Scilight

Artificial Intelligence and Emerging Technologies
https://www.sciltp.com/journals/aiet

Article

CQC-Net: A Lightweight Model Utilizing Consistent Orientational Quaternion Wavelet Convolution for Skin Lesion Segmentation

Hao Tang [1], Guoheng Huang [1,*], Xiaochen Yuan [2], Qi Tao [3], Guo Zhong [4,*] and Baiying Lei [5,*]

[1] The School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
[2] The Faculty of Applied Sciences, Macao Polytechnic University, Macau 999078, China
[3] The Department of Mechanical Engineering (Robotics), Guangdong Technion-Israel Institute of Technology, Shantou 515000, China
[4] The School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou 510006, China
[5] The School of Biomedical Engineering, Shenzhen University, Shenzhen 518000, China
* Correspondence: kevinwong@gdut.edu.cn (H.G.); yb77410@umac.mo (G.Z.); leiby@szu.edu.cn (B.L.)

How To Cite: Tang, H.; Huang, G.; Yuan, X.; et al. CQC-Net: A Lightweight Model Utilizing Consistent Orientational Quaternion Wavelet Convolution for Skin Lesion Segmentation. *Artificial Intelligence and Emerging Technologies* **2026**, *1*(1), 3. https://doi.org/10.53941/aiet.2026.100003

Received: 25 March 2026
Revised: 28 March 2026
Accepted: 31 March 2026
Published: 31 March 2026

Abstract: Accurate and efficient segmentation is crucial for melanoma diagnosis. Recent approaches have shifted from focusing solely on spatial information to incorporating frequency information, such as via wavelet transforms, to balance performance and model complexity. While these methods have demonstrated success, they often overlook intrinsic directional properties of wavelets (vertical, horizontal, and diagonal components) and the interplay between low- and high-frequency components. To address these gaps, we propose an orientation-consistent quaternion convolution (OCQC) module and a quaternion enhanced feedforward network (QEFN), both operating in the quaternion wavelet domain. The OCQC module leverages directional properties of quaternion wavelet transforms, applying direction-specific quaternion convolutional kernels to different frequency components to avoid redundant feature learning. The QEFN uses quaternion depthwise separable convolutions (QDSC) and inverse QDSC (IQDSC) to project features into higher dimensions and reconstruct them, facilitating interaction among frequency bands. By integrating the quaternion wavelet transform (QWT), inverse QWT (IQWT), OCQC, and QEFN, we propose the consistent quaternion convolution neural network (CQC-Net). Extensive experiments show that our method achieves competitive performance while maintaining efficiency, with only 0.86 M parameters and 2.96 G Flops.

Keywords: U-Net; quaternion wavelet; quaternion convolution; lightweight; skin lesion segmentation

1. Introduction

Melanoma is a relatively common type of cancer, and its incidence rate continues to rise [1]. Although public awareness has improved early detection and reduced mortality, early-stage melanoma often exhibits subtle features, making manual dermoscopic inspection time consuming and dependent on expert physicians. To overcome these limitations, deep learning-based approaches have gained prominence in recent years. Architectures like FCN [2], U-Net [3], and Deeplab [4] are widely used for lesion segmentation, with U-Net being particularly influential. Variants such as U-Net++citea7 and U-NetV2 [5] refine skip connections, while TransUNet [6] and Swin-UNet [7] introduce transformer blocks for improved performance. Nonetheless, due to the blurred boundaries of many lesions, several methods [8–12] have been developed to capture subtle differences and locate boundary regions by stacking deep networks or employing various attention mechanisms. Although effective, these approaches often lead to significant memory consumption and high computational complexity, making them less suitable for deployment in

medical applications. A promising solution lies in combining frequency and spatial information using wavelets.

Specifically, conventional discrete wavelet transforms (DWT) decompose an image into four sub-bands without introducing trainable parameters: Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH). The LL band captures low-frequency components reflecting structural and background information, while the LH, HL, and HH bands represent vertical, horizontal, and diagonal high-frequency details such as edges and textures, respectively. These frequency components complement spatial domain representations, helping maintain model performance with reduced complexity. Building on this concept, some methods adopt wavelet transforms as downsampling modules in U-shaped architectures [13–15] to preserve image details, while others leverage high-frequency information to enhance boundary sensitivity [16,17] or separate frequencies for self-supervised learning [18].

However, based on our review of prior works, we identify two key limitations: First, existing methods often process the LH, HL, and HH sub-bands collectively as high-frequency information, neglecting their distinct directional characteristics. Second, the grouping of high-frequency information is achieved by concatenation along the channel dimension, where convolutional operations treat each channel independently, overlooking the complex inter-channel relationships critical for capturing the rich information within each frequency band.

To overcome these issues, we explore the potential of quaternion convolution, which encodes interactions among components and reduces computational cost via Hamilton product [19]. To investigate the first issue, we introduce the orientation-consistent quaternion convolution (OCQC) module, as demonstrated in Figure 1. OCQC module divides wavelet features into four parts: LL, LH, HL, and HH then applies quaternion convolution kernels of different shapes and scales for each subband. This design fully utilizes each frequency component and avoids redundant representations. To address the second issue, we develop the quaternion enhanced feedforward network (QEFN), which includes two quaternion depthwise separable convolutions (QDSC) and one inverse version (IQDSC). QDSC projects features into higher-dimensional space, while IQDSC reconstructs them, enabling effective fusion of multi-frequency features.

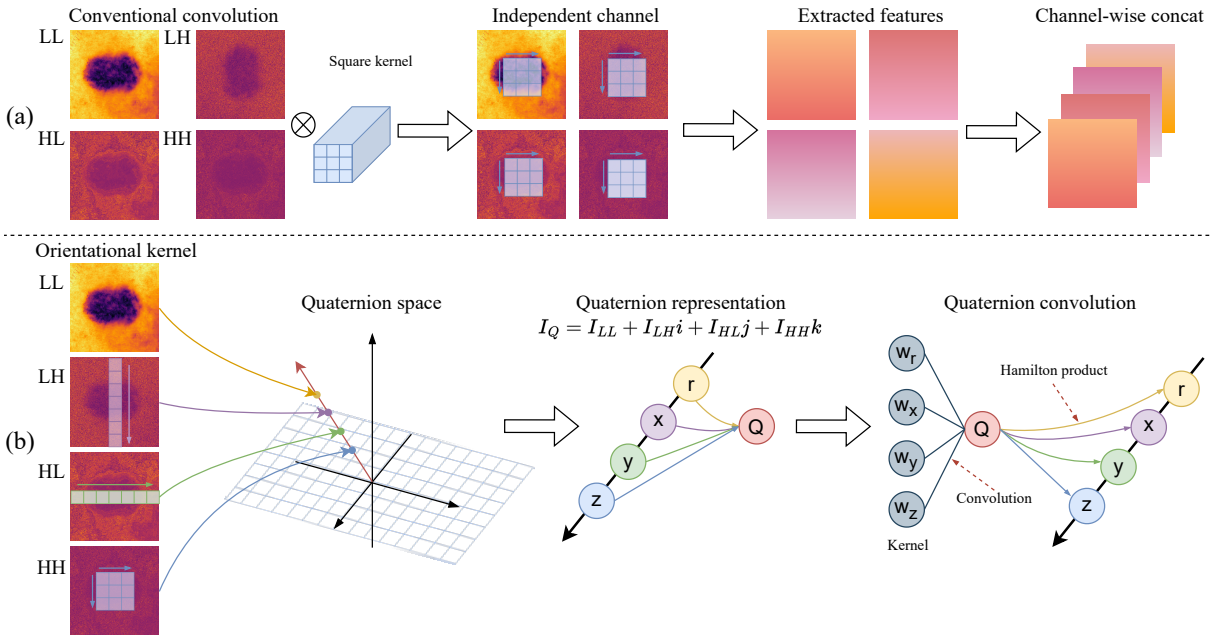

Figure 1. (**a**) Conventional convolution applies identical kernels independently across channels without inter-channel fusion; (**b**) Our orientation-consistent quaternion convolution captures directional features with varied kernel shapes and fuses multi-channel information via quaternion operations, yielding richer structural representations.

To further optimize wavelet representation for quaternion networks, we incorporate the quaternion wavelet transform (QWT) [20], which provides additional sub-bands and translation invariance over standard wavelets. This facilitates robust handling of variably positioned lesions, thereby improving generalization. Implemented in our Consistent Quaternion Convolutional Neural Network (CQC-Net), these techniques yield competitive performance with a lightweight architecture. Our main contributions are as follows:

- We propose OCQC, which employs quaternion convolution kernels of varying shapes to extract directional features from individual sub-bands, reducing redundancy while remaining lightweight.
- We design QEFN to project features into higher-dimensional spaces and reconstruct them, enhancing the utilization of diverse frequency information.

- We integrate QWT instead of conventional wavelet transforms, which hierarchically subdivides the original four sub-bands and removes the shift variance of standard wavelets, improving generalization and enabling effective handling of lesions with varying shapes and positions.
- Built upon the techniques above, CQC-Net offers lightweight design with performance comparable to larger models. Extensive experiments show its effectiveness, achieving IoU scores of 86.15%, 80.78%, and 82.09% on ISIC 2016, ISIC 2017, and ISIC 2018, respectively, outperforming a range of state-of-the-art methods.

