1 Related works

1.1 Quaternion nerual networks

Quaternion neural networks (QNNs) have been applied to tasks such as segmentation [15], 3D audio processing [3], and human pose estimation [16]. Gaudet et al. [6] introduced quaternion convolutional neural networks (QCNNs), showing they outperform real-valued CNNs with fewer parameters. Tay et al. [10] proposed quaternion attention, extending real-valued attention to the quaternion domain with improved efficiency and performance. This approach has been successfully applied in highlight removal [11] and single-image dehazing [5]. Motivated by these advances, we adopt QNNs in this work. Existing methods often map RGB channels to the imaginary parts of quaternions with a zero real part, which adds unnecessary computation and noise. To address this, we extend QNNs to the wavelet domain, leveraging only the intrinsic information within the image.

1.2 Wavelet-based segmentation

Wavelet transform is widely used in time-frequency analysis and computer vision due to its reversibility and information preservation [8]. It has recently gained attention in image segmentation. Banu et al. [2] introduced discrete wavelet transform for downsampling in U-shaped networks, arguing it preserves image details better than traditional pooling, achieving near-lossless results. Zhao et al. [14] suggested retaining only the LL component during downsampling, as it carries the most essential information, while high-frequency components may introduce noise. Agnes et al. [1] extended this by using inverse wavelet transform for upsampling, enabling nearly lossless end-to-end processing. Imtiaz et al. [7] integrated high-frequency features into a boundary-awareness module using a gated unit to enhance edge detail extraction. Zhang et al. [13] proposed a high-low frequency attention mechanism, processing LL and high-frequency components separately before concatenation to capture both types of features. Lin et al. [9] combined a wavelet and convolutional encoder to better segment vascular stents in CT images. However, most methods handle high- and low-frequency information separately or discard components, overlooking sub-band properties and their interactions. This work explores leveraging these characteristics to further enhance segmentation performance.

2 Basic quaternion algebra

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

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

Here, r, x, y, and z are real numbers, while $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ are the quaternion unit bases. In this representation, r is the real part, and $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, with $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ik} = -1$, represents the imaginary part. 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:

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

Scalar Multiplication: The Multiplication with scalar β is defined as:

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

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 is the fundamental multiplication rule in quaternion algebra. The multiplication of two quaternions Q and $R = p + l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$ is defined as:

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

In contrast to the traditional real-valued dot product, which only aggregates information from the same position, each component of the Hamilton product incorporates information from other components.

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.

3.1 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.

Let $I(x,y)$ denote the input image of height m and width n, and $F_l[n]$ and $F_h[n]$ represent the low-pass and high-pass filters, respectively. The decomposition begins with row-wise filtering, where the image is convolved with $h[n]$ and $g[n]$ to produce low-frequency and high-frequency components:

$$\begin{aligned} I_{\text{low}}(x,y) &= \sum_n I(x,n) * F_l[n], \\ I_{\text{high}}(x,y) &= \sum_n I(x,n) * F_h[n] \end{aligned} \tag{6}$$

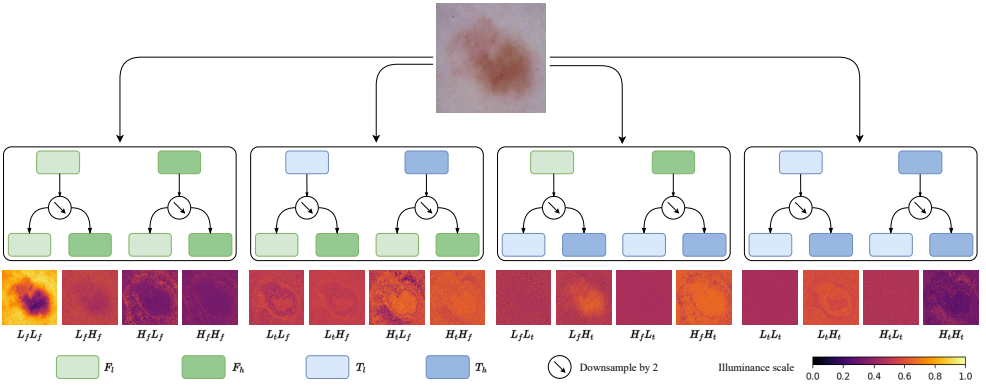

Figure 1: Dual-tree quaternion wavelet transform (QWT). The input image is decomposed into sixteen sub-bands by interleaving low-pass and high-pass filters.

