting, however, that machine learning techniques often require meticulous data preprocessing and manual feature selection, posing challenges to achieving fully automated water region recognition. Additionally, there is a pressing need to further improve segmentation accuracy in these methods to ensure reliable and precise WBE.

B. Convolutional Neural Networks for WBE

In recent years, the remarkable effectiveness of deep learning in the field of remote sensing has increasingly positioned it as a valuable tool for analyzing WBE. Particularly, RGB satellite images, with their high contrast and exceptional clarity, provide a rich source of texture and structural information. As a result, the combination of this imagery and deep learning has become widely utilized and easily accessible for extracting water bodies. Convolutional neural networks (CNNs) have proven to be exceptionally adept at extracting local features, including geometric information crucial for object recognition in images. This adaptability has led to the widespread application of CNNs in various contexts within satellite remote sensing image interpretation, such as road extraction [18], oil spill detection [19], land classification [20], bridge detection [21], oriented object detection [22], and WBE [23]. The field of image segmentation methods employing CNNs is continuously evolving. One influential contribution in this field is the fully convolutional network (FCN) [24], which replaces traditional fully connected layers with convolutional layers, which effectively addresses semantic segmentation challenges. Ronneberger et al. [25] introduced the U-Net encoder–decoder model with a symmetrical structure, which has since become a fundamental paradigm for a wide range of visual tasks. Furthermore, Chen et al. [26] developed DeepLab, a framework that integrates deep CNNs (DCNNs) and probability graph models, incorporating dilated convolutions to capture larger receptive fields and enhance spatial information features without sacrificing image resolution. To address challenges in object localization, the framework utilizes the conditional random fields (CRFs) algorithm. In recent years, an increasing number of researchers have applied CNNs to WBE tasks, building upon the abovementioned models. Wang et al. [27] proposed HA-Net that uses the two-branch encoder and hybrid-scale channel attention block to enhance the robustness of the water body segmentation algorithm. Weng et al. [28] proposed SR-SegNet, incorporating residual blocks and depthwise separable residual convolutions into the SegNet [29] model. Wang et al. [30] introduced SADA-Net, a model designed to establish connections between shape and semantic information. This model utilizes gated convolutional layers to enhance the extraction of water bodies in complex scenes. In efforts to accelerate the inference speed of water body segmentation, Nie et al. [31] proposed SE-BiSeNet, which extracts spatial and contextual information from two branches and employs the atmospheric spatial pyramid pooling (ASPP) to expand the receptive field while minimizing the number of model parameters.

C. Transformer Network for WBE

Although CNN-based methods can learn local details for precise water body segmentation well, they often struggle to capture global contextual information necessary for inferring water body contours. To address this limitation, techniques such as dilated convolution have been employed to enhance the receptive field. However, inspired by the remarkable success of the transformer architecture [32] in natural language processing (NLP), as demonstrated by Dosovitskiy et al. [33], the vision transformer (ViT) has emerged as an alternative approach for image analysis. The ViT offers a unique capability to capture global context information in images, providing a promising avenue for improving water body segmentation accuracy. Zhong et al. [34] developed a multiscale transformer block and developed NT-Net based on it, which improves the consistency of lake water boundaries. Additionally, Zhang et al. [35] presented MF-SegFormer, a model that combines features produced by SegFormer to enhance the identification of small water bodies and improve the delineation of water body edges. Qi et al. [36] integrated the geometric active controller model with the ViT to address the challenges of blurred boundaries and complex segmentation of water bodies in low-contrast regions. Transformer models have not only found widespread application in the domain of water extraction but have also proven their effectiveness in numerous other remote sensing tasks. These models have generated numerous noteworthy works, demonstrating their versatility and potential in addressing diverse challenges within the field of remote sensing. Jiang et al. [37] proposed the graph generative structure-aware transformer, which generates absolute positional encoding serving the transformer through a graph neural network, achieving efficient performance in hyperspectral image classification. Duan et al. [38] leveraged the region homogeneity and spectral correlation inherent in hyperspectral imagery to construct a novel double-aware transformer for hyperspectral unmixing. Su et al. [39] proposed a novel multiscale mixed residual transformer (MMRT), successfully enabling dynamic retrieval of global ocean underground density. He et al. [40] proposed a highly robust U-Net model, which integrates a parallel Swin transformer with CNN for semantic segmentation of remote sensing images.

D. Background of Quaternion Algebra

In this section, we give a brief introduction to quaternion which is proposed by Hamilton [41]. Quaternion algebra is an extension of the complex field $\mathbb{C}$. The original intention of quaternions was to facilitate the representation of vector operations in 3-D space, encompassing any stretching or rotational

transformation imaginable within that realm. Its widespread application in computer graphics and physical simulations has revolutionized rotation interpolation operations, surpassing the limitations encountered with Euler angles. Teramae et al. [42] developed a novel framework for optimizing human pose trajectories, leveraging the powerful representation of quaternion. Fathian et al. [43] proposed a new formulation based on rotation quaternion representation, which solves the problem of recovering the rotation and translation changes of a moving camera from captured images. A quaternion $\hat{\mathbf{q}}$ consists of both a real part and three imaginary components, typically expressed in the following form:

$$\hat{\mathbf{q}} = r + q_1 i + q_2 j + q_3 k, \quad r, q_1, q_2, q_3 \in \mathbb{R} \tag{2}$$

where $\mathbb{R}$ denotes the real field, while i, j, and k represent orthogonal imaginary units. These imaginaries are subject to a set of specific mathematical rules, as outlined below

$$\begin{aligned} ijk &= i^2 = j^2 = k^2 = -1 \\ ij &= -ji = k \\ jk &= -kj = i \\ ki &= -ik = j. \end{aligned} \tag{3}$$

In the context of (2), the quaternion denoted as $\hat{\mathbf{q}}$ is categorized as a pure quaternion when its real scalar component, designated as r, assumes a value of 0. The operation rules of quaternion algebra encompass several key principles.

1) *Addition:* When adding two quaternions, the real and imaginary components are summed independently. Mathematically, if we have two quaternions, $\hat{\mathbf{q}} = q_r + q_1 i + q_2 j + q_3 k$ and $\hat{\mathbf{p}} = p_r + p_1 i + p_2 j + p_3 k$, their sum is given by

$$\begin{aligned} \hat{\mathbf{q}} + \hat{\mathbf{p}} = q_r + p_r + (q_1 + p_1)i \\ + (q_2 + p_2)j + (q_3 + p_3)k. \end{aligned} \tag{4}$$

2) *Scalar Multiplication:* Scalar α needs to be multiplied by all components

$$\alpha\hat{\mathbf{q}} = \alpha r + \alpha q_1 i + \alpha q_2 j + \alpha q_3 k. \tag{5}$$

3) *Conjugate:* The conjugate of $\hat{\mathbf{q}}$ is defined as follows:

$$\hat{\mathbf{q}}^* = r - (q_1 i + q_2 j + q_3 k). \tag{6}$$

4) *Modulus:* The modulus of $\hat{\mathbf{q}}$ is defined as follows:

$$|\hat{\mathbf{q}}| = \left(r^2 + q_1^2 + q_2^2 + q_3^2\right)^{1/2}. \tag{7}$$

5) *Inverse:* The inverse of $\hat{\mathbf{q}}$ is defined as follows:

$$\hat{\mathbf{q}}^{-1} = \hat{\mathbf{q}}^*/|\hat{\mathbf{q}}|^2. \tag{8}$$

6) *Normalization:* Normalizing a quaternion involves scaling it so that its magnitude (the square root of the sum of squares of all components) becomes 1. A normalized quaternion is often used to represent rotations. The normalization of unit quaternion $\hat{\mathbf{q}}^{\triangleleft}$ is defined as follows:

$$\hat{\mathbf{q}}^{\triangleleft} = \frac{\hat{\mathbf{q}}}{\sqrt{r^2 + q_1^2 + q_2^2 + q_3^2}}. \tag{9}$$

7) *Multiplication:* Quaternion multiplication is not commutative, and it follows the Hamilton product rule [44]. Given two quaternions $\hat{\mathbf{q}}$ and $\hat{\mathbf{p}}$, their product is calculated as follows:

$$\begin{aligned} \hat{\mathbf{q}} \bigotimes \hat{\mathbf{q}} = (q_r p_r - q_1 p_1 - q_2 p_2 - q_3 p_3) \\ + (q_r p_1 + q_1 p_r + q_2 p_3 - q_3 p_2)i \\ + (q_r p_2 - q_1 p_3 + q_2 p_r + q_3 p_1)j \\ + (q_r p_3 + q_1 p_2 - q_2 p_1 + q_3 p_r)k. \end{aligned} \tag{10}$$

In the realm of remote sensing, quaternion algebra has seen application in hyperspectral image classification [45], [46], whereas its utilization in RGB satellite remote sensing images remains relatively limited.

III. Proposed Approach

In this section, we introduce the QTU-Net, our proposed model. We commence by presenting the overall structure of the QTU-Net in Section III-A. Subsequently, in Section III-B, we delve into the specifics of the QIM. Then we provide an in-depth description of the proposed MSAA module in Section III-D. Finally, we introduce three different variants of QTU-Net in Section III-F.