2. Related Works

2.1. Wavelet-Based Segmentation

Wavelet transform is highly effective for time-frequency analysis and has been widely applied in computer vision due to its reversibility and ability to preserve information [21,22]. Recently, it has been applied to image segmentation. anu et al. [13] first proposed using discrete wavelet transform for downsampling in U-shaped architectures, arguing that wavelet-based downsampling preserves image details better than common pooling operations, achieving near-lossless results and improved segmentation. Zhao et al. [14] suggested retaining only the LL component during downsampling, as high-frequency components introduce noise. Building on this, Agnes et al. [15] introduced inverse wavelet transform for upsampling in the decoder, enabling nearly lossless transformations. Imtiaz et al. [16] incorporated high-frequency information into a boundary-awareness module with a gated unit, enhancing edge detail extraction. Zhang et al. [17] proposed a high-low frequency attention mechanism, processing the LL component and high-frequency features separately to improve focus. Lin et al. [23] combined DWT with a conventional convolutional encoder for better segmentation of vascular stents in CT images. Despite these advances, these methods typically separate or discard high- and low-frequency information, overlooking the interactions between sub-bands. This paper aims to explore explore how to better leverage these interactions to enhance segmentation performance.

2.2. Quaternion Nerual Networks

Quaternion neural networks (QNNs) have been applied in various domains, including segmentation [24], 3D audio processing [25], and human pose estimation [26]. Gaudet et al. [27] introduced quaternion convolutional neural networks (QCNNs), showing they outperform real-valued CNNs in accuracy with fewer parameters. Tay et al. [28] extended real-valued attention to the quaternion domain, proposing quaternion attention, which achieved both efficiency and high performance. The method was successfully applied to specular highlight removal [29] and single-image dehazing [30], improving performance while reducing parameters. Motivated by the success of quaternions, we use QNNs as the foundation for this study. Current visual applications decompose images into R, G, and B channels as quaternion imaginary parts, with a zero-filled channel as the real part, causing unnecessary overhead. To address this, we extend QNNs to the wavelet domain, utilizing only the intrinsic image information.

3. Methodology

3.1. Basic Quaternion Algebra

A quaternion Q in the quaternion domain $\mathbb{H}$ is a hypercomplex number of rank 4, extending complex numbers non-commutatively. It can be represented as:

$$Q = r + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{1}$$

where r, x, y, z are real numbers, and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are quaternion units bases with $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$. Key operations on quaternions are defined as follows:

Addition: The addition of two quaternions Q and $R = p + l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$ is defined as:

$$Q + R = (r + p) + (x + l)\mathbf{i} + (y + m)\mathbf{j} + (z + n)\mathbf{k} \tag{2}$$

Scalar Multiplication: The Multiplication with scalar β is:

$$\beta Q = \beta r + \beta x\mathbf{i} + \beta y\mathbf{j} + \beta z\mathbf{k} \tag{3}$$

Conjugate: The conjugate Q^{H} is defined as:

$$Q^{\mathrm{H}} = r - x\mathbf{i} - y\mathbf{j} - z\mathbf{k} \tag{4}$$

Hamilton Product: The Hamilton product of Q and R is:

$$\begin{aligned} Q \otimes R = &(rp - xl - ym - zn) \\ &+ (xp + rl - zm + yn)\mathbf{i} \\ &+ (yp + zl + rm - xn)\mathbf{j} \\ &+ (zp - yl + xm + rn)\mathbf{k} \end{aligned} \tag{5}$$

3.2. Overview

The overview of CQC-Net is illustrated in Figure 2. The network features a U-shaped symmetric structure with Quaternion Wavelet Embedding, an encoder, a decoder, and a Quaternion Wavelet Merging module. The Quaternion wavelet embedding involves the Quaternion Wavelet Transform (QWT) and Quaternion Group Normalization (QGN). Given an input image $x \in \mathbb{R}^{H \times W \times 3}$, QWT decomposes the image into 16 sub-bands, resulting in a feature map of size $\frac{H}{2} \times \frac{W}{2} \times 48$. The feature map undergoes extraction through four stages of CQC blocks, each with the OCQC module, QEFN, and residual connections. Max pooling halves the dimensions at each stage, while a 1×1 quaternion depthwise convolution doubles the number of channels. The encoder input dimensions sequentially evolve as $\frac{H}{2} \times \frac{W}{2} \times 48$, $\frac{H}{4} \times \frac{W}{4} \times 96$, $\frac{H}{8} \times \frac{W}{8} \times 192$, and $\frac{H}{16} \times \frac{W}{16} \times 384$. In the decoder, the dimensions follow the reverse progression, with outputs of $\frac{H}{16} \times \frac{W}{16} \times 384$, $\frac{H}{8} \times \frac{W}{8} \times 192$, $\frac{H}{4} \times \frac{W}{4} \times 96$, and $\frac{H}{2} \times \frac{W}{2} \times 48$. The Quaternion Wavelet Merging module then applies the Inverse Quaternion Wavelet Transform (IQWT) to reconstruct the output to $H \times W \times 3$, followed by a 3×3 convolution to reduce the channels to 1, generating the predicted binary mask.

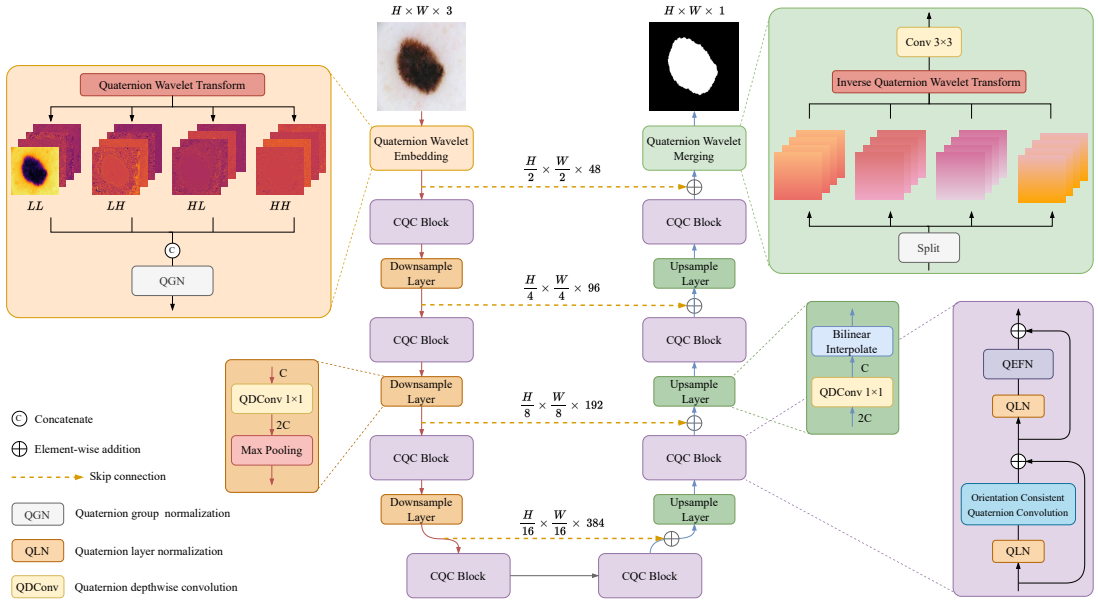

Figure 2. Overview of the Proposed CQC-Net. Our method follows the encoder-decoder architecture of a U-shaped network, where each encoder stage is connected to its corresponding decoder via skip connections. The network consists of several key components, including Quaternion Wavelet Embedding, the OCQC module, max-pooling, bilinear interpolation, and Quaternion Wavelet Merging.

3.3. Quaternion Wavelet Embedding

Our quaternion wavelet embedding consists of two components: QWT and QGN, designed to generate wavelet representations that are suitable for quaternion networks.

3.3.1. Quaternion Wavelet Transform (QWT)

The conventional DWT decomposes an input signal into multiple frequency sub-bands, effectively capturing both spatial and frequency information. This process employs a combination of wavelet filters: low-pass filters, which capture coarse features, and high-pass filters, which extract fine details.