These intermediate outputs are then subjected to columsecse filtering, yielding four distinct sub-bands:

$$\begin{aligned} LL &= \sum_m I_{\text{low}}(m,y) * F_l[m], \\ LH &= \sum_m I_{\text{low}}(m,y) * F_h[m], \\ HL &= \sum_m I_{\text{high}}(m,y) * F_l[m], \\ HH &= \sum_m I_{\text{high}}(m,y) * F_h[m] \end{aligned} \tag{7}$$

Here, LL represents low-frequency components in both dimensions, while LH, HL, and HH capture horizontal, vertical, and diagonal high-frequency details, respectively. Each sub-band is then downsampled by a factor of 2.

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 1. 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:

$$\begin{aligned} QWT_{ff} &= \{L_fL_f, L_fH_f, H_fL_f, H_fH_f\}, real, \\ QWT_{tf} &= \{L_tL_f, L_tH_f, H_tL_f, H_tH_f\}, (i), \\ QWT_{ft} &= \{L_fL_t, L_fH_t, H_fL_t, H_fH_t\}, (j), \\ QWT_{tt} &= \{L_tL_t, L_tH_t, H_tL_t, H_tH_t\}, (k) \end{aligned} \tag{8}$$

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:

$$\begin{aligned} \mathbf{QWT}(I) = QWT_{ff} + QWT_{tf}\mathbf{i} \\ + QWT_{ft}\mathbf{j} + QWT_{tt}\mathbf{k} \end{aligned} \tag{9}$$

where the individual quaternion sub-bands are defined as:

$$\begin{aligned} LL &= L_f L_f + L_t L_f \mathbf{i} + L_f L_t \mathbf{j} + L_t L_t \mathbf{k}, \\ LH &= L_f H_f + L_t H_f \mathbf{i} + L_f H_t \mathbf{j} + L_t H_t \mathbf{k}, \\ HL &= H_f L_f + H_t L_f \mathbf{i} + H_f L_t \mathbf{j} + H_t L_t \mathbf{k}, \\ HH &= H_f H_f + H_t H_f \mathbf{i} + H_f H_t \mathbf{j} + H_t H_t \mathbf{k} \end{aligned} \tag{10}$$

3.2 Quaternion Group Normalization (QGN)

Group normalization, proposed by Wu et al. [12], partitions the input data into multiple groups and computes the mean and variance within each group for normalization. This method accelerates network convergence and improves accuracy. However, conventional group normalization does not guarantee equal variance across the quaternion components due to independent variance calculations within each group. To address this issue, we construct an augmented quaternion vector $\tilde{\mathbf{q}}$.

Given a quaternion $\mathbf{q} \in \mathbb{H}$, the augmented quaternion vector is defined as:

$$\tilde{\mathbf{q}} = [\mathbf{q}, \mathbf{q}^{\hat{i}}, \mathbf{q}^{\hat{j}}, \mathbf{q}^{\hat{k}}]^T \tag{11}$$

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

$$\begin{aligned} \mathbf{q}^{\hat{i}} &= q_r^{\hat{i}} + q_i^{\hat{i}}\mathbf{i} - q_j^{\hat{i}}\mathbf{j} - q_k^{\hat{i}}\mathbf{k} \\ \mathbf{q}^{\hat{j}} &= q_r^{\hat{j}} + q_i^{\hat{j}}\mathbf{i} - q_j^{\hat{j}}\mathbf{j} - q_k^{\hat{j}}\mathbf{k} \\ \mathbf{q}^{\hat{k}} &= q_r^{\hat{k}} + q_i^{\hat{k}}\mathbf{i} - q_j^{\hat{k}}\mathbf{j} - q_k^{\hat{k}}\mathbf{k} \end{aligned} \tag{12}$$