A. Overview

The input to QTU-Net is an RGB satellite image, denoted as $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, where $H \times W$ represents the image dimensions, and C signifies the three RGB channels. Our objective in the WBE task is to perform semantic segmentation, determining whether each pixel in the image corresponds to a water body region. QTU-Net, illustrated in Fig. 2, features a U-shaped symmetric encoder–decoder architecture. Within the encoder, the input data first undergoes the QIM to generate quaternion data, which captures the weight of RGB components in semantic expression. Subsequently, the encoder encompasses five feature extraction layers. Each ith feature extraction layer integrates a quaternion convolution block and multiple layers (L_i) of quaternion attention blocks, enhancing local details and spatial context information extraction. The quaternion attention block is designed utilizing the MSAA module, which improves the identification of water-like regions in images by computing similarity information across various scale regions. Each layer performs downsampling to reduce feature map size. The decoder section mirrors the encoder's architecture, featuring five layers, each equipped with a quaternion convolutional block that combines shallow feature maps with deeper ones. The decoder also executes two upsampling operations per layer, progressively restoring image resolution and integrating shallow and deep semantic information. This structured architecture facilitates detailed image analysis and water body region identification.

B. Quaternion Initialization Module

The function of the QIM is to generate quaternion data from RGB three-channel data and identify the weights of the three RGB components in semantic expression. As shown in Fig. 3, QIM is designed based on ECA [47]. To initialize the potential

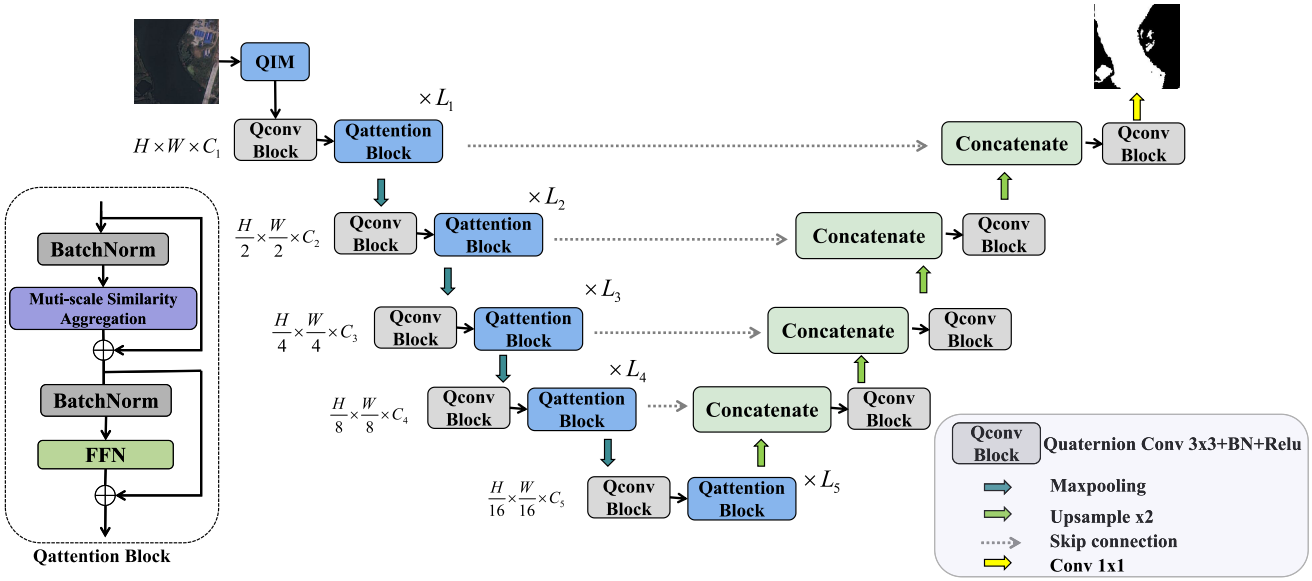

Fig. 2. Architecture of QTU-Net, which is based on U-Net and is composed of QIM, QConv block, and Qattention block. QIM is a quaternion data initialization module that generates quaternion data from RGB images. The QConv block sequentially performs 3 × 3 quaternion convolution operations, BatchNorm operations, and Relu() activation operations. The Qattention block, based on quaternion convolution and the MSAA components, captures spatial contextual features from images.

connection between the three channels, we first use a 1 × 1 2-D convolution operation to increase the dimensionality of the input data $\mathbf{X} \in \mathbb{R}^{3\times H\times W}$ to 64 ($\mathbf{X_h} \in \mathbb{R}^{64\times H\times W}$). Subsequently, ECA is applied to obtain the attention weights between channels, and the output is mapped to the three imaginary parts (i, j, and k) through another 1 × 1 2-D convolution operation. The process can be summarized as follows:

$$\begin{aligned}\mathbf{W} &= \sigma(\text{Conv1D}_3(\text{GAP}(\mathbf{X_h})))\\ \mathbf{q}_{ijk} &= \text{Conv2D}_1(\mathbf{X_h} \odot \mathbf{W})\end{aligned} \tag{11}$$

where $\text{GAP}(\mathbf{X_h}) = (1/HW)\sum_{i=1,j=1}^{H,W}(\mathbf{X_h})_{ij}$ is the global average pooling operation, $\mathbf{W} \in \mathbb{R}^{64\times 1\times 1}$, and $\mathbf{q}_{ijk} \in \mathbb{R}^{3\times H\times W}$. σ is the sigmoid function and $\odot$ represents the element-wise product. The real part of the quaternion $\mathbf{q}_r \in \mathbb{R}^{1\times H\times W}$ is the sum of the weights of all channels

$$\mathbf{q}_r = \text{Conv2D}_1(\mathbf{X_h}). \tag{12}$$

Then the final output quaternion data $\hat{\mathbf{q}}$ can be obtained by a concatenation operation of channel dimension

$$\hat{\mathbf{q}} = \text{concat}(\mathbf{q}_{ijk}, \mathbf{q}_r). \tag{13}$$

Subsequently, the quaternion data $\hat{\mathbf{q}} \in \mathbb{R}^{4\times H\times W}$ serves as the input data for the lower level feature extraction network.

C. Quaternion Convolution

The classic convolution operations are performed in the real-valued domain and have proven to play an important role in image analysis and feature extraction. Given an input feature map tensor $\mathbf{I}$ with dimension $H \times W$ and a convolution kernel tensor $\mathbf{K}$ with dimension $K_H \times K_W$, the mathematical representation of the classic convolution is as follows:

$$(\mathbf{I} * \mathbf{K})(x, y) = \sum_{i=0}^{K_H-1}\sum_{j=0}^{K_W-1} \mathbf{I}(x+i, y+j)\cdot\mathbf{K}(i, j). \tag{14}$$

In this equation, $\mathbf{I} * \mathbf{K}$ denotes the convolution operation, and it calculates the output value at position (x, y) in the output feature map. This is achieved by summing the products of input values from the image with the corresponding kernel weights. For quaternion convolution operation, it is a convolution operation using quaternion algebra. Given an input quaternion tensor $\mathbf{Q}$ with dimension $H \times W \times 4$, representing quaternion data, and a quaternion convolution kernel tensor $\mathbf{X_q}$ with dimension $X_H \times X_W \times 4$, the quaternion convolution operation can be defined as follows:

$$(\mathbf{Q} * \mathbf{X})(x, y) = \sum_{i=0}^{X_H-1}\sum_{j=0}^{X_W-1} \mathbf{Q}(x+i, y+j) \otimes \mathbf{X_q}(i, j). \tag{15}$$

In quaternion convolution, the value of each position in quaternion tensor $\mathbf{Q}$ is a quaternion data $\hat{\mathbf{q}} = r + q_1 i + q_2 j + q_3 k$. Similarly, the value for each position of the quaternion filtering matrix $\mathbf{X_q}$ is $\hat{\mathbf{x}} = r_x + x_1 i + x_2 j + x_3 k$. Then the numerical product of the corresponding positions can be calculated as follows:

$$\begin{aligned}\hat{\mathbf{q}} \bigotimes \hat{\mathbf{x}} &= (rr_x - q_1x_1 - q_2x_2 - q_3x_3)\\ &\quad + (rr_x + q_1x_1 + q_2x_3 - q_3x_2)i\\ &\quad + (rx_2 - q_1x_3 + q_2r_x + q_3x_1)j\\ &\quad + (rx_3 + q_1x_2 - q_2x_1 + q_3r_x)k\\ &= \begin{bmatrix} r & -q_1 & -q_2 & -q_3\\ q_1 & r & -q_3 & q_2\\ q_2 & q_3 & r & -q_1\\ q_3 & -q_2 & q_1 & r\end{bmatrix}_{4\times 4} * \begin{bmatrix} x_r\\ x_1\\ x_2\\ x_3\end{bmatrix}_{4\times 1}.\end{aligned} \tag{16}$$

Indeed, within quaternion convolution operations, the consecutive fusion of every four channels of a feature map can be regarded as a quaternion data representation. This approach contrasts with conventional convolutional operations, where the components of the three RGB channels are consistently constrained within a quaternion domain. Consequently,