The QWT enhances DWT by introducing four filters for decomposition: F_l, F_h, T_l, and T_h. Here, F_l and F_h are the conventional low-pass and high-pass filters, while T_l and T_h are their respective Hilbert transforms, as

shown in Figure 3. This extended filtering mechanism results in a total of 16 sub-bands and the introduction of Hilbert transform-based filters enhances the transform by providing shift invariance:

$$\begin{aligned} QWT_{ff} &= \{L_fL_f, L_fH_f, H_fL_f, H_fH_f\}, real, \\ QWT_{tf} &= \{L_tL_f, L_tH_f, H_tL_f, H_tH_f\}, (i), \\ QWT_{ft} &= \{L_fL_t, L_fH_t, H_fL_t, H_fH_t\}, (j), \\ QWT_{tt} &= \{L_tL_t, L_tH_t, H_tL_t, H_tH_t\}, (k), \end{aligned} \tag{6}$$

where f and t represent the real-valued filter F and its Hilbert transform T, respectively. The resulting sub-bands are then projected into quaternion space, as follows:

$$\mathbf{QWT}(I) = QWT_{ff} + QWT_{tf}\mathbf{i} + QWT_{ft}\mathbf{j} + QWT_{tt}\mathbf{k}, \tag{7}$$

where the individual quaternion sub-bands are defined as:

$$\begin{aligned} LL &= L_fL_f + L_tL_f\mathbf{i} + L_fL_t\mathbf{j} + L_tL_t\mathbf{k}, \\ LH &= L_fH_f + L_tH_f\mathbf{i} + L_fH_t\mathbf{j} + L_tH_t\mathbf{k}, \\ HL &= H_fL_f + H_tL_f\mathbf{i} + H_fL_t\mathbf{j} + H_tL_t\mathbf{k}, \\ HH &= H_fH_f + H_tH_f\mathbf{i} + H_fH_t\mathbf{j} + H_tH_t\mathbf{k}, \end{aligned} \tag{8}$$

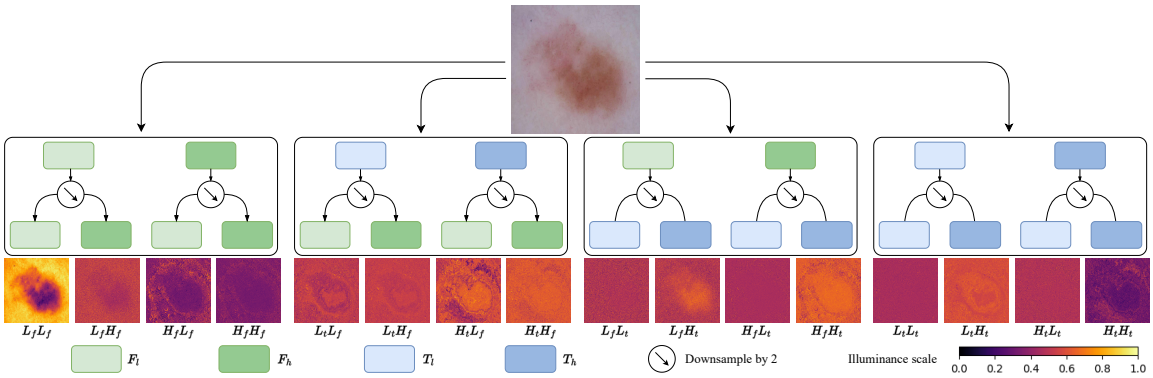

Figure 3. Dual-tree quaternion wavelet transform (QWT). The input image is decomposed into sixteen sub-bands by interleaving low-pass and high-pass filters.

3.3.2. Quaternion Group Normalization (QGN)

Group normalization [31] partitions the input into multiple groups and normalizes within each group, accelerating convergence and improving accuracy. However, conventional group normalization does not ensure equal variance across quaternion components. To address this, we introduce an augmented quaternion vector $\tilde{\mathbf{q}}$. For a quaternion $\mathbf{q} \in \mathbb{H}$, the augmented vector is defined as:

$$\tilde{\mathbf{q}} = [\mathbf{q}, \mathbf{q}^{\hat{\imath}}, \mathbf{q}^{\hat{\jmath}}, \mathbf{q}^{\hat{\kappa}}]^T, \tag{9}$$

where the three perpendicular involutions of $\mathbf{q}$ are defined as:

$$\begin{aligned} \mathbf{q}^{\hat{\imath}} &= q_r^{\hat{\imath}} + q_i^{\hat{\imath}}\mathbf{i} - q_j^{\hat{\imath}}\mathbf{j} - q_k^{\hat{\imath}}\mathbf{k}, \\ \mathbf{q}^{\hat{\jmath}} &= q_r^{\hat{\jmath}} + q_i^{\hat{\jmath}}\mathbf{i} - q_j^{\hat{\jmath}}\mathbf{j} - q_k^{\hat{\jmath}}\mathbf{k}, \\ \mathbf{q}^{\hat{\kappa}} &= q_r^{\hat{\kappa}} + q_i^{\hat{\kappa}}\mathbf{i} - q_j^{\hat{\kappa}}\mathbf{j} - q_k^{\hat{\kappa}}\mathbf{k}. \end{aligned} \tag{10}$$

We define the augmented covariance matrix as:

$$\tilde{\mathcal{C}}_{\mathbf{qq}} = \mathrm{E}\left\{\tilde{\mathbf{q}}\tilde{\mathbf{q}}^{\mathrm{H}}\right\} = \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathcal{C}_{\mathbf{qq}^{\hat{\imath}}} & \mathcal{C}_{\mathbf{qq}^{\hat{\jmath}}} & \mathcal{C}_{\mathbf{qq}^{\hat{\kappa}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{\imath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\imath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\jmath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\kappa}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{\jmath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\imath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\jmath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\kappa}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{\kappa}}} & \mathcal{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\imath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\jmath}}} & \mathcal{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\kappa}}} \end{bmatrix}, \tag{11}$$

where $\mathcal{C}$ represents the covariance between the real, i, j, and k components of the quaternion. To simplify this formulation for practical applications, we assume $\mathbb{Q}$-properness [32], which states that the quaternion vector $\mathbf{q}$ is uncorrelated with its involutions $\mathbf{q}^{\hat{\imath}}$, $\mathbf{q}^{\hat{\jmath}}$, and $\mathbf{q}^{\hat{\kappa}}$, i.e., $\mathcal{C}_{\mathbf{q}\mathbf{q}^{\hat{\imath}}} = \mathcal{C}_{\mathbf{q}\mathbf{q}^{\hat{\jmath}}} = \mathcal{C}_{\mathbf{q}\mathbf{q}^{\hat{\kappa}}} = 0$. Under this assumption, Equation (11) simplifies to:

$$\tilde{\mathcal{C}}_{\mathbf{q}\mathbf{q}} = \begin{bmatrix} \mathcal{C}_{\mathbf{q}\mathbf{q}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{\imath}}\mathbf{q}^{\hat{\imath}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{\jmath}}\mathbf{q}^{\hat{\jmath}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{\kappa}}\mathbf{q}^{\hat{\kappa}}} \end{bmatrix} = \sum_{\delta \in \{r,i,j,k\}} \mathrm{E}\left\{\mathbf{q}_\delta^2\right\} \mathbf{I}. \tag{12}$$

While this approach assumes $\mathbb{Q}$-properness, it demonstrates that the variance of a quaternion can be approximated by the variances of its four components. Accordingly, the quaternion mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$ are calculated as follows:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{C}\sum_{c=1}^{C}(q_{r,c} + q_{i,c}\mathbf{i} + q_{j,c}\mathbf{j} + q_{k,c}\mathbf{k}) \\ &= \overline{q_r} + \overline{q_i}\mathbf{i} + \overline{q_j}\mathbf{j} + \overline{q_k}\mathbf{k}, \\ \boldsymbol{\sigma}^2 &= \frac{1}{4}\sum_{\delta \in \{r,i,j,k\}} \left(\mathrm{E}\{\mathbf{q}_\delta^2\} - (\mathrm{E}\{\mathbf{q}_\delta\})^2\right) \\ &= \frac{1}{4C}\sum_{\delta \in \{r,i,j,k\}}\sum_{c=1}^{C}\left(\mathbf{q}_{\delta,c} - \bar{\mathbf{q}}_\delta\right)^2. \end{aligned} \tag{13}$$

Thus, we can define the complete QGN process as follows:

$$\mathbf{QGN}(x) = \gamma\left(\frac{\mathbf{x} - \boldsymbol{\mu}}{\sqrt{\boldsymbol{\sigma}^2 + \epsilon}}\right) + \beta. \tag{14}$$

Finally, the quaternion wavelet embedding is:

$$\hat{I} = \mathbf{QGN}(\mathbf{QWT}(I)). \tag{15}$$

3.4. Quaternion Depthwise Convolution (QDConv)

QDConv is a core component in our framework, employed in OCQC module, QEFN, and both upsampling and downsampling paths. It preserves the expressive capacity of quaternion convolution while significantly reducing parameters.

Given an quaternion input $x = x^r + x^i\mathbf{i} + x^j\mathbf{j} + x^k\mathbf{k}$ and a quaternion kernel $W = W^r + W^i\mathbf{i} + W^j\mathbf{j} + W^k\mathbf{k}$, the quaternion convolution is implemented via Hamilton product:

$$\begin{aligned} W \otimes x = \ & (W^r x^r - W^i x^i - W^j x^j - W^k x^k) \\ & + (W^r x^i + W^i x^r + W^j x^k - W^k x^j)\mathbf{i} \\ & + (W^r x^j - W^i x^k + W^j x^r + W^k x^i)\mathbf{j} \\ & + (W^r x^k + W^i x^j - W^j x^i + W^k x^r)\mathbf{k} \end{aligned}. \tag{16}$$

Compared to standard convolution with 16 matrices for 4-channel input and output, quaternion convolution requires only 4, achieving a 75% parameter reduction. To further improve efficiency, QDConv partitions the input $\mathbf{y} \in \mathbb{R}^{h\times w\times c}$ into $g = c/4$ groups. Let $W_g = W_g^r + W_g^i\mathbf{i} + W_g^j\mathbf{j} + W_g^k\mathbf{k}$ denote the quaternion convolution kernel for group g:

$$\begin{aligned} \mathbf{y}_g &= \mathbf{y}_g^r + \mathbf{y}_g^i\mathbf{i} + \mathbf{y}_g^j\mathbf{j} + \mathbf{y}_g^k\mathbf{k}, \\ \mathbf{y}_g^{out} &= \mathbf{y}_g \otimes W_g, \quad g = 1, \ldots, G. \end{aligned} \tag{17}$$

The final output of QDConv is obtained by concatenating the outputs of all groups:

$$\mathbf{y}^{\text{out}} = \left[\mathbf{y}_1^{\text{out}}, \mathbf{y}_2^{\text{out}} \ldots, \mathbf{y}_G^{\text{out}}\right]. \tag{18}$$

3.5. Consistent Quaternion Convolution (CQC) block

Our CQC block integrates quaternion layer normalization (QLN), an OCQC module, a QEFN, and residual connections into a cohesive design. The overall structure illustrated in Figure 2 is based on the MetaFormer framework [33].