We define the augmented covariance matrix as:

$$\tilde{\mathcal{C}}_{\mathbf{qq}} = \mathrm{E}\left\{\tilde{\mathbf{q}}\tilde{\mathbf{q}}^{\mathrm{H}}\right\} = \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathcal{C}_{\mathbf{qq}^{\hat{i}}} & \mathcal{C}_{\mathbf{qq}^{\hat{j}}} & \mathcal{C}_{\mathbf{qq}^{\hat{k}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{i}}} & \mathcal{C}_{\mathbf{q}^{\hat{i}}\mathbf{q}^{\hat{i}}} & \mathcal{C}_{\mathbf{q}^{\hat{i}}\mathbf{q}^{\hat{j}}} & \mathcal{C}_{\mathbf{q}^{\hat{i}}\mathbf{q}^{\hat{k}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{j}}} & \mathcal{C}_{\mathbf{q}^{\hat{j}}\mathbf{q}^{\hat{i}}} & \mathcal{C}_{\mathbf{q}^{\hat{j}}\mathbf{q}^{\hat{j}}} & \mathcal{C}_{\mathbf{q}^{\hat{j}}\mathbf{q}^{\hat{k}}} \\ \mathcal{C}^{\mathrm{H}}_{\mathbf{qq}^{\hat{k}}} & \mathcal{C}_{\mathbf{q}^{\hat{k}}\mathbf{q}^{\hat{i}}} & \mathcal{C}_{\mathbf{q}^{\hat{k}}\mathbf{q}^{\hat{j}}} & \mathcal{C}_{\mathbf{q}^{\hat{k}}\mathbf{q}^{\hat{k}}} \end{bmatrix} \tag{13}$$

where $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose operator, and $\mathcal{C}$ 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 [4], which states that the quaternion vector $\mathbf{q}$ is uncorrelated with its involutions $\mathbf{q}^{\hat{i}}$, $\mathbf{q}^{\hat{j}}$, and $\mathbf{q}^{\hat{k}}$, i.e., $\mathcal{C}_{\mathbf{qq}^{\hat{i}}} = \mathcal{C}_{\mathbf{qq}^{\hat{j}}} = \mathcal{C}_{\mathbf{qq}^{\hat{k}}} = 0$. Under this assumption, Equation 13 simplifies to:

$$\tilde{\mathcal{C}}_{\mathbf{qq}} = \begin{bmatrix} \mathcal{C}_{\mathbf{qq}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{i}}\mathbf{q}^{\hat{i}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{j}}\mathbf{q}^{\hat{j}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{C}_{\mathbf{q}^{\hat{k}}\mathbf{q}^{\hat{k}}} \end{bmatrix} = \sum_{\delta \in \{r,i,j,k\}} \mathrm{E}\left\{\mathbf{q}_\delta^2\right\}\mathbf{I} \tag{14}$$

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, we calculate

the mean μ of the quaternion input and determine the quaternion variance σ^2 as the average of the variances of its individual components, expressed as follows:

$$\begin{aligned}\mu &= \frac{1}{C}\sum_{c=1}^{C}(q_{r,c}+q_{i,c}\mathbf{i}+q_{j,c}\mathbf{j}+q_{k,c}\mathbf{k}) \\ &= \overline{q_r}+\overline{q_i}\mathbf{i}+\overline{q_j}\mathbf{j}+\overline{q_k}\mathbf{k}\end{aligned} \tag{15}$$

$$\begin{aligned}\sigma^2 &= \frac{1}{4}\sum_{\delta\in\{r,i,j,k\}}\left(\mathrm{E}\{\mathbf{q}_\delta^2\}-(\mathrm{E}\{\mathbf{q}_\delta\})^2\right) \\ &= \frac{1}{4C}\sum_{\delta\in\{r,i,j,k\}}\sum_{c=1}^{C}\left(\mathbf{q}_{\delta,c}-\bar{\mathbf{q}}_\delta\right)^2\end{aligned} \tag{16}$$

here, $\overline{(\cdot)}$ denotes the mean operation and C represents the channel dimension of each quaternion component. Thus, we can define the complete QGN process as follows:

$$\mathbf{QGN}(x) = \gamma\left(\frac{\mathbf{x}-\mu_q}{\sqrt{\sigma^2+\varepsilon}}\right)+\beta \tag{17}$$

where β is a learnable quaternion shift parameter, γ is a learnable scalar scaling parameter.