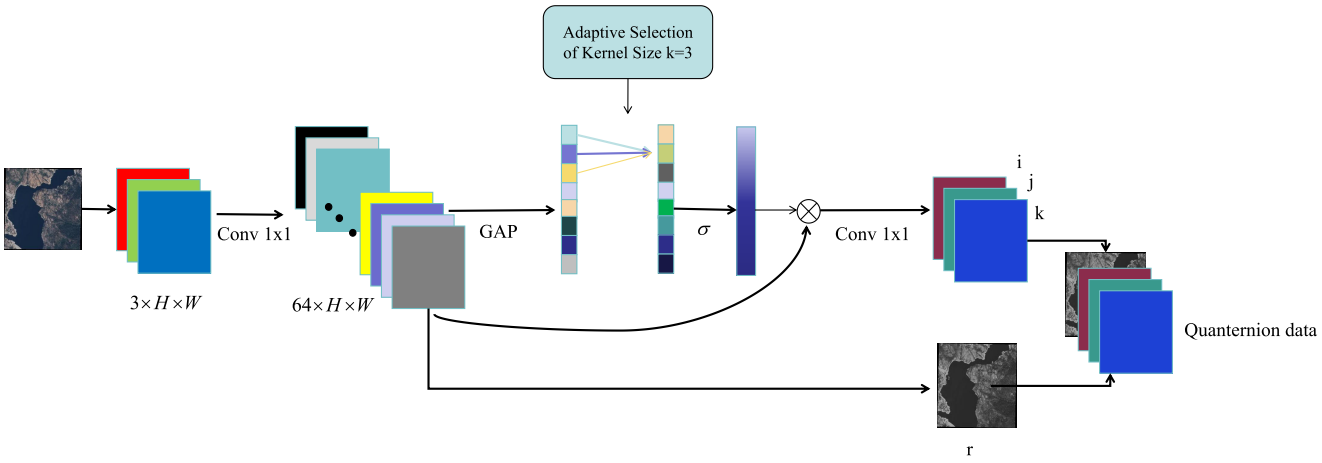

Fig. 3. Entire operating procedures of the QIM.

this limitation reduces their degrees of freedom, yet notably enhances their capacity to encapsulate intricate interrelations among the three channels.

D. MSAA Module

The role of MSAA module is to identify water bodies with high local similarity by obtaining spatial information on a larger scale. This module innovatively uses a multiscale downsampling method to aggregate global information and compare it with local details to obtain local similarity. The details of MSAA are shown in Fig. 4. We first specify a series of downsampling rates $D = (d_1, d_2, \ldots, d_i)$. In this article, the default value of D is $(2, 4, 8)$. Given a feature map $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, the feature maps that aggregate information at different scales can be calculated as follows:

$$\mathbf{F}_i = \mathrm{Up}(\mathrm{Avg}(\mathbf{F}, d_i), d_i) \tag{17}$$

where Avg denotes the average pooling operation of the scale d_i, Up is the d_i-fold nearest-neighbor interpolation upsampling method, and $\mathbf{F}_i \in \mathbb{R}^{C\times H\times W}$. Then the local similarity between each $\mathbf{F}_i$ and the original feature map $\mathbf{F}$ will be calculated

$$\begin{aligned}\mathbf{S}_i &= \mathrm{CosSim}_c(\mathbf{F}, \mathbf{F}_i)\\ \mathrm{CosSim}_d(\mathbf{A}, \mathbf{B}) &= \frac{\sum_{i=1}^{d} \mathbf{A}_i \cdot \mathbf{B}_i}{\sqrt{\sum_{i=1}^{d} \mathbf{A}_i^2} \cdot \sqrt{\sum_{i=1}^{d} \mathbf{B}_i^2}}\end{aligned} \tag{18}$$

where CosSim_c represents the cosine similarity function calculated on the c channel dimension. Attention scores and output results can be obtained as follows:

$$\begin{aligned}\mathbf{S} &= \mathrm{SUM}(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_i)\\ \mathbf{Atten} &= \sigma(\mathrm{Conv2D}_1(\mathrm{Relu}(\mathbf{S})))\\ \mathrm{MSAA}(\mathbf{F}) &= \mathbf{F} \odot \mathbf{Atten}.\end{aligned} \tag{19}$$

where σ denotes the sigmoid function, $\odot$ is the element-wise product operation, and $\mathbf{S}, \mathbf{Atten} \in \mathbb{R}^{1\times H\times W}$.

To guarantee a high signal value for water calculation within this module when searching for water information in a limited range, we establish a minimum downsampling ratio of 2. The progressively expanding scale will subsequently reinforce the high-signal water body information computed at the minimum region similarity, effectively attenuating the surrounding irrelevant ground feature signals. Inspired by ConvNext [48] and LSKNet [49], we design a Qattention block based on MSAA and quaternion convolution. The details of the Qattention block are shown in Fig. 5. Each Qattention block consists of two parts: MSAA operation and feed-forward network (FFN). The FFN is employed to effectively blend channels and enhance the refinement of features. It is noteworthy that we employ the 1×1 quaternion convolution operation to replace the commonly utilized fully connected layer.

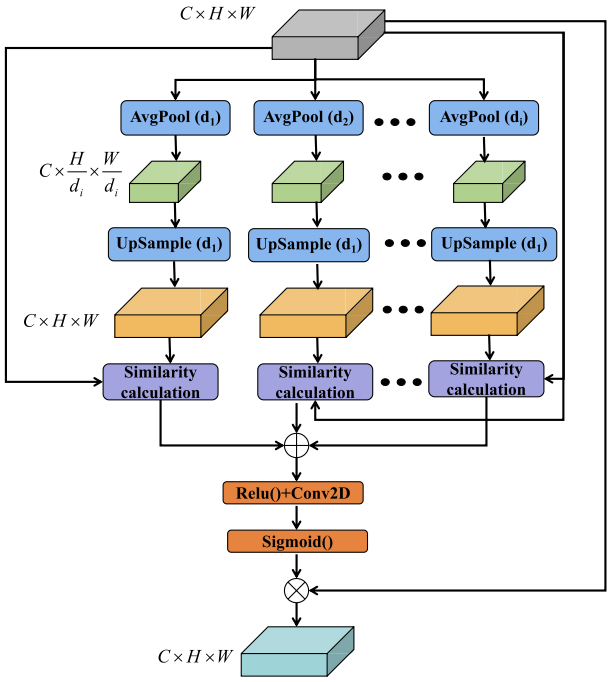

Fig. 4. Details of the proposed MSAA.

E. Decoder

The principal role of the decoder is to merge shallow and deep feature maps while progressively restoring the feature map to the size of the original image. The core of the

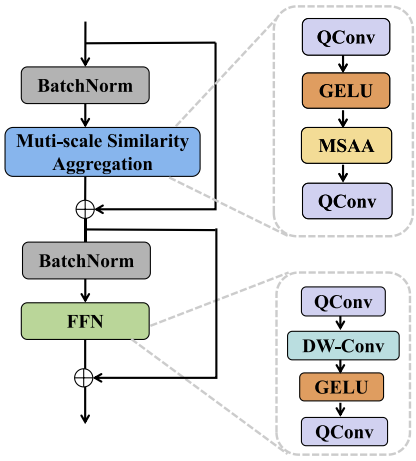

Fig. 5. Structure of Qattention Block.

decoder is four levels of quaternion convolution blocks. Each block has two inputs: a feature map from the encoder in the same level and a feature map 2× upsampling from the lower level. Skip connections between the two feature maps can retain more detailed information. In each block, the model sequentially performs the operation of 2× upsampling, concat, the operation of 3 × 3 convolution, and Relu().

F. Quaternion Transformer-Based U-Net

QTU-Net is a U-shaped symmetric neural network for WBE. The encoder generates five distinct levels of feature maps $\mathbf{F}_i \in \mathbb{R}^{C_i \times (H/2^{i-1}) \times (W/2^{i-1})}$, where C_i represents the feature-channel number of the ith level feature map. Each ith feature extraction layer integrates a quaternion convolution block and multiple layers (L_i) of quaternion attention blocks. The detailed configurations of different variants of QTU-Net used in this article are shown in Table I.

IV. Experiments

A. Datasets and Experimental Settings

1) Wuhan Dense Labeling Dataset: The Wuhan dense labeling dataset (WHDLD) [50], [51] serves as the first water body dataset employed in our experimental study. WHDLD is composed of RGB satellite images, captured by the GaoFen-1 and ZiYuan-3 satellites. The study area under consideration spans the urban landscape of Wuhan, China. In alignment with the multicategory labels thoughtfully provided by Wuhan University, our research focuses intently on the segmentation and comprehensive analysis of water bodies, constituting the principal subject of our inquiry. WHDLD encompasses a total of 4940 RGB images, each meticulously standardized with spatial dimensions measuring 256 × 256 pixels and a resolution of 2 m. To facilitate our research and ensure robust model evaluation, we partitioned the dataset into distinct subsets. Specifically, 3458 images have been allocated to our training set, 988 images to the validation set and 494 images to the testing set.

2) Water Bodies Dataset: The Water Bodies dataset, obtainable from Kaggle and downloadable at the following: https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies, encompasses a collection of satellite images capturing various water bodies. These images are acquired by the Sentinel-2 satellite, and the masks highlighting the water bodies are generated using the NDWI. In total, the dataset comprises 2841 images, each of distinct dimensions. To ensure uniformity in our analysis, we have resized all images to a standard dimension of 256 × 256 pixels. Our research employs a subset of 1989 images for the training set, 568 images for the validation set, and 284 images for the test set, facilitating the development and comprehensive evaluation of our model.