Overall, given an input $x \in \mathbb{R}^{h\times w\times c}$, the OCQC module follows the process:

$$\begin{aligned} x_{\text{ocqc}} &= \mathbf{OCQC}(\mathbf{QLN_1}(x)) + x, \\ x_{\text{output}} &= \mathbf{QEFN}\big(\mathbf{QLN_2}(x_{(ocqc)})\big) + x_{\text{ocqc}}. \end{aligned} \tag{19}$$

3.5.1. Quaternion Layer Normalization (QLN)

As demonstrated in the MetaFormer architecture, a lightweight normalization method is essential at the input of both the token mixer and the FFN. To meet this requirement, we leverage the intrinsic normalization rules of quaternions to construct QLN. Given an input quaternion $\mathbf{x} \in \mathbb{H}$, the normalization process begins by computing the quaternion norm: Given an input quaternion $\mathbf{x} \in \mathbb{H}$, the normalization process begins by computing the quaternion norm:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}_r^2 + \mathbf{x}_i^2 + \mathbf{x}_j^2 + \mathbf{x}_k^2 + \epsilon} \tag{20}$$

where $\epsilon > 0$ is a small constant added to avoid division by zero. Then, each component of the quaternion needs to be normalized:

$$\mathbf{x}'_r = \frac{\mathbf{x}_r}{\|\mathbf{x}\|}, \quad \mathbf{x}'_i = \frac{\mathbf{x}_i}{\|\mathbf{x}\|}, \quad \mathbf{x}'_j = \frac{\mathbf{x}_j}{\|\mathbf{x}\|}, \quad \mathbf{x}'_k = \frac{\mathbf{x}_k}{\|\mathbf{x}\|} \tag{21}$$

According to Equation (3), we can define the complete quaternion layer normalization (QLN) process as follows:

$$\mathbf{QLN}(\mathbf{x}) = \alpha\Big(\frac{\mathbf{x}}{\|\mathbf{x}\|}\Big) + \beta \tag{22}$$

where $\alpha \in \mathbb{R}^d$ is a learnable scaling parameter, and $\beta \in \mathbb{R}^d$ is a learnable shifting parameter.

3.5.2. Orientation-Consistent Quaternion Convolution (OCQC) Module

The proposed OCQC module is shown in Figure 4, which consists of four branches. From top to bottom, these branches include identity mapping and three QDConvs with various kernels. Each branch is designed to extract features from different sub-bands produced by the quaternion wavelet embedding.

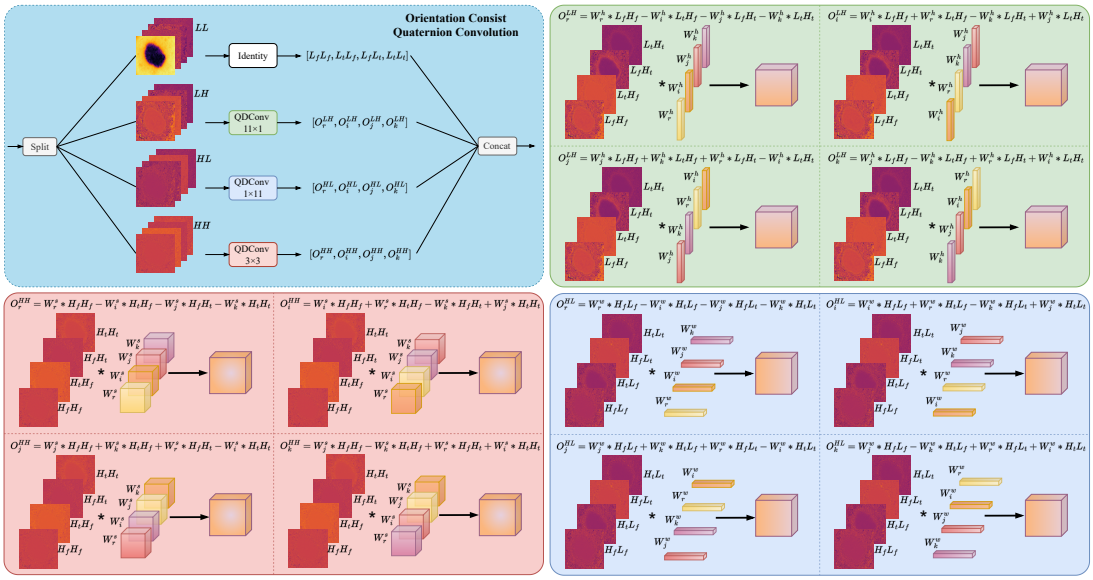

Figure 4. The details of the first OCQC module: It consists of four branches, each operating with quaternion convolution kernels of different forms, and the results are finally concatenated together.

Taking the first OCQC module as an example, the input X is split into four groups along channel dimension corresponding to the LL, LH, HL, and HH sub-bands. The LL sub-band, capturing primary image structure, undergoes identity mapping to preserve geometric integrity without additional parameters. The LH sub-band, representing vertical high-frequency details, is processed with a 11×1 QDConv kernel to extract features while

minimizing redundant information. Similarly, the HL sub-band, containing horizontal high-frequency details, uses a 1×11 QDConv kernel. The HH sub-band, encoding diagonal high-frequency texture, employs a 3×3 QDConv kernel to capture bidirectional features.

Formally, the input X is divided into four groups:

$$X_{id}, X_h, X_w, X_s = X_{:c}, X_{c:2c}, X_{2c:3c}, X_{3c:}, \tag{23}$$

where c is the channel count per group. These groups are processed by different branches:

$$\begin{aligned} X'_{id} &= X_{id},\ X'_h = QDConv^{c\to c}_{k_h\times 1}(X_h), \\ X'_w &= QDConv^{c\to c}_{1\times k_w}(X_w),\ X'_s = QDConv^{c\to c}_{k_s\times k_s}(X_s). \end{aligned} \tag{21}$$

Finally, these features are concatenated along the channel dimension:

$$X' = [X'_{id}, X'_h, X'_w, X'_s]. \tag{10}$$

Within each sub-band, QDConv enables interaction between features from filters F and T. For instance, $O^{LH}_{\delta}, \delta \in \{r, i, j, k\}$ integrates information from $L_f H_f$, $L_f H_t$, $L_t H_f$, and $L_t H_t$, with analogous computations for other sub-bands.

3.5.3. Quaternion Enhanced Feedforward Network (QEFN)

The OCQC module effectively extracts internal information within individual sub-bands but does not account for interactions between different frequency bands. To address this limitation, we propose QEFN, which mainly consists of two QDSCs and one IQDSC, as shown in Figure 5.

Taking the first QEFN in the network as an example, the initial QDSC employs a 3×3 QDConv to extract spatial features and promote interactions within groups of four non-overlapping channels, as illustrated in Figure 6 (a). Subsequently, a 1×1 QPConv doubles the channel count by treating the entire channel as a single group, evenly divided into four parts corresponding to the quaternion components, enabling feature integration across the LL, LH, HL, and HH sub-bands via the Hamilton product. In the high-dimensional space, a second QDSC enhances both local and global channel correlations, with GELU activation applied after each QDSC to introduce nonlinearity and capture complex relationships. Finally, an IQDSC restores the original channel dimensions, followed by QGN to refine the output.

Overall, the process of QEFN can be expressed by the following formula:

$$X'' = \mathbf{QGN}(\text{IQDSC}(\sigma(\text{QDSC}(\sigma(\text{QDSC}(X')))))), \tag{11}$$

where σ represents the GELU activation function.