Finally, the overall process of the quaternion wavelet embedding can be expressed as:

$$\hat{i} = \mathbf{QGN}(\mathbf{QWT}(I)) \tag{18}$$

4 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:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}_r^2+\mathbf{x}_i^2+\mathbf{x}_j^2+\mathbf{x}_k^2+\varepsilon} \tag{19}$$

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

$$\mathbf{x}_r' = \frac{\mathbf{x}_r}{\|\mathbf{x}\|},\quad \mathbf{x}_i' = \frac{\mathbf{x}_i}{\|\mathbf{x}\|},\quad \mathbf{x}_j' = \frac{\mathbf{x}_j}{\|\mathbf{x}\|},\quad \mathbf{x}_k' = \frac{\mathbf{x}_k}{\|\mathbf{x}\|} \tag{20}$$

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

$$\mathbf{QLN}(\mathbf{x}) = \alpha\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)+\beta \tag{21}$$

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

References

[1] S Akila Agnes, A Arun Solomon, and K Karthick. Wavelet u-net++ for accurate lung nodule segmentation in ct scans: Improving early detection and diagnosis of lung cancer. *Biomedical Signal Processing and Control*, 87:105509, 2024.

[2] A Shamim Banu and S Deivalakshmi. Awunet: leaf area segmentation based on attention gate and wavelet pooling mechanism. *Signal, Image and Video Processing*, 17(5): 1915–1924, 2023.

[3] Michela Ricciardi Celsi, Simone Scardapane, and Danilo Comminiello. Quaternion neural networks for 3d sound source localization in reverberant environments. In *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2020. doi: 10.1109/MLSP49062.2020.9231809.

[4] Clive Cheong Took and Danilo P. Mandic. Augmented second-order statistics of quaternion random signals. *Signal Processing*, 91(2):214–224, 2011. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2010.06.024. URL https://www.sciencedirect.com/science/article/pii/S0165168410002719.

[5] Vladimir Frants, Sos Agaian, and Karen Panetta. Qcnn-h: Single-image dehazing using quaternion neural networks. *IEEE Transactions on Cybernetics*, 53(9):5448–5458, 2023.

[6] Chase J. Gaudet and Anthony S. Maida. Deep quaternion networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2018. doi: 10.1109/IJCNN.2018.8489651.

[7] Tamjid Imtiaz, Shaikh Anowarul Fattah, and Sun-Yuan Kung. Bawgnet: Boundary aware wavelet guided network for the nuclei segmentation in histopathology images. *Computers in Biology and Medicine*, 165:107378, 2023.

[8] Zongying Lai, Xiaobo Qu, Yunsong Liu, Di Guo, Jing Ye, Zhifang Zhan, and Zhong Chen. Image reconstruction of compressed sensing mri using graph-based redundant wavelet transform. *Medical image analysis*, 27:93–104, 2016.

[9] Mingfeng Lin, Quan Lan, Chenxi Huang, Bin Yang, and Yuexin Yu. Wavelet-based u-shape network for bioabsorbable vascular stents segmentation in ivoct images. *Frontiers in Physiology*, 15:1454835, 2024.

[10] Yi Tay, Aston Zhang, Anh Tuan Luu, Jinfeng Rao, Shuai Zhang, Shuohang Wang, Jie Fu, and Siu Cheung Hui. Lightweight and efficient neural natural language processing with quaternion networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1494–1503. Association for Computational Linguistics, July 2019. doi: 10.18653/v1/P19-1145. URL https://aclanthology.org/P19-1145/.

[11] The Van Le and Jin Young Lee. Specular highlight removal using quaternion transformer. *Computer Vision and Image Understanding*, 249:104179, 2024.

[12] Yuxin Wu and Kaiming He. Group normalization. In *Computer Vision – ECCV 2018*, pages 3–19, 2018. ISBN 978-3-030-01261-8.

[13] Jianming Zhang, Zhigao Zeng, Pradip Kumar Sharma, Osama Alfarraj, Amr Tolba, and Jin Wang. A dual encoder crack segmentation network with haar wavelet-based high–low frequency attention. *Expert Systems with Applications*, 256:124950, 2024.

[14] Yawu Zhao, Shudong Wang, Yulin Zhang, Sibo Qiao, and Mufei Zhang. Wranet: wavelet integrated residual attention u-net network for medical image segmentation. *Complex intelligent systems*, 9(6):6971–6983, 2023.

[15] Zewen Zheng, Guoheng Huang, Xiaochen Yuan, Chi-Man Pun, Hongrui Liu, and Wing-Kuen Ling. Quaternion-valued correlation learning for few-shot semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(5): 2102–2115, 2022.

[16] Zhizhen Zhou, Yejing Huo, Guoheng Huang, An Zeng, Xuhang Chen, Lian Huang, and Zinuo Li. Qean: quaternion-enhanced attention network for visual dance generation. *The Visual Computer*, pages 1–13, 2024.