3) GLH-Water Dataset: The GLH-Water dataset [52] is a large-scale dataset for global surface water detection in large-size VHR optical satellite imagery. Comprising 250 VHR images with a ground sampling distance (GSD) of 0.3 m and a resolution of 12 800 × 12 800 pixels, it encompasses a diverse array of water bodies around the world. Each image has undergone meticulous manual labelling and expert scrutiny to ensure the accuracy of annotations. The dataset has been systematically partitioned into three subsets: 80% for training, 10% for validation, and another 10% for testing. To manage computing resources efficiently, we chose to crop each original large image into nonoverlapping 512 × 512 pixel segments. This approach balances the effectiveness of the training process with robust evaluation. Importantly, the validation and test sets are geographically dispersed across various regions worldwide, reflecting real-world conditions and enabling a comprehensive assessment of the model's performance. What sets the GLH-Water dataset apart from others is its emphasis on large-scale imagery, a substantial number of samples, extensive geographical coverage, a broad temporal span of image acquisition, and the inclusion of diverse surface water types. These unique characteristics render it an invaluable resource for researchers and practitioners in surface water detection and related fields.

4) Loss Function: In our research, a mixed loss comprising Dice loss and cross-entropy loss is utilized in the experiments. The mixed loss holds the capability to balance the weighting of local and global information effectively. Dice loss plays a pivotal role in addressing the challenge of imbalanced class proportions between the target and background. At the same time, the regularization effect engendered by cross-entropy loss serves as a robust preventative measure against overfitting phenomena, enhancing the overall robustness and stability of the training process. The total loss can be calculated as follows:

$$\begin{aligned}
&\text{Loss} = \lambda_1 \text{Loss}_{\text{ce}} + \lambda_2 \text{Loss}_{\text{dice}} \\
&\text{Loss}_{\text{dice}}(Y, P) = 1 - \frac{2\sum_i^{H\times W} Y_i P_i}{\sum_i^{H\times W} Y_i^2 + \sum_i^{H\times W} P_i^2 + \epsilon} \\
&\text{Loss}_{\text{ce}}(Y, P) \\
&= -\sum_{i=1}^{H}\sum_{j=1}^{W} \left[P_{I,j} \times \log Y_{i,j} + \left(1 - P_{i,j}\right) \log\left(1 - Y_{i,j}\right)\right]
\end{aligned} \tag{20}$$

where H and W represent the height and width of satellite images, respectively. Y denotes the true label and P denotes the prediction results. The parameters λ_1 and λ_2 are weight

TABLE I
VARIANTS OF QTU-NET USED IN THIS ARTICLE

Model	QTU-Net-T	QTU-Net	QTU-Net-L
$\{L_1, L_2, L_3, L_4, L_5\}$	{1, 1, 1, 1, 1}	{2, 2, 2, 2, 2}	{2, 2, 2, 2, 2}
$\{C_1, C_2, C_3, C_4, C_5\}$	{16, 32, 64, 128, 256}	{16, 32, 64, 128, 256}	{32, 64, 128, 256, 512}
Parameters (MB)	12.57	25.15	100.58
FLOPs (G)	12.21	24.31	96.90

factors that have been judiciously assigned values within the range of 0–1. Specifically, λ_1 is configured to a value of 0.4, while λ_2 has been set to 0.6. These values have been meticulously selected to achieve the desired balance and influence between the weight factors in the context of our experiments.

5) Implementation Details: In all of our experiments on WHDLD and Water Bodies datasets, we have incorporated random rotation and flipping techniques to enhance data diversity. The model proposed in our research is trained on a single NVIDIA GeForce RTX 4090 GPU with 24 GB of memory. In order to expedite the convergence of the training process, we have employed the Adam gradient descent optimizer, which is initialized with a learning rate of $5e-4$, a momentum value of 0.9, and a weight decay of $1e-5$. These optimizations are instrumental in achieving faster and more efficient training outcomes. For the GLH-Water dataset, we adopted a pretraining strategy for the encoder part of the QTU-Net model on the ImageNet-1K [53] dataset for 90 epochs, utilizing 8 NVIDIA A40 GPUs. The pretraining process employed the SGD optimizer with a batch size of 128 and an initial learning rate of 0.1. In contrast, for the backbones of the other comparative models, we consistently utilized pretrained weights provided by the official PyTorch repository. Subsequently, all models were trained on the GLH-Water dataset for 15 epochs, leveraging two NVIDIA RTX A6000 GPUs with a uniform batch size of 16. The Adam optimizer was utilized with a momentum of 0.9, a decayed weight of 1×10^{-5}, an initial learning rate of 2×10^{-4}, and a linear learning rate decay strategy. These training strategies are designed to ensure the effective learning of features specific to water body detection, leading to robust and accurate segmentation results.

6) Metrics: In the experiments, we conducted a comprehensive evaluation using three metrics to assess the performance of our models [Hausdorff distance (HD), Dice similarity coefficient (DSC), and Intersection over Union (IoU)]. The HD metric served as a measure of the maximum dissimilarity between two sets, allowing us to assess the localization accuracy achieved by our segmentation approach. This metric is particularly useful in evaluating edge matching algorithms. Concurrently, the DSC, which measures the spatial overlap between predicted and ground-truth masks, provides insights into the overall segmentation accuracy. Additionally, the IoU enabled us to assess the spatial intersection between the predicted and true binary masks. These metrics collectively afforded us a holistic understanding of our models' segmentation prowess, encompassing aspects of localization accuracy, spatial overlap, and segmentation precision, thereby facilitating a comprehensive evaluation of their performance. The calculation methods for DSC, HD, and IoU are as follows:

$$\begin{aligned}
\mathrm{DSC} &= \frac{2 \times \mathrm{TP}}{\mathrm{FP} + \mathrm{FN} + 2 \times \mathrm{TP}} \\
\mathrm{IoU} &= \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{FN} + \mathrm{TP}} \\
\mathrm{HD}(\boldsymbol{Y}, \boldsymbol{P}) &= \max(h(\boldsymbol{Y}, \boldsymbol{P}), h(\boldsymbol{P}, \boldsymbol{Y})) \\
h(\boldsymbol{Y}, \boldsymbol{P}) &= \max_{y \in \boldsymbol{Y}} \left\{ \min_{p \in \boldsymbol{P}} \|y - p\| \right\} \\
h(\boldsymbol{P}, \boldsymbol{Y}) &= \max_{p \in \boldsymbol{P}} \left\{ \min_{y \in \boldsymbol{Y}} \|p - y\| \right\}
\end{aligned} \tag{21}$$

where $\boldsymbol{P}$ denotes the prediction results and $\boldsymbol{Y}$ denotes the true label. TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

B. Comparison Methods

1) *U-Net [25]:* A symmetrical encoder–decoder semantic segmentation model that integrates both deep and shallow features for improved segmentation outcomes.
2) *FCN [24]:* Feature extraction is performed using full convolutional layers, and the ultimate segmentation result is attained through skip connections.
3) *SegNet [29]:* It is an encoder–decoder model that integrates pooling index operations during upsampling to achieve efficient segmentation.
4) *R2U-Net [54]:* It enhances feature representation by cyclically stacking residual convolutional layers.
5) *DeepLabV3+ [55]:* It employs dilated convolution and multiscale information fusion, making it suitable for the segmentation of large-scale high-resolution images.
6) *Swin-Unet [56]:* It substitutes the convolutional blocks in U-Net with Swin transformer [57] blocks to extract both local and global features, resulting in a segmentation model with a pure transformer architecture. Each Swin transformer block comprises a window-based multihead self-attention module and a shifted window-based multihead self-attention module.
7) *TransUNet [58]:* It utilizes ResNet for extracting low-level features, a ViT for high-level feature extraction, and integrates local details with global information.
8) *MF-SegFormer [35]:* It incorporates a feature fusion module to enhance edge features building upon SegFormer [59], and enhances the extraction of small water bodies through the atrous spatial pyramid pooling module.