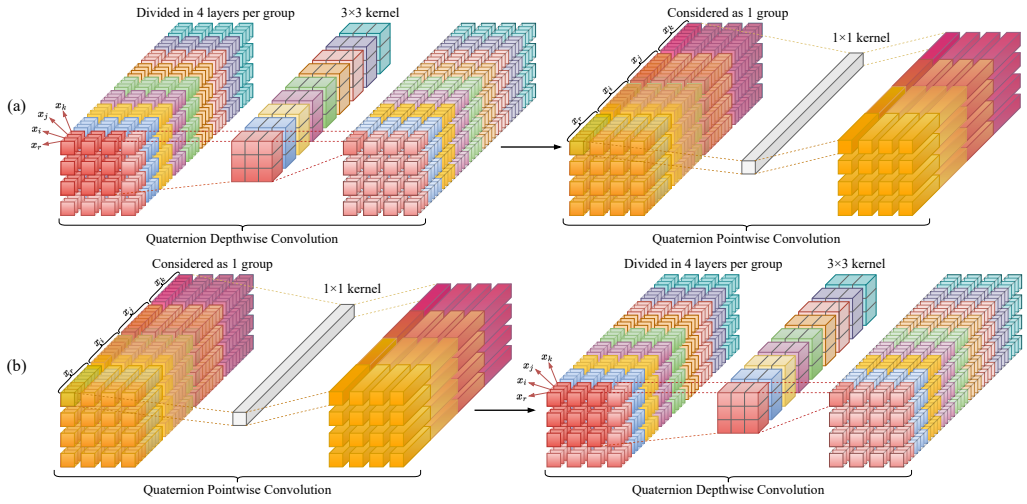

Figure 5. The architectures of QDSC and IQDSC. (**a**) QDSC consists of a QDConv and a QPConv. The QDConv facilitates interactions among every four non-overlapping channels, while the QPConv promotes interactions across all channels; (**b**) in contrast, the components of IQDSC are arranged in reverse.

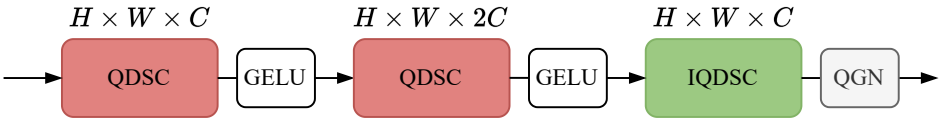

Figure 6. The architecture of the proposed QEFN, which consists of two QDSCs, one IQDSC, GeLU activation functions, and QGN. Notably, the second QDSC projects the channel dimensions of the input features to twice their original size, which are then restored to their initial dimensions by the IQDSC.

3.5.4. Architecture Design

The OCQC module maintains sub-band independence by avoiding global channel interactions, whereas QEFN facilitates their mixing. However, residual connections preserve original sub-band features and directional information across stages. Therefore, we maintain the same design across each CQC block: the OCQC module first extracts direction-specific features, preserving channel independence. As prior QEFN introduces channel interactions, the strip kernels in OCQC module cannot capture features from all directions. Thus, the 3×3 QDConvs in QEFN complements this limitation by capturing additional features.

3.6. Loss Function

The total loss combines Dice loss, Binary Cross-Entropy (BCE) loss, and Focal loss. Dice loss measures overlap between predictions and ground truth, effectively handling class imbalance:

$$L_{\text{Dice}} = 1 - \frac{2\sum_i (p_i g_i)}{\sum_i p_i + \sum_i g_i} \tag{12}$$

where p_i and g_i denote the predicted map and ground truth.

BCE loss serves as a standard pixel-wise classification loss:

$$L_{\text{BCE}} = -\frac{1}{N}\sum_i [g_i \log(p_i) + (1 - g_i)\log(1 - p_i)] \tag{13}$$

Focal loss emphasizes hard-to-classify pixels:

$$L_{\text{Focal}} = -\frac{1}{N}\sum_i \alpha(1 - p_i)^\gamma g_i \log(p_i) \tag{14}$$

To avoid manual tuning, we adopt uncertainty modeling [34], treating λ_i as learnable:

$$L_{\text{total}} = \frac{1}{\lambda_1^2} L_{\text{Dice}} + \frac{1}{\lambda_2^2} L_{\text{BCE}} + \frac{1}{\lambda_3^2} L_{\text{Focal}} + \log(\lambda_1 \lambda_2 \lambda_3) \tag{15}$$

where the logarithmic term regularizes λ_i. The parameters are initialized as $\lambda_1 = \lambda_2 = \lambda_3 = 1$ and updated via backpropagation.

4. Experiments

4.1. Datasets

We conducted experiments on three widely used skin lesion datasets: ISIC 2016 [35], ISIC 2017 [36], and ISIC 2018 [37], curated by the International Skin Imaging Collaboration (ISIC) to support automated melanoma diagnosis. The ISIC 2016 dataset contains 1,250 images, with 900 used for training and 350 for validation. ISIC 2017 and 2018 include 2,150 and 2,694 images, respectively. Following prior work [38,39], the ISIC 2017 and ISIC 2018 datasets were divided into training and validation sets in a 7:3 ratio.

4.2. Implementation Details

Our model was trained on a server running Python 3.8, CUDA 12.4, and the PyTorch 2.5 framework, equipped with an RTX 4070 Super GPU (12GB). We utilized the AdamW optimizer with an initial learning rate of 1×10^{-3}, combined with the CosineAnnealingLR [40] scheduler, which decays the learning rate from a maximum of 1×10^{-3} to a minimum of 1×10^{-5} over 50 iterations. Before inputting images into the model, we resized them to 256×256 pixels and applied data augmentation techniques, including vertical and horizontal flips with a probability of 0.5 and

random rotations between 0° and 360° with a probability of 0.5. Each dataset was trained for 300 epochs.

The performance of lesion segmentation is generally measured by comparing predicted masks with ground truth annotations. We adopted five commonly used metrics: intersection over union (IoU), dice similarity coefficient (DSC), accuracy (Acc), specificity (Spe), and sensitivity (Sen).

4.3. Comparison With Other Segmentation Methods

We conducted comparisons with a wide range of commonly used methods, including CNN-based approaches [3, 10, 41–43], Transformer-based methods [7, 9], Mamba-based approaches [38, 39], and hybrid methods combining CNN and attention mechanisms [5, 8].

4.3.1. ISIC 2016

The results are summarized in Table 1. Most methods perform well on this dataset with minor differences, but our method surpasses UNetV2 in mIoU, DSC, Acc, and sensitivity, achieving state-of-the-art performance.

Visual comparisons are shown in Figure 7. In the first image, UNet++ and U-Lite over-segment the right side, while Swin-UNet under-segments the red-circled regions and produces jagged edges. Most methods mispredict the left concave region, and VM-UNetV2 struggles with lighter lesions. Our method restores the concavity and avoids over-segmentation. In the second image, several methods misclassify parts of the black background or underpredict the bottom-left concavity, while our method captures both accurately. In the third image, most methods produce holes; among the exceptions, Att-UNet, UNext, and C^{2}SGD have contour deviations, TransFuse misses the left protrusion, and VM-UNetV2 predicts isolated points while our method closely matches the ground truth.

This robustness comes from QEFN's integration of coarse contours and fine details, improving boundary perception and maintaining smooth lesion interiors without holes.

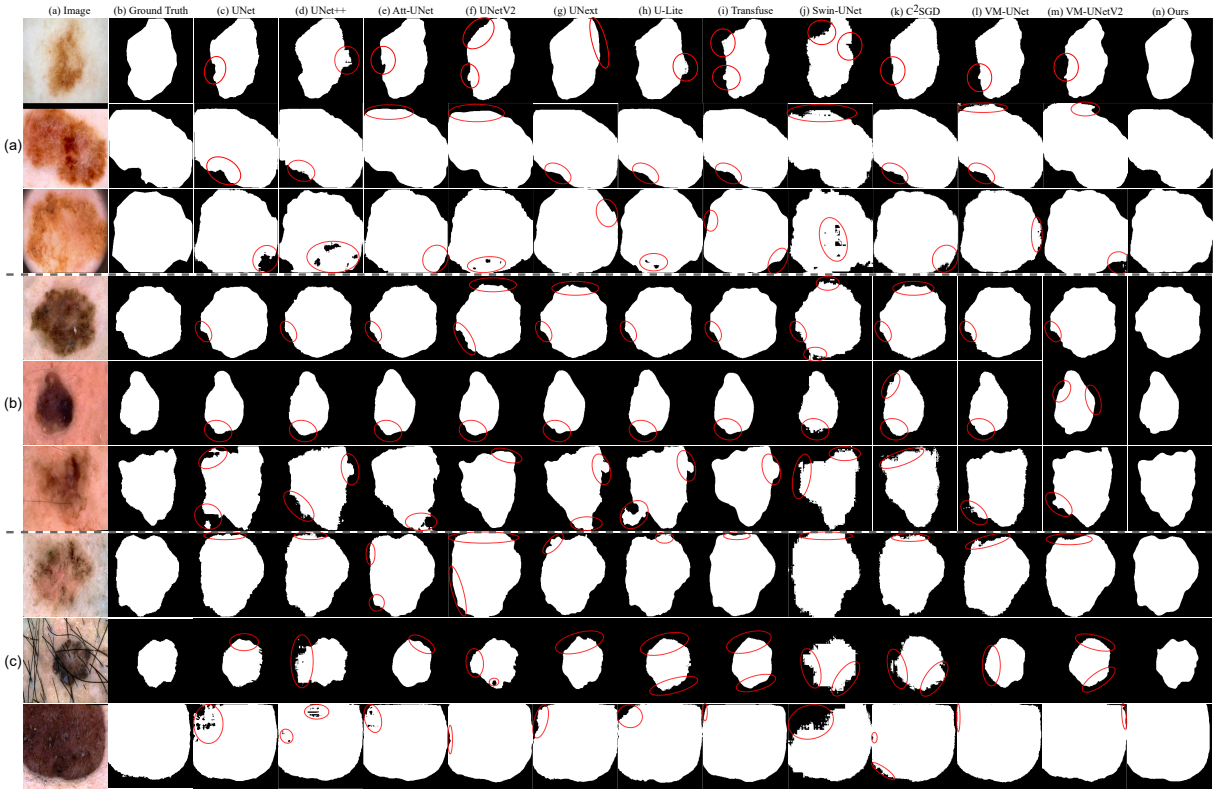

Figure 7. Visual comparison of segmentation results for the ISIC datasets. Panels (**a**); (**b**); and (**c**) depict the outcomes for ISIC 2016, ISIC 2017, and ISIC 2018, respectively.

4.3.2. ISIC 2017

As depicted in Table 2, our method outperforms VM-UNetV2, achieving improvements of 0.50% in mIoU, 0.31% in DSC, 0.22% in accuracy (Acc), and 0.34% in specificity (Spe), thereby establishing state-of-the-art performance.

Table 1. Quatitative Comparison on ISIC 2016 dataset (bold indicates the best, underline indicates the second best, * indicates the use of pre-trained weights.).

Model	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
UNet [3]	84.33	91.50	95.22	96.82	91.16
UNet++ [41]	83.58	91.05	94.90	96.04	92.01
Att-UNet [8]	84.55	91.63	95.27	96.64	91.77
UNetV2 * [5]	<u>86.07</u>	<u>92.52</u>	<u>95.79</u>	97.17	92.27
UNext [42]	85.25	92.04	95.49	96.67	92.49
U-Lite [43]	85.08	91.94	95.42	96.49	<u>92.69</u>
Transfuse-L * [6]	84.66	91.69	95.36	97.17	90.76
C^2SGD* [10]	85.28	92.05	95.52	96.92	91.96
Swin-UNet * [7]	83.61	91.08	95.20	96.60	91.67
VM-UNet * [38]	85.35	92.10	95.48	96.32	**93.35**
VM-UNetV2 * [39]	85.55	92.21	95.64	**97.31**	91.41
CQC-Net (Ours)	**86.15**	**92.56**	**95.80**	<u>97.27</u>	92.09

Table 2. Quatitative Comparison on ISIC 2017 dataset (bold indicates the best, underline indicates the second best, * indicates the use of pre-trained weights.).

Model	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
UNet [3]	75.57	86.09	95.48	97.87	83.58
UNet++ [41]	78.18	87.76	95.97	97.91	86.32
Att-UNet [8]	76.34	86.58	95.56	97.56	85.60
UNetV2 * [5]	78.04	87.66	95.92	97.81	86.54
UNext [42]	77.61	87.39	95.11	<u>98.17</u>	84.66
U-Lite [43]	78.13	87.72	95.97	97.96	86.07
Transfuse-L * [6]	79.91	88.83	96.30	97.98	87.92
C^2SGD* [10]	79.46	88.55	96.10	97.30	**90.13**
Swin-UNet* [7]	73.83	84.95	94.99	97.12	84.42
VM-UNet * [38]	80.19	89.01	96.34	97.71	<u>88.54</u>
VM-UNetV2 * [39]	<u>80.28</u>	<u>89.06</u>	<u>96.36</u>	97.96	88.42
CQC-Net (Ours)	**80.78**	**89.37**	**96.58**	**98.30**	87.79

In the first example, most methods fail to accurately segment the lower side of the protruding region on the left, while UNetV2, UNext, and Swin-Unet suffer from over-segmentation at the top of the lesion. In the second image, both our method and VM-UNetV2 successfully recover the lower protruding region; however, VM-UNetV2 misclassifies the lighter area on the right as part of the lesion, causing over-segmentation. In the third image, UNet, UNet++, Att-UNet, UNext, and U-Lite produce significantly oversized masks with irregular edges. Although UNetV2, TransFuse, and C^2SGD generate masks closer in area and contour to the ground truth, they still fail to capture fine edge details.

These examples demonstrate cases where lesion edges exhibit low color contrast with surrounding skin. Our method effectively addresses this challenge by enhancing high-frequency edge information in horizontal, vertical, and diagonal directions, resulting in superior perception and segmentation of subtle boundaries.

4.3.3. ISIC 2018

Our method shows a clear advantage over competing approaches. Compared with VM-UNet, it improves mIoU by 1.04%, DSC by 0.64%, Acc by 0.28%, and sensitivity by 0.26% over the VM-UNet, as shown in Table 3.

In the first example, the top region of the lesion has similar color to the surrounding skin, making it difficult for most methods to segment accurately. While Att-UNet partially succeeds, it misses the protruding regions marked by the two red circles. In contrast, our method produces contours closely matching the ground truth even in visually ambiguous areas. In the second example, significant hair interference causes over-segmentation in UNet++, UNetV2, Swin-UNet, and C^2SGD. UNet, TransFuse, and VM-UNetV2 fail to detect the concave lesion area due to hair occlusion. Our method effectively suppresses hair features and accurately segments the lesion, as shown in the feature maps in Figure 8. During encoding, the model progressively isolates lesion features; by the third decoder

layer, hair-related features are largely eliminated. In the third example involving a large lesion, most methods produce masks with holes, similar to the ISIC 2016 case. VM-UNetV2 exhibits over-segmentation in the red-circled region, whereas our method generates a mask nearly identical to the ground truth.

Table 3. Quatitative Comparison on ISIC 2018 dataset (bold indicates the best, underline indicates the second best, * indicates the use of pre-trained weights.).

Model	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
UNet [3]	78.01	87.64	94.03	96.26	87.07
UNet++ [41]	76.31	86.56	93.76	<u>97.35</u>	82.90
Att-UNet [8]	77.76	87.49	93.96	96.28	86.74
UNetV2 * [5]	77.90	87.58	93.89	95.64	88.45
UNext [42]	79.02	88.28	94.34	96.52	87.56
U-Lite [43]	79.55	88.61	94.51	96.69	87.74
Transfuse-L * [6]	79.82	88.78	94.70	**97.44**	86.18
C^2SGD * [10]	80.75	89.35	94.78	96.31	<u>90.02</u>
Swin-UNet * [7]	76.33	86.58	93.44	95.53	86.95
VM-UNet * [38]	<u>81.05</u>	<u>89.53</u>	<u>94.93</u>	96.78	89.17
VM-UNetV2 * [39]	80.74	89.35	94.84	96.73	88.96
CQC-Net (Ours)	**82.09**	**90.17**	**95.21**	96.79	**90.28**

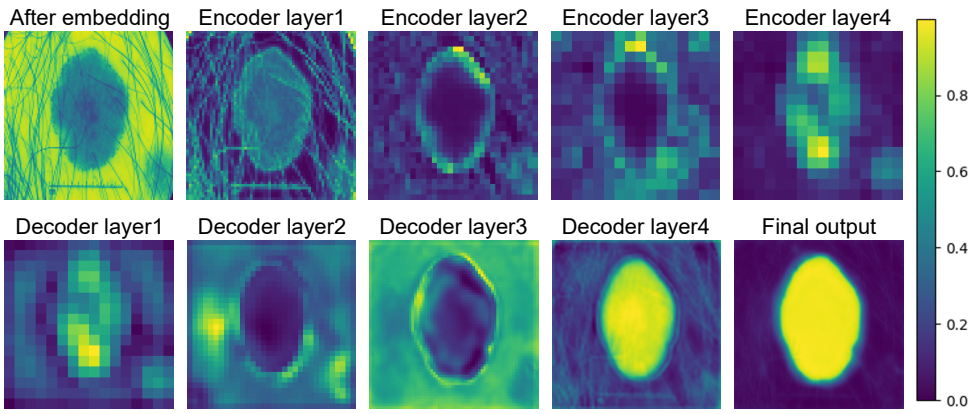

Figure 8. Example feature maps extracted at various stages of our model. The visualizations illustrate the progressive evolution of features, capturing hierarchical information from low-level details in the early stages to high-level semantic representations as they propagate through the encoder and decoder.

Furthermore, Table 4 shows that our method has the lowest number of parameters and requires tens to hundreds of times less memory than other approaches, making it highly suitable for real-world medical deployment.

Table 4. FLOPs and Params Comparation of each method (bold indicates the best).

Model	UNet	UNet++	Att-UNet	UNetV2	UNext	U-Lite	Transfuse-L	C^2SGD	Swin-UNet	VM-UNet	VM-UNetV2	Proposed
Input size	256×256	256×256	256×256	256×256	256×256	256×256	256×192	256×256	224×224	256×256	256×256	256×256
FLOPs(G)↓	54.73	199.64	66.63	3.90	**0.57**	0.76	60.52	7.97	5.91	4.73	4.40	2.96
Params(M)↓	31.03	47.18	34.88	24.90	1.47	0.88	143.54	22.00	27.15	27.50	22.77	**0.86**
Memory(MB)↓	118.49	180.03	133.17	95.13	5.65	3.40	548.03	84.11	105.50	104.75	87.01	**3.40**

4.4. Ablation Studies

We conducted several ablation experiments on the ISIC 2017 and ISIC 2018 datasets to validate our design.