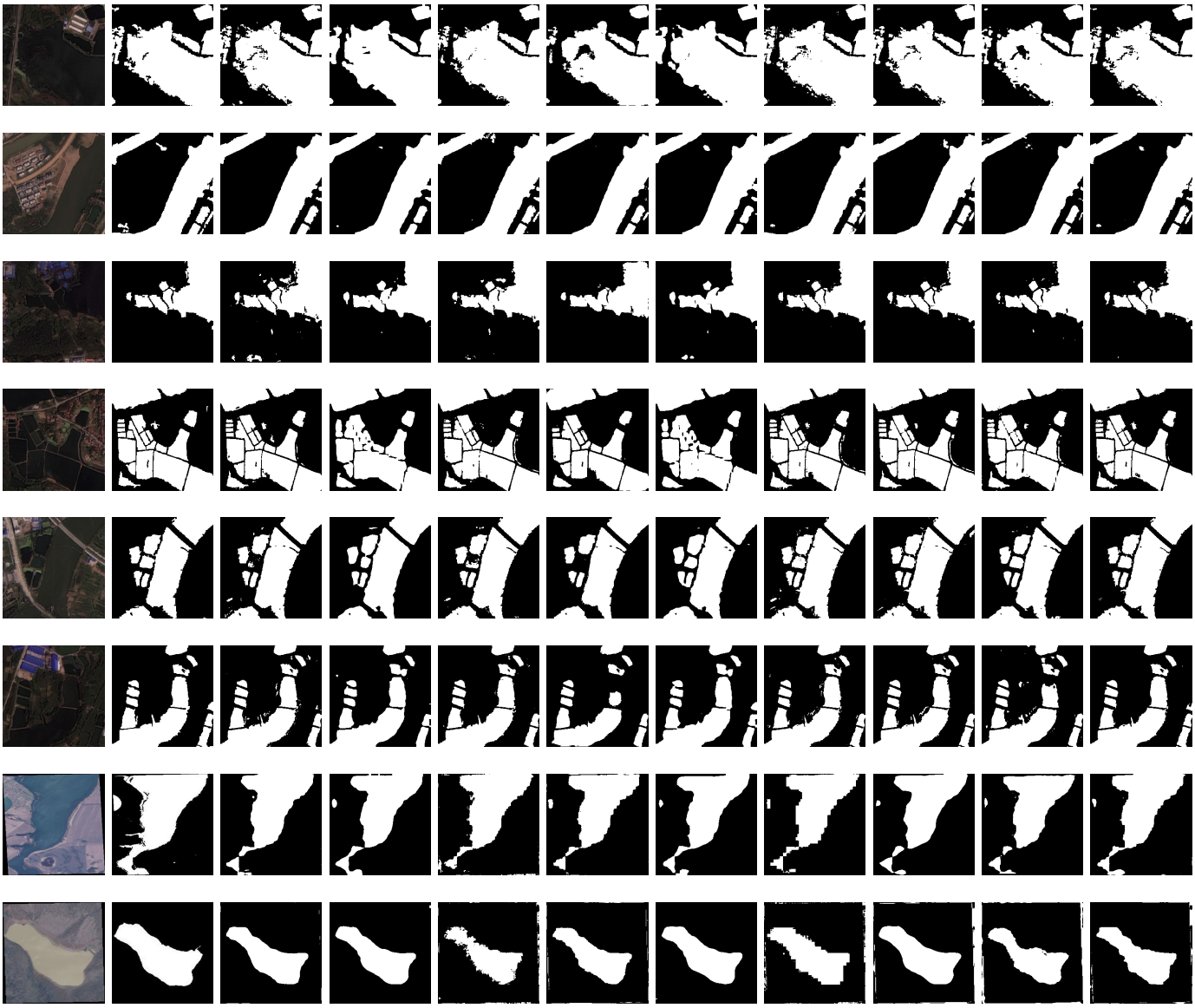

Fig. 6. Experimental results of different semantic segmentation methods on WHDLD and Water Bodies datasets: (a) original image, (b) true segmentation results, (c) U-Net, (d) FCN, (e) SegNet, (f) R2U-Net, (g) DeepLabV3+, (h) Swin-Unet, (i) TransUNet, (j) MF-SegFormer, and (k) proposed QTU-Net.

C. Experiment Results

In this section, we conducted a comprehensive comparative analysis of our proposed method against eight state-of-the-art segmentation approaches. These encompass five CNN-based methods, namely U-Net [25], FCN [24], SegNet [29], R2U-Net [54], and DeepLabV3+ [55], as well as three transformer-based methods, which include Swin-Unet [56], TransUNet [58], and MF-SegFormer [35]. The results of these extensive experiments are methodically presented in Tables II and III, demonstrating the performance evaluation outcomes on the WHDLD and Water Bodies datasets, respectively. This comparative analysis underscores the efficacy of our proposed method relative to other leading segmentation techniques, providing insights into its strengths and capabilities in the domain of water bodies extraction. Notably, in the tables, the best results achieved are emphasized using bold typeface.

From Tables II and III, we can find that our proposed QTU-Net has achieved state-of-the-art performance across the three evaluation metrics, underscoring its excellence in WBE task. Specifically, in terms of the DSC metric, QTU-Net surpasses SegNet by an impressive 1.5% on the WHDLD dataset and outperforms TransUNet by 0.53% on the Water Bodies dataset. For the IoU metric, QTU-Net exhibits a noteworthy superiority, surpassing TransUNet by 2.04% on WHDLD and by 0.56% on the Water Bodies dataset. In the context of the HD metric, QTU-Net achieves remarkable results, outperforming TransUNet by a substantial margin of 3.68% on the WHDLD dataset. These findings unequivocally establish the robust segmentation capabilities of our proposed model in the domain of WBE. This is because QTU-Net extracts the water body pixels from quaternion domain, which could explore the dependencies between channels. And the proposed QTU-Net, established on MSAA component, captures local

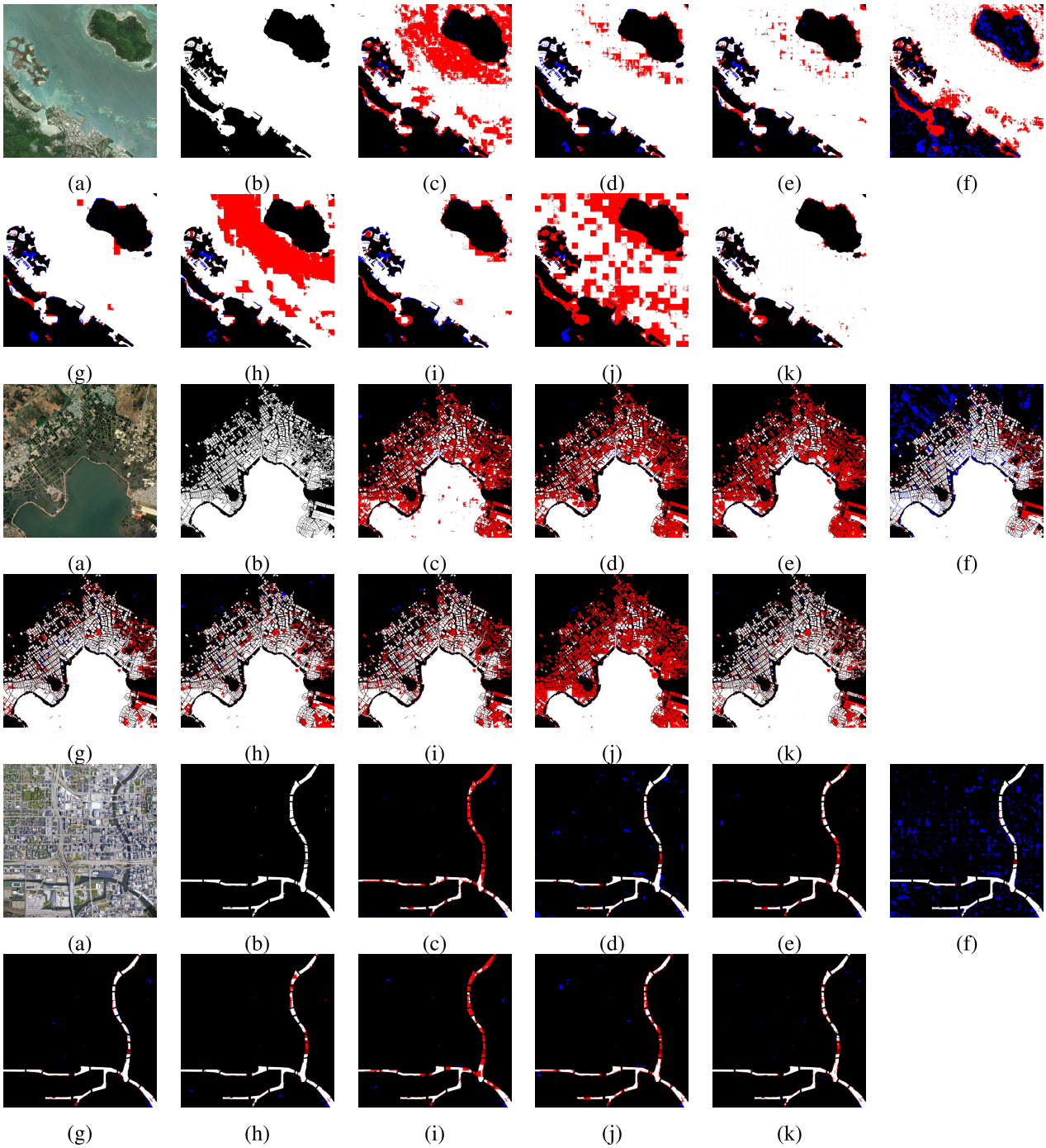

Fig. 7. Experimental results of different semantic segmentation methods on GLH-Water dataset: (a) original image, (b) true segmentation results, (c) U-Net, (d) FCN, (e) SegNet, (f) R2U-Net, (g) DeepLabV3+, (h) Swin-Unet, (i) TransUNet, (j) MF-SegFormer, and (k) proposed QTU-Net. Blue pixels represent false positive areas, while red pixels represent false negative areas.

similarity information from multiscale regions, enhancing spatial features in the image and producing improved segmentation results. It is evident that Swin-Unet and TransUNet demonstrate superior performance in water segmentation tasks compared to traditional CNN segmentation models. This is due to the enhanced global modeling capability provided by their transformer layers.