4.4.1. Evaluation of the OCQC Module Design

To verify the effectiveness of our design on OCQC module, we compared the original implementation with two alternative setups: (a) Rearranging the convolution branches such that the order from top to bottom becomes the identity mapping branch, 3×3 square convolution, 1×11 horizontal strip convolution, and 11×1 vertical strip convolution. This setup tests our theory of directional consistency. (b) Replacing both strip convolutions with 11×11 square convolutions to assess the hypothesis that introducing excess directional feature leads to redundancy.

The results of our experiments are presented in Table 5. In setup (a), the mIoU, DSC, and Acc decreased by 1.95%, 1.21%, and 0.38%, respectively, compared to the baseline. This performance drop is due to the loss of directional consistency among the convolution branches. Specifically, in the second branch, the introduction of two directions with a limited vertical receptive field made it difficult to capture the vertical high-frequency information of the LH sub-band; In the third branch, the vertical strip convolution kernel failed to capture the horizontal correlations in the HL sub-band; The last branch only captured horizontal high-frequency information from the HH sub-band while neglecting vertical information.

Table 5. Quatitive evaluation on various setting (bold indicates the best).

Datasets	Settings	Params(K)↓	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
ISIC 2017	(a)	865.78	78.75	88.11	96.10	98.06	86.33
	(b)	944.98	78.76	88.12	96.07	97.88	87.05
	(c)	**422.26**	79.41	88.52	96.26	98.29	86.14
	(d)	3300.87	78.78	88.13	95.99	97.44	88.80
	vanilla	865.78	**80.78**	**89.37**	**96.58**	**98.30**	**87.79**
ISIC 2018	(a)	865.78	80.84	89.41	94.86	96.87	88.65
	(b)	944.98	80.56	89.23	94.80	96.97	88.10
	(c)	**422.26**	80.74	89.34	94.95	**97.49**	87.03
	(d)	3300.87	79.68	88.69	94.39	95.67	90.38
	vanilla	865.78	**82.09**	**90.17**	**95.21**	96.79	**90.28**

Setup (b) resulted in a further decline in performance compared to (a), while also increasing computational complexity and parameter count. In this configuration, the second and third branches introduced excessive direction-independent redundant information, which interfered with the network's learning and caused a significant performance drop.

The comparison between (a), (b), and the original design clearly demonstrates the effectiveness of our OCQC module and the rationale behind its specific design choices.

4.4.2. Evaluation of the QEFN

We compared our QEFN with FFN constructed using two layers of QPConv as linear layers. The results are summarized in setting (c) of Table 5.

Using only linear layers in the FFN resulted in a noticeable performance drop. This decline can be attributed to the limitations of QPConv, which has a 1×1 receptive field and is unable to effectively model the spatial dependencies that the OCQC module fails to fully capture. Moreover, the conventional FFN lacks the capacity to learn rich semantic information in high-dimensional spaces, further impacting overall performance.

4.4.3. Evaluation of the Quaternion Neural Networks

To demonstrate the effectiveness of quaternion networks, we constructed a network using conventional CNNs following the same overall structure as our quaternion-based network. The performance comparison is illustrated in Table 5 setting (d).

The results clearly show that using conventional CNNs significantly increases the number of model parameters. This is because quaternion networks intrinsically encode multi-dimensional information within a single quaternion-valued operation, effectively reducing redundancy and compressing the model, as analyzed in Equation (16). In contrast, conventional CNNs require independent convolutions for each channel, leading to a higher computational burden.

Furthermore, conventional CNNs lack the inherent ability to facilitate interactions across different frequency bands, which is a key feature of quaternion operations. This inability limits the model's capacity to capture the complex interdependencies among frequency components, resulting in suboptimal feature representation. Consequently, the performance of conventional CNNs is inferior, highlighting the advantages of quaternion networks in terms of both efficiency and accuracy.

4.4.4. Evaluation of Various Loss Functions

To identify the most suitable loss function for our method, we compared four configurations: the conventional BCE + Dice, BCE + Dice + Focal, Uncertainty BCE + Dice, and Uncertainty BCE + Dice + Focal. The results are

shown in Table 6. It is evident that incorporating Focal loss without adaptive weighting significantly degrades performance. This is because Focal loss assigns higher weights to hard-to-segment samples, which can disproportionately influence the gradients, causing Dice loss and BCE to be overlooked. As a result, the model struggles to optimize global regions effectively and becomes overly focused on difficult samples, neglecting overall segmentation quality.

Table 6. Quatitive evaluation of variouss loss functions (bold indicates the best).

Datasets	Loss function	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
ISIC 2017	Dice + BCE	80.32	89.09	96.33	97.73	89.39
	Dice + BCE + Focal	79.01	88.28	96.06	97.55	88.63
	Uncertainty Dice + BCE	80.48	89.19	96.42	98.06	88.25
	Uncertainty Dice + BCE + Focal	**80.78**	**89.37**	**96.58**	**98.30**	**87.79**
ISIC 2018	Dice + BCE	81.48	89.80	95.06	96.95	89.21
	Dice + BCE + Focal	81.10	89.56	94.81	95.91	**91.41**
	Uncertainty Dice + BCE	80.86	89.42	94.91	**97.3**	88.34
	Uncertainty Dice + BCE + Focal	**82.09**	**90.17**	**95.21**	96.79	90.28

By introducing uncertainty modeling, the model adaptively learns the weights of each loss term, dynamically balancing their contributions during training. This allows the model to achieve an optimal trade-off and results in superior performance.

5. Discussion

Figure 9 compares the frequency spectra of two input images and their corresponding outputs after final IQWT reconstruction. The original images show smooth color transitions with a broad frequency spectrum. After feature extraction, the reduction of gray areas and expansion of dark regions indicate more effective decoupling of high- and low-frequency information. This property is particularly advantageous for handling diverse and complex lesion boundaries commonly encountered in clinical practice. Moreover, CQC-Net has a compact memory footprint, enabling low-cost deployment in resource-constrained environments.

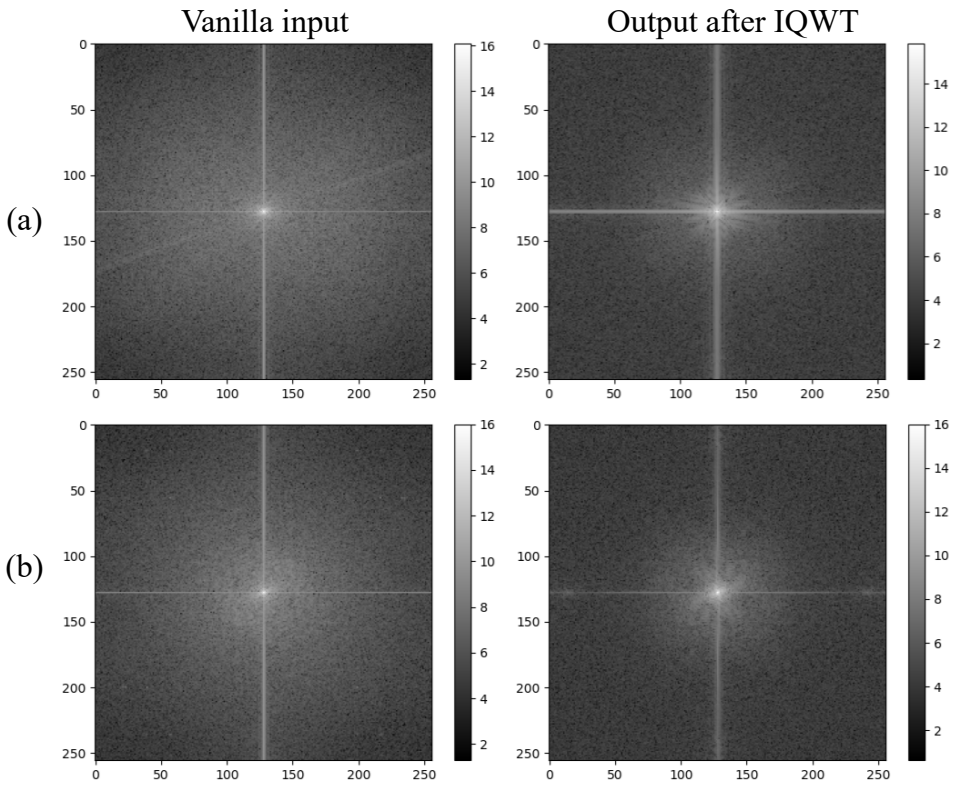

Figure 9. Spectral visualization of two input images and their corresponding outputs after the final IQWT in the network. After feature extraction through the network, the high- and low-frequency components have been separated, with the boundaries becoming more distinct.

Despite its strong performance, CQC-Net still faces challenges when trained on limited datasets, which affects its generalization to rare lesion cases. Future work will incorporate few-shot learning techniques to improve adaptability with minimal supervision. We also plan to explore self-supervised pretraining and semi-supervised learning frameworks to further enhance robustness in real-world clinical settings where labeled data is scarce. These improvements will make CQC-Net a more practical and effective tool for clinical analysis.

6. Conclutions

In this paper, we proposed a light weight model called consistent quaternion convolution neural network (CQC-Net) for skin lesion segmentation, leveraging quaternion wavelet transform and quaternion convolutional operations. By integrating orientation-consistent quaternion convolution (OCQC) module for directional feature extraction and quaternion enhanced feedforward network (QEFN) for inter-frequency interaction, our method effectively combines low-frequency global contour information with high-frequency edge details. This design ensures robust contour perception and accurate segmentation, even in challenging cases with hair interference, blurred boundaries, or light-colored lesions. Extensive experiments on the ISIC 2016, ISIC 2017, and ISIC 2018 datasets demonstrate that our method achieves state-of-the-art performance while maintaining a lightweight architecture with minimal computational complexity.

Author Contributions

H.T.: methodology, writing—original draft; G.H.: writing—review and editing, supervision, funding acquisition; Q.T.: writing—review and editing; X.Y.: writing—review and editing; G.Z.: writing—review and editing; B.L.: supervision. All authors have reviewed and approved the final version of the manuscript for publication.

Funding

This work was supported by Key Areas Research and Development Program of Guangzhou Grant 2023B01J0029, Science and technology research in key areas in Foshan under Grant 2020001006832, the Science and technology projects of Guangzhou under Grant 202007040006, the Guangdong Provincial Key Laboratory of Cyber-Physical System under Grant 2020B1212060069.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the reported results can be found at the ISIC archive: https://challenge.isic-archive.com/data/.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT to assist in drafting and refining sections of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Long, G.V.; Swetter, S.M.; Menzies, A.M.; et al. Cutaneous Melanoma. *Lancet* **2023**, *402*, 485–502.
2. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651.
3. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Berlin/Heidelberg, 2015; pp. 234–241.
4. Chen, L.C.; Papandreou, G.; Kokkinos, I.; et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848.

5. Peng, Y.; Sonka, M.; Chen, D.Z. U-Net V2: Rethinking the Skip Connections of U-Net for Medical Image Segmentation. *arXiv* **2023**, arXiv:2311.17791.
6. Chen, J.; Mei, J.; Li, X.; et al. TransUNet: Rethinking the U-Net Architecture Design for Medical Image Segmentation through the Lens of Transformers. *Med. Image Anal.* **2024**, *97*, 103280.
7. Cao, H.; Wang, Y.; Chen, J.; et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In *Computer Vision—ECCV 2022 Workshops*; Springer: Berlin/Heidelberg, 2023; pp. 205–218.
8. Oktay, O.; Schlemper, J.; Folgoc, L.L.; et al. Attention U-Net: Learning Where to Look for the Pancreas. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands, 4–6 July 2018.
9. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, Proceedings of the 24th International Conference, Strasbourg, France, 27 September–1 October 2021*; Springer International Publishing: Cham, Switzerland, 2021; pp. 14–24.
10. Hu, S.; Liao, Z.; Xia, Y. Devil Is in Channels: Contrastive Single Domain Generalization for Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2023, Proceedings of the 26th International Conference, Vancouver, BC, Canada, 8–12 October 2023*; Springer International Publishing: Cham, Switzerland, 2023; pp. 14–23.
11. Chen, W.; Wang, K.; Qian, C.; et al. PPFormer: A Novel Model for Polyp Segmentation in Digestive Endoscopy. *IEEE Trans. Med. Robot. Bionics* **2024**, *6*, 548–555.
12. Yang, L.; Zhai, C.; Wang, H.; et al. A Dual-Branch Fusion Network for Surgical Instrument Segmentation. *IEEE Trans. Med. Robot. Bionics* **2024**, *6*, 1542–1554.
13. Banu, A.S.; Deivalakshmi, S. AWUNet: Leaf Area Segmentation Based on Attention Gate and Wavelet Pooling Mechanism. *Signal Image Video Process.* **2023**, *17*, 1915–1924.
14. Zhao, Y.; Wang, S.; Zhang, Y.; et al. WRANet: Wavelet Integrated Residual Attention U-Net Network for Medical Image Segmentation. *Complex Intell. Syst.* **2023**, *9*, 6971–6983.
15. Agnes, S.A.; Solomon, A.A.; Karthick, K. Wavelet U-Net++ for Accurate Lung Nodule Segmentation in CT Scans: Improving Early Detection and Diagnosis of Lung Cancer. *Biomed. Signal Process. Control* **2024**, *87*, 105509.
16. Imtiaz, T.; Fattah, S.A.; Kung, S.Y. BAWGNet: Boundary Aware Wavelet Guided Network for the Nuclei Segmentation in Histopathology Images. *Comput. Biol. Med.* **2023**, *165*, 107378.
17. Zhang, J.; Zeng, Z.; Sharma, P.K.; et al. A Dual Encoder Crack Segmentation Network with Haar Wavelet-Based High–Low Frequency Attention. *Expert Syst. Appl.* **2024**, *256*, 124950.
18. Zhou, Y.; Huang, J.; Wang, C.; et al. XNet: Wavelet-Based Low and High Frequency Fusion Networks for Fully- and Semi-Supervised Semantic Segmentation of Biomedical Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 21085–21096.
19. Hamilton, W.R. II. On Quaternions; or on a New System of Imaginaries in Algebra. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1844**, *25*, 10–13.
20. Chan, W.L.; Choi, H.; Baraniuk, R. Quaternion Wavelets for Image Analysis and Processing. In Proceedings of the 2004 International Conference on Image Processing (ICIP), Singapore, 24–27 October 2004; Volume 5, pp. 3057–3060.
21. Lai, Z.; Qu, X.; Liu, Y.; et al. Image Reconstruction of Compressed Sensing MRI Using Graph-Based Redundant Wavelet Transform. *Med. Image Anal.* **2016**, *27*, 93–104.
22. Choi, Y.J.; Lee, Y.W.; Kim, B.G. Wavelet Attention Embedding Networks for Video Super-Resolution. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 7314–7320.
23. Lin, M.; Lan, Q.; Huang, C.; et al. Wavelet-Based U-Shape Network for Bioabsorbable Vascular Stents Segmentation in IVOCT Images. *Front. Physiol.* **2024**, *15*, 1454835.
24. Zheng, Z.; Huang, G.; Yuan, X.; et al. Quaternion-Valued Correlation Learning for Few-Shot Semantic Segmentation. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 2102–2115.
25. Celsi, M.R.; Scardapane, S.; Comminiello, D. Quaternion Neural Networks for 3D Sound Source Localization in Reverberant Environments. In Proceedings of the 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), Espoo, Finland, 21–24 September 2020; pp. 1–6.
26. Zhou, Z.; Huo, Y.; Huang, G.; et al. QEAN: Quaternion-Enhanced Attention Network for Visual Dance Generation. *Vis. Comput.* **2024**, *41*, 961–973.
27. Gaudet, C.J.; Maida, A.S. Deep Quaternion Networks. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
28. Tay, Y.; Zhang, A.; Luu, A.T.; et al. Lightweight and Efficient Neural Natural Language Processing with Quaternion Networks. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 Augus 2019; pp. 1494–1503.
29. Van Le, T.; Lee, J.Y. Specular Highlight Removal Using Quaternion Transformer. *Comput. Vis. Image Underst.* **2024**, *249*, 104179.
30. Frants, V.; Agaian, S.; Panetta, K. QCNN-H: Single-Image Dehazing Using Quaternion Neural Networks. *IEEE Trans. Cybern.* **2023**, *53*, 5448–5458.

31. Wu, Y.; He, K. Group Normalization. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
32. Cheong Took, C.; Mandic, D.P. Augmented Second-Order Statistics of Quaternion Random Signals. *Signal Process.* **2011**, *91*, 214–224.
33. Yu, W.; Luo, M.; Zhou, P.; et al. MetaFormer Is Actually What You Need for Vision. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 10809–10819.
34. Cipolla, R.; Gal, Y.; Kendall, A. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7482–7491.
35. Gutman, D.; Codella, N.C.; Celebi, E.; et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2016**, arXiv:1605.01397.
36. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 168–172.
37. Codella, N.C.F.; Rotemberg, V.M.; Tschandl, P.; et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2019**, arXiv:1902.03368.
38. Ruan, J.; Xiang, S. VM-UNet: Vision Mamba UNet for Medical Image Segmentation. *arXiv* **2024**, arXiv:2402.02491.
39. Zhang, M.; Yu, Y.; Jin, S.; et al. VM-UNET-V2: Rethinking Vision Mamba UNet for Medical Image Segmentation. In *Bioinformatics Research and Applications, Proceedings of the 20th International Symposium, ISBRA 2024, Kunming, China, 19–21 July 2024*; Springer International Publishing: Cham, Switzerland, 2024; pp. 335–346.
40. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
41. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 1856–1867.
42. Valanarasu, J.M.J.; Patel, V.M. UNeXt: MLP-Based Rapid Medical Image Segmentation Network. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022, Proceedings of the 25th International Conference, Singapore, 18–22 September 2022*; Springer International Publishing: Cham, Switzerland, 2024; pp. 335–346. pp. 23–33.
43. Dinh, B.D.; Nguyen, T.T.; Tran, T.T.; Pham, V.T. 1M Parameters Are Enough? A Lightweight CNN-Based Model for Medical Image Segmentation. In Proceedings of the 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, Taiwan, 31 October–3 November 2023; pp. 1279–1284.