Moreover, in our visual assessment of the segmentation results across all methods on the two datasets (as depicted in Fig. 6), it is evident that QTU-Net excels in capturing both large and small water bodies with remarkable accuracy. The segmentation maps produced by QTU-Net exhibit superior delineation of water body contours and edge details. In contrast, other CNN-based segmentation methods, including U-Net and DeepLabV3+, struggle to deliver precise segmentation results for the edge regions of water and land, providing only approximate water body contours. Transformer-based methods, represented by MF-SegFormer in the sixth image, are prone to producing false positives during WBE, as indicated in the segmentation image. The phenomenon under observation

TABLE II
RESULTS ON WHDLD DATASET

Method	Backbone	DSC↑	IoU↑	HD↓
U-Net	-	0.7647	0.6805	36.13
FCN	ResNet50	0.7519	0.6612	37.80
SegNet	-	0.7885	0.6720	41.84
R2U-Net	-	0.6975	0.6082	51.75
DeepLabV3+	ResNet50	0.7525	0.6631	36.51
Swin-Unet	-	0.7840	0.6832	36.38
TransUNet	ResNet50	0.7833	0.6928	35.61
MF-SegFormer	MiT-B5	0.7312	0.6211	40.39
QTU-Net	**-**	**0.8041**	**0.7132**	**31.93**

TABLE III
RESULTS ON WATER BODIES DATASET

Method	Backbone	DSC↑	IoU↑	HD↓
U-Net	-	0.7303	0.6170	40.45
FCN	ResNet50	0.7181	0.6091	67.84
SegNet	-	0.7424	0.6254	**36.22**
R2U-Net	-	0.7035	0.5912	42.48
DeepLabV3+	ResNet50	0.7156	0.6075	66.95
Swin-Unet	-	0.7466	0.6323	37.30
TransUNet	ResNet50	0.7473	0.6347	36.60
MF-SegFormer	MiT-B5	0.7339	0.6183	36.44
QTU-Net	**-**	**0.7526**	**0.6403**	36.75

arises from the capacity of QTU-Net to employ quaternion convolution operations in modeling the internal channels of RGB. This modeling capability facilitates the extraction of semantic information among different features in remote sensing images from a novel standpoint, thereby mitigating the occurrence of missed and erroneous detections. Additionally, the MSAA operation in the spatial domain enhances the model's improved ability to capture details at the edges of water bodies. These visual comparisons vividly reinforce the superior performance of QTU-Net in water body segmentation tasks.

To further demonstrate the high applicability of our proposed QTU-Net in WEB tasks, we conducted experiments on the large-scale global water detection satellite dataset GLH-Water. Comprising ultrahigh-resolution satellite images of surface water from diverse geographical regions, the GLH-Water dataset underscores the robustness and versatility of our approach. As detailed in Table IV, our method exhibits remarkable competitiveness. Specifically, in terms of the DSC metric, QTU-Net outperforms DeepLabV3+ by 1.98% and TransUnet by 0.17%. Regarding the HD metric, owing to the vast image dimensions and the variability in water area distribution within the dataset, the distribution of predicted water anomalies is inconsistent, thereby elevating the HD values. In Fig. 7, we illustrate the visual predictions of each model. Overall, QTU-Net and the DeepLabV3+ model exhibit superior segmentation performance on this dataset, demonstrating greater adaptability to more complex

TABLE IV
RESULTS ON GLH WATER DATASET

Method	Backbone	DSC↑	IoU↑	HD↓
U-Net	-	0.6839	0.5195	1118.33
FCN	ResNet50	0.7924	0.6562	1157.17
SegNet	-	0.7727	0.6296	1216.95
R2U-Net	-	0.5599	0.3888	1760.69
DeepLabV3+	ResNet50	0.8219	0.6977	**858.08**
Swin-Unet	-	0.7698	0.6257	1087.5
TransUNet	ResNet50	0.8400	0.7241	869.73
MF-SegFormer	MiT-B5	0.7593	0.6120	1165.89
QTU-Net	**-**	**0.8417**	**0.7267**	1063.49

scenarios. Notably, in the first image, QTU-Net achieves the segmentation of most water bodies with minimal false-negative pixels, in contrast to TransUnet, SegNet, U-Net, and other models. This proficiency stems from the strong color contrast between the lake and its surrounding structures, which our quaternion convolution effectively captures, enabling robust color modeling and improved discrimination of water bodies from other objects. The second and third images represent chaotic small water body scenes and strip water body scenes, respectively. Our MSAA module adeptly extracts similarity features across multiple scales, thus empowering the model to exhibit excellent fitting performance for diverse water body shapes and sizes. In conclusion, our QTU-Net model consistently demonstrates robust performance on large-scale water datasets, making it a broadly applicable approach for WEB tasks.

D. Ablation Studies

In this section, to further validate the effectiveness of our proposed method, we conducted numerous ablation studies and reported the experimental results on the WHDLD dataset.

1) Ablation Study on Quaternion Convolution, QIM, and Qattention Block: In this research, we use the U-Net model as the baseline, mirroring the identical number ($C = \{16, 32, 64, 128, 256\}$) of channels per layer as our QTU-Net architecture. Our investigation unfolded in distinct phases. First, we replaced the conventional convolution operation in the U-Net model with the quaternion convolution operation to gauge its impact on the segmentation task. Subsequently, we introduced the QIM to investigate its role in initializing quaternion elements, a crucial aspect of our methodology. Finally, we incorporated the innovative Qattention block, built upon the principles of MSAA, to further enhance the segmentation capabilities of our model. This structured approach allowed us to systematically assess the contributions and effectiveness of each component introduced in the progression of our research. The results are shown in Table V. It can be seen that QCov and QIM can improve the performance of the model by about 1% on the three indicators of DSC, IoU, and HD. This improvement can be attributed to the efficacy of quaternion convolution operations, which facilitate the generation of adaptive interactions among the three RGB channels, thereby enabling the capture of intricate and

TABLE V
ABLATION STUDY ON QUATERNION U-NET

QConv	QIM	Qattention	DSC↑	IoU↑	HD↓	P(MB)↓	F(G)↓
×	×	×	0.7647	0.6805	36.13	0.82	2.00
✓	×	×	0.7737	0.6823	35.25	**0.21**	**0.51**
✓	✓	×	0.7834	0.6981	34.42	0.21	0.85
✓	✓	✓	**0.7882**	**0.7048**	**30.15**	12.57	12.21

TABLE VI
ABLATION STUDY OF DIFFERENT ATTENTION MECHANISMS IN QTU-NET

Method	DSC↑	IoU↑	HD↓	P(MB)↓	F(G)↓
SE	0.7869	0.6997	33.67	25.17	24.31
SK	0.7945	0.7097	30.65	39.91	38.46
CBAM	0.7912	0.7087	**30.44**	25.17	24.33
Shuffle	0.8001	0.7093	32.12	25.15	24.31
Triplet	0.7919	0.7073	34.76	25.15	24.33
MSAA	**0.8041**	**0.7132**	31.93	**25.15**	**24.31**

TABLE VII
ABLATION STUDY OF NUMBER OF SCALES IN MSAA

D	DSC↑	IoU↑	HD↓	P(MB)↓	F(G)↓	FPS↑
{2,4}	**0.8140**	0.7126	32.20	24.31	25.15	**51.12**
{2,4,8}	0.8041	0.7132	31.93	24.31	25.15	43.45
{2,4,8,16}	0.8086	**0.7157**	**31.24**	24.31	25.15	42.61

nuanced connections between them. The Qattention block has greatly improved the model. By leveraging the principles of MSAA, it adeptly aggregates multiscale feature information, thereby empowering the model to obtain more precise water body delineations from a holistic similarity perspective. The introduction of QIM results in a negligible increase in the number of model parameters, indicating that our QIM is a relatively lightweight but effective module that can generate quaternion data containing RGB channel initialization weights. There is an interesting phenomenon that when we replace the original convolution operation of U-Net with quaternion convolution, the number of parameters and FLOPs decrease by about four times. This is due to the special operation form of quaternion convolution. Consider a traditional convolutional operation with an input channel depth of inC, an output channel depth of outC, and a convolutional kernel size of K, resulting in a parameter quantity of $\text{inC} \times \text{outC} \times K \times K$. With this configuration maintained, the quaternion convolution's parameter count is specified as $(\text{inC}/4) \times (\text{outC}/4) \times K \times K \times 4 = ((\text{inC} \times \text{outC} \times K \times K)/4)$. This indicates that quaternion convolution operation produces better water body recognition performance at a smaller parameter level, and it has great potential to be explored in lightweight WBE tasks in the future.

2) Ablation Study of Different Attention Mechanism: In order to further validate the effectiveness of MSAA, we compared this attention mechanism with five other plug-and-play, lightweight attention mechanisms, including squeeze and extend (SE) [60], selective kernel (SK) [61], convolutional block attention module (CBAM) [62], shuffle attention [63], and triple attention [63]. We followed the network structure of QTU-Net in our research, replacing the original MSAA with other attention mechanisms, and recorded the experimental results in Table VI. The SE attention mechanism is designed to generate interaction between channels by obtaining channel weights. However, quaternion convolution operations take into account the entire RGB color space and explore complex connections among its three components, which can lead to better channel interaction. While other attention mechanisms also consider spatial contextual information, they disregard similarity information between different scale regions. Our MSAA outperforms these methods in water extraction tasks, as demonstrated by our results. The SK attention mechanism employs convolutional operations utilizing varying kernel sizes, which inherently leads to an increase in both parameter count and computational complexity, when compared to other types of attention mechanisms. In contrast, MSAA and similar mechanisms are designed to be highly efficient, requiring minimal parameters while still managing to yield substantial performance enhancements.

3) Ablation Study of Number of Scales in MSAA: Within the MSAA, the process of aggregating similarity information across different scales bestows varying degrees of receptive fields upon the network. This, in turn, empowers the network to access global contextual information spanning different ranges. However, it is imperative to strike a delicate balance in the adjustment of the receptive field size. Aggregating information on an excessively large scale can introduce extraneous information that is unrelated to the segmented target, potentially leading to noise and decreased precision. Conversely, aggregating information on too small a scale may fall short of acquiring the essential contextual information necessary for the task at hand. In this study, we specified three different scales (downsampling rates) $D = (d_1, d_2, \ldots, d_i)$ and recorded the experimental results in Table VII. It can be seen that as the number of scales within the MSAA mechanism gradually increases, there is a discernible and consistent improvement in the average performance of the QTU-Net model across the three key performance indicators. Given that MSAA is an attention mechanism characterized by a minimal parameter count, its underlying principle capitalizes on the prior knowledge that water bodies exhibit local aggregation tendencies within a given scene. Consequently, augmenting the downsampling frequency within MSAA does not substantially augment the overall quantity of model parameters or FLOPs. However, the incremental computational demands do result in a heightened duration for both the model's inference and training phases.

4) Ablation Study of Parameters: In this section, we explore the influence of parameter quantities on the network presented in this article. To this end, we present three distinct variants of the QTU-Net model, each characterized by different parameter quantities. The configuration details of these three network structures are thoughtfully elaborated in the section for comprehensive reference. The experimental results and

TABLE VIII
ABLATION STUDY OF DIFFERENT PARAMETERS IN QTU-NET

Method	Parameters (MB)	FLOPs(G)	DSC↑	IoU↑	HD↓
QTU-Net-T	12.57	12.21	0.7882	0.7048	30.15
QTU-Net	25.15	24.31	0.8041	0.7132	31.93
QTU-Net-L	100.58	96.60	**0.8098**	**0.7187**	**29.48**

findings pertaining to these variants are thoughtfully reported in Table VIII. The analysis reveals a notable trend: as the number of network layers deepens and the quantity of parameters increases, there is a consistent and discernible rise in the values of DSC and IoU. This observation suggests that augmenting the number of parameters to some extent enhances the model's performance in WBE tasks. It is important to note that this performance improvement comes at the cost of heightened computational complexity, potentially leading to reduced inference speed. Hence, there exists a trade-off between model performance and computational efficiency, necessitating a careful consideration of these factors in the design and deployment of the network for specific applications.

5) Ablation Study of the Weight Parameters in Loss: In this study, the loss function utilized across all experiments is a hybrid loss, which combines cross-entropy loss and Dice loss. To determine the optimal weighted ratio between these two components, we conducted a quantitative exploration of the numerical integration of λ_1 and λ_2. Within each experiment, we assigned values to λ_1 at increments of 0.2, while setting λ_2 to $1 - \lambda_1$. Simultaneously, we evaluated whether the two types of losses could achieve the most effective training outcome at a 1:1 ratio by designating a specific ratio where $\lambda_1 = \lambda_2 = 0.5$. We employed IoU as the metric for selecting these two hyperparameters.

It can be seen from Fig. 8 that our network's peak performance is realized when the parameters (λ_1, λ_2) are set to $(0.4, 0.6)$. Interestingly, we observe that utilizing cross-entropy loss alone still yields a relatively commendable network performance. This can be explained by the fact that cross-entropy loss facilitates meticulous, pixel-by-pixel optimization of the network, conferring enhanced robustness. Furthermore, considering the substantial presence of nonwater body areas within our actual dataset, there exists a probability that these could overshadow the pixels of the targeted water bodies. By strategically increasing the weightage of Dice loss, we can offset the limitations inherent to cross-entropy loss when managing the issue of class imbalance in our samples. Nevertheless, an exclusive reliance on Dice loss leads to a significant deterioration in network performance, resulting in a final metric that is 2.02% inferior to the benchmark set by the optimal network performance. This decline can be attributed to the lack of pixel-level discrimination due to the exclusion of cross-entropy loss, which in turn curtails the network's capacity for precise data fitting.

E. Complexity Analysis and Limitation

In this section, we calculated the parameters and FLOPs of the QTU-Net we proposed and other comparison models

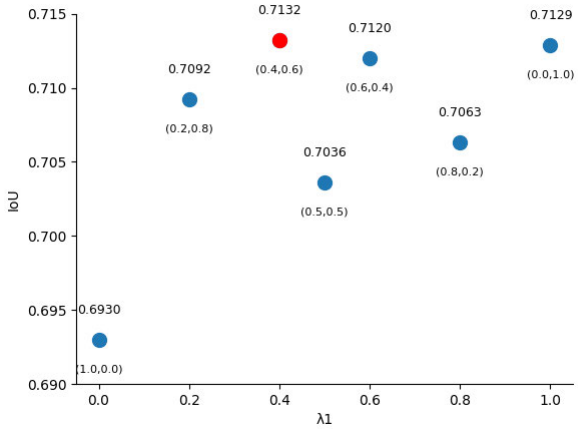

Fig. 8. Impact of values of λ_1 and $\lambda 2$ on the final result.

TABLE IX
COMPLEXITY ANALYSIS OF ALL METHODS

Method	Backbone	Parameters (MB)↓	FLOPs (G)↓	FPS↑
U-Net	-	**0.82**	**2.00**	**274.92**
FCN	ResNet50	34.71	32.94	100.78
SegNet	-	29.44	40.16	165.16
R2U-Net	-	39.09	152.92	61.30
DeepLabV3+	ResNet50	41.03	39.63	93.06
Swin-Unet	-	27.14	7.73	114.19
TransUNet	ResNet50	93.23	32.23	60.37
MF-SegFormer	MiT-B5	24.26	81.87	74.66
QTU-Net	-	25.15	24.31	43.45

in this article. Furthermore, we compared their runtime performance in the inference process when applied to an RGB satellite image of size 256×256 pixels. We introduce frames per second (FPS) as a valuable metric for assessing model runtime efficiency. It can be calculated as follows:

$$\text{FPS} = \frac{1}{\text{Model Runtime (seconds)}}. \tag{22}$$

The results are shown in Table IX. The U-Net model, characterized by its simplicity and minimal number of parameters, naturally achieves the fastest inference speed among the models examined. Conversely, our proposed QTU-Net model, while demonstrating commendable performance in terms of parameters and FLOPs, exhibits a comparatively slower inference speed. This discrepancy in speed can be attributed to the presence of a significant number of depth-wise convolutional operations within the QTU-Net architecture, which substantially increase the input–output (IO) read speed, consequently limiting its overall inference speed. It is important to note that the QTU-Net model's inference speed may be more favorable when deployed on machines with enhanced memory access bandwidth performance. However, a comprehensive exploration of this aspect falls outside the scope of the current article's discussion. Our ongoing research endeavors are dedicated to addressing and optimizing this challenge, with the aim of enhancing the model's inference speed, thus ensuring its applicability in a broader range of real-world scenarios.

V. Conclusion

In this article, we propose a novel QTU-Net for WBE from RGB satellite image in the quaternion domain. The interaction of the three components of RGB enhances the semantic expression of different regions. To capture the intricate connections within the RGB components and initialize their weights effectively, we devise an QIM utilizing quaternion convolution. To obtain water bodies of different scales, we designed an MSAA module. This module utilizes multiscale regional similarity to distinguish water bodies from land and capture water bodies of different sizes. Based on the above modules, we have constructed a U-shaped symmetric network called QTU-Net. Extensive experimental results on the extraction of three sets of water bodies indicate that QTU-Net outperforms state-of-the-art semantic segmentation models. Ablation studies confirm the effectiveness and superior performance of the proposed modules. Our future work will focus on further improving the design of the model and accelerating its inference speed to face the challenge of real-time water monitoring.

References

[1] W. Jiang et al., "Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images," *Remote Sens.*, vol. 10, no. 5, p. 755, May 2018.

[2] Y. Chen, R. Fan, X. Yang, J. Wang, and A. Latif, "Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning," *Water*, vol. 10, no. 5, p. 585, May 2018.

[3] C. Chen, X. He, Y. Lu, and Y. Chu, "Application of Landsat time-series data in island ecological environment monitoring: A case study of Zhoushan Islands, China," *J. Coastal Res.*, vol. 108, no. 1, pp. 193–199, Sep. 2020.

[4] H. Song et al., "HA-Unet: A modified unet based on hybrid attention for urban water extraction in SAR images," *Electronics*, vol. 11, no. 22, p. 3787, Nov. 2022.

[5] B. Brisco, "Mapping and monitoring surface water and wetlands with synthetic aperture radar," in *Remote Sensing of Wetlands: Applications and Advances*. Boca Raton, FL, USA: CRC Press, 2015.

[6] D. Yamazaki, M. A. Trigg, and D. Ikeshima, "Development of a global 90m water body map using multi-temporal Landsat images," *Remote Sens. Environ.*, vol. 171, pp. 337–351, Dec. 2015.

[7] H. Tang, S. Lu, M. H. Ali Baig, M. Li, C. Fang, and Y. Wang, "Large-scale surface water mapping based on Landsat and Sentinel-1 images," *Water*, vol. 14, no. 9, p. 1454, May 2022.

[8] H. Cao, H. Zhang, C. Wang, and B. Zhang, "Operational flood detection using Sentinel-1 SAR data over large areas," *Water*, vol. 11, no. 4, p. 786, Apr. 2019.

[9] S. Klemenjak, B. Waske, S. Valero, and J. Chanussot, "Automatic detection of rivers in high-resolution SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1364–1372, Oct. 2012.

[10] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, May 1996, doi: 10.1080/01431169608948714.

[11] Q. Guo, R. Pu, J. Li, and J. Cheng, "A weighted normalized difference water index for water extraction using Landsat imagery," *Int. J. Remote Sens.*, vol. 38, no. 19, pp. 5430–5445, Oct. 2017.

[12] Y. Zhou, H. Zhao, H. Hao, and C. Wang, "A new multi-spectral threshold normalized difference water index (MST-NDWI) water extraction method—A case study in Yanhe watershed," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 2557–2564, May 2018.

[13] X. Zhang and X. Liu, "Comparative study on extraction of banded water and surface water in urban area based on MNDWI," in *Proc. 3rd Int. Conf. Geol., Mapping Remote Sens. (ICGMRS)*, Apr. 2022, pp. 33–40.

[14] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[15] M. C. R. Cordeiro, J.-M. Martinez, and S. Peña-Luque, "Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with level 2A processors," *Remote Sens. Environ.*, vol. 253, Feb. 2021, Art. no. 112209.

[16] P. Yousefi, H. A. Jalab, R. W. Ibrahim, N. F. M. Noor, M. N. Ayub, and A. Gani, "Water-body segmentation in satellite imagery applying modified kernel kmeans," *Malaysian J. Comput. Sci.*, vol. 31, no. 2, pp. 143–154, Apr. 2018.

[17] G. Sarp and M. Ozcelik, "Water body extraction and change detection using time series: A case study of lake Burdur, Turkey," *J. Taibah Univ. Sci.*, vol. 11, no. 3, pp. 381–391, May 2017.

[18] C. Wang et al., "Toward accurate and efficient road extraction by leveraging the characteristics of road shapes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4404616.

[19] Q. Zhu et al., "Oil spill contextual and boundary-supervised detection network based on marine SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5213910.

[20] M. Talha, F. A. Bhatti, S. Ghuffar, and H. Zafar, "ADU-Net: Semantic segmentation of satellite imagery for land cover classification," *Adv. Space Res.*, vol. 72, no. 5, pp. 1780–1788, Sep. 2023.

[21] Y. Li, J. Luo, Y. Zhang, Y. Tan, J.-G. Yu, and S. Bai, "Learning to holistically detect bridges from large-size VHR remote sensing imagery," 2023, *arXiv:2312.02481*.

[22] J. Luo, X. Yang, Y. Yu, Q. Li, J. Yan, and Y. Li, "PointOBB: Learning oriented object detection via single point supervision," 2023, *arXiv:2311.14757*.

[23] Y. Xie, R. Chen, M. Yu, X. Rui, and X. Du, "Improvement and application of UNet network for avoiding the effect of urban dense high-rise buildings and other feature shadows on water body extraction," *Int. J. Remote Sens.*, vol. 44, no. 12, pp. 3861–3891, Jun. 2023.

[24] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[26] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2017.

[27] Z. Wang, X. Gao, and Y. Zhang, "HA-Net: A lake water body extraction network based on hybrid-scale attention and transfer learning," *Remote Sens.*, vol. 13, no. 20, p. 4121, Oct. 2021.

[28] L. Weng, Y. Xu, M. Xia, Y. Zhang, J. Liu, and Y. Xu, "Water areas segmentation from remote sensing images using a separable residual SegNet network," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 4, p. 256, Apr. 2020.

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[30] B. Wang, Z. Chen, L. Wu, X. Yang, and Y. Zhou, "SADA-Net: A shape feature optimization and multiscale context information-based water body extraction method for high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1744–1759, 2022.

[31] P. Nie, X. Cheng, Z. Song, M. Mao, T. Wang, and L. Meng, "Rethinking BiSeNet: A lightweight network for urban water extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4203910.

[32] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–23.

[33] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[34] H.-F. Zhong, Q. Sun, H.-M. Sun, and R.-S. Jia, "NT-Net: A semantic segmentation network for extracting lake water bodies from optical remote sensing images based on transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5627513, doi: 10.1109/TGRS.2022.3197402.

[35] T. Zhang et al., "Water body extraction of the weihe river basin based on MF-SegFormer applied to Landsat8 OLI data," *Remote Sens.*, vol. 15, no. 19, p. 4697, Sep. 2023.

[36] H. Qi, X. Kong, L. Cheng, J. Hu, and J. Gu, "Addressing fine-grained lake water body extraction: A hybrid approach combining vision transformer and geodesic active contour," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4204614.

[37] M. Jiang et al., "GraphGST: Graph generative structure-aware transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5504016.
[38] Y. Duan, X. Xu, T. Li, B. Pan, and Z. Shi, "UnDAT: Double-aware transformer for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522012.
[39] H. Su, J. Qiu, Z. Tang, Z. Huang, and X.-H. Yan, "Retrieving global ocean subsurface density by combining remote sensing observations and multiscale mixed residual transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4201513.
[40] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding UNet for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408715.
[41] W. R. Hamilton, *Elements of Quaternions*. Minneapolis, MN, USA: Green&Company, 1866.
[42] T. Teramae, T. Matsubara, T. Noda, and J. Morimoto, "Quaternion-based trajectory optimization of human postures for inducing target muscle activation patterns," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 6607–6614, Oct. 2020.
[43] K. Fathian, J. P. Ramirez-Paredes, E. A. Doucette, J. W. Curtis, and N. R. Gans, "QuEst: A quaternion-based approach for camera motion estimation from minimal feature points," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 857–864, Apr. 2018.
[44] S. Zhang, L. Yao, L. Vinh Tran, A. Zhang, and Y. Tay, "Quaternion collaborative filtering for recommendation," 2019, *arXiv:1906.02594*.
[45] H. Li, H. Li, and L. Zhang, "Quaternion-based multiscale analysis for feature extraction of hyperspectral images," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1418–1430, Mar. 2019.
[46] H. Li, H. Huang, Z. Ye, and H. Li, "Hyperspectral image classification using adaptive weighted quaternion Zernike moments," *IEEE Trans. Signal Process.*, vol. 70, pp. 701–713, 2022.
[47] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.
[48] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Nov. 2022, pp. 11976–11986.
[49] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," 2023, *arXiv:2303.09030*.
[50] Z. Shao, K. Yang, and W. Zhou, "Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset," *Remote Sens.*, vol. 10, no. 6, p. 964, Jun. 2018.
[51] Z. Shao, W. Zhou, X. Deng, M. Zhang, and Q. Cheng, "Multilabel remote sensing image retrieval based on fully convolutional network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 318–328, Jan. 2020.
[52] Y. Li, B. Dang, W. Li, and Y. Zhang, "GLH-Water: A large-scale dataset for global surface water detection in large-size very-high-resolution satellite imagery," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, no. 20, pp. 22213–22221.
[53] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2009, pp. 248–255.
[54] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.
[55] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[56] H. Cao et al., "Swin-UNet: UNet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 205–218.
[57] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
[58] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.
[59] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12077–12090.
[60] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[61] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.
[62] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
[63] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239.

Mingzhi Wang received the B.S. degree from Harbin Institute of Technology, Weihai, China, in 2022, where he is currently pursuing the M.S. degree.

His research interests include computer vision, deep learning, and remote sensing.

Chunshan Li (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2005, 2009, and 2014, respectively.

He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Weihai, China. His research interests include deep learning, big data analysis, remote sensing, and LLM-based intelligent decision.

Xiaofei Yang (Member, IEEE) received the B.S. degree from Suihua University, Suihua, China, in 2011, and the M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2014 and 2019, respectively.

He was a Postdoctoral Researcher with the Department of Computer and Information Science, University of Macau, Macau, China, from 2020 to 2023. He is currently with the School of Electronic and Communication Engineering, Guangzhou University, Guangzhou, China. His research interests include semisupervised learning, deep learning, remote sensing, transfer learning, and graph mining.

Dianhui Chu received the Ph.D. degree from the Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, in 2014.

He is now a Professor at Harbin Institute of Technology, Weihai, China. His research interests include service computing, service engineering, big data, and software architecture.

Zhiquan Zhou received the M.S. and Ph.D. degrees in information and communication engineering from Harbin Institute of Technology, Harbin, China, in 1998 and 2004, respectively.

He is currently a Professor with the School of Information Science and Engineering, Harbin Institute of Technology, Weihai, China. His research interests include sensor design and signal processing, ocean surveillance, and communication systems.

Raymond Y. K. Lau (Senior Member, IEEE) is a Professor at the Department of Information Systems, City University of Hong Kong, Hong Kong, China. He is the author of over 200 refereed international journals and conference papers. His research work has been published in renowned journals such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE INTELLIGENT SYSTEMS, and *ACM Transactions on Information Systems*. His research interests include financial technology (fintech), social media analytics, and artificial intelligence (AI) for business.

Mr. Lau is a Senior Member of ACM. He was ranked on Stanford University's Top 2% Scientists List from 2020 to 2023.