WAQNIQA: Wavelet-Augmented Quaternion Network for No-Reference Image Quality Assessment

Yejing Huo, Guoheng Huang*, Zhiwen Yu *Senior Member, IEEE*, Xiaochen Yuan *Senior Member, IEEE*, Chi-Man Pun*, *Senior Member, IEEE* , Lianglun Cheng, Hanyu Shi, Xuhang Chen, Zehong Chen

***Abstract*—No-reference image quality assessment (NR-IQA) plays a pivotal role in computer vision by enabling image quality evaluation without reference images. While recent CNN and Transformer-based methods have advanced feature extraction, they face significant limitations. CNNs exhibit local feature bias, limiting their ability to capture global dependencies and complex structures critical to understanding diverse distortions. Transformers, despite modeling nonlocal dependencies through multi-head attention, suffer from quadratic computational complexity with spatial dimensions, hindering efficient multiscale analysis. Moreover, their attention mechanisms frequently overlook critical inter-channel dependencies, which are vital for capturing fine details in texture-rich images. Coupled with difficulties in handling high-noise environments and complex textures, this results in limited real-world accuracy and poor generalization across diverse datasets and unknown distortions. To address these challenges, we propose WAQNIQA, a wavelet-augmented Quaternion network for NR-IQA featuring two synergistic modules: the Wavelet-Infused Adaptive Attention (WIAA) module, which combines wavelet transforms and attention mechanisms for robust multi-scale analysis, and the Quaternion Collaborative Feature Enhancement (QCFE) module, which enhances feature representation through Quaternion learning. This integration balances global image characteristics with local detail preservation for comprehensive quality assessment. We also introduce PowerGridIQ, the first NR-IQA dataset tailored for power grid scenarios. Extensive experiments on PowerGridIQ and six public datasets demonstrate that WAQNIQA consistently outperforms existing state-of-the-art methods. Additionally, WAQNIQA shows competitive performance on the AGIQA-1K dataset for AI-Generated Content image quality assessment, showcasing its robustness and generalization capabilities without explicit text-to-image semantic alignment training. Our code is available at `https://github.com/king-huoye/WAQNIQA`.**

***Index Terms*—Image quality assessment, Quaternion, Wavelet transform, Deep learning**

I. Introduction

Every day, millions of images are uploaded to social platforms such as Twitter and Facebook [1], often subjected to various processes like capturing, compression, storage, and transmission. These operations can degrade image quality by introducing artifacts, losing visual details, and compromising overall clarity. The presence of low-quality images can significantly impact user experience, impair visual perception, and, in some cases, lead to undesirable consequences [2]–[4]. Consequently, the development of robust image quality estimation [5]–[7] models is critical. These models, particularly those powered by artificial intelligence, can autonomously predict the visual quality of an image, enabling quality assessments and ensuring an optimal experience across diverse platforms.

Image quality assessment is typically categorized into three types: Full-Reference IQA (FR-IQA) [8], [9], Reduced-Reference IQA (RR-IQA) [10], and No-Reference IQA (NR-IQA) [11], [12]. FR-IQA derives the quality score of a distorted image by comparing it with the original high-definition image. RR-IQA, on the other hand, assesses the quality of a test image by comparing it with partial feature information from a reference image. However, the original reference image is often unavailable in many practical scenarios. To overcome this limitation, NR-IQA provides a perceptual quality score without needing any reference image, making it more applicable to real-world situations. Therefore, this paper will discuss NR-IQA to address the challenge of evaluating image quality without relying on reference data. Traditional NR-IQA methods are generally categorized into two types: those based on natural scene statistics (NSS) [13]–[15] and those based on machine learning [16], [17]. NSS-based methods assume that undistorted images follow certain statistical patterns, and deviations from these patterns are used to assess image quality. Machine learning-based approaches, such as those using dictionary learning and support vector regression, attempt to model image quality by learning visual features from data. However, both approaches often struggle with real-world data and complex distortions. With the rise of

This work was supported in part by the Key Areas Research and Development Program of Guangzhou under Grant 2023B01J0029, in part by the Guangdong Provincial Key Laboratory of Cyber-Physical System under Grant 2020B1212060069, in part by the Science and Technology Development Fund, Macau SAR, under Grants 0193/2023/RIA3 and 0079/2025/AFJ, and in part by the University of Macau under Grant MYRG-GRG2024-00065-FST-UMDF.

Yejing Huo, Guoheng Huang, Hanyu Shi and Lianglun Cheng are with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: 2112305121@mail2.gdut.edu.cn; kevinwong@gdut.edu.cn; 13420626887@163.com; llcheng@gdut.edu.cn). * is corresponding author.

Zhiwen Yu is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510650, China, and also with Pengcheng Laboratory, Shenzhen 518066, China (e-mail: zhwyu@scut.edu.cn).

Xiaochen Yuan is with the Faculty of Applied Sciences, Macao Polytechnic University, Macau, China (e-mail: xcyuan@mpu.edu.mo).

Chi-Man Pun is with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: cmpun@umac.mo).

Xuhang Chen is with the School of Computer Science and Engineering, Huizhou University, Guangdong, China (email: xuhangc@hzu.edu.cn).

Zehong Chen is with the department of Jiangmen Power Grid, Guangdong Grid Co, Jiangmen,Guangdong(email:281150668@qq.com)

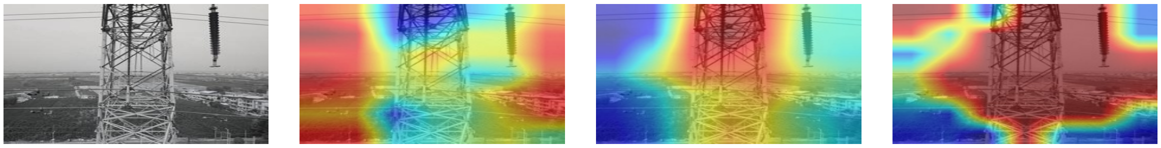

(a) Distorted image (b) CNN-based (c) Transformer-based (d) Ours

Fig. 1. Visualization comparison between different NR-IQA methods. CNN-based methods focus intensely on specific parts of the tower structure and power lines, but their hyper-local attention misses broader context and overlooks global structures and textures, particularly in noisy or complex areas. Transformer-based methods, while better at capturing global dependencies, lack the precision needed to highlight crucial local details, especially in high-texture regions like the metal lattice of the tower and along power lines. In contrast, WAQNIQA demonstrates a more balanced and nuanced attention distribution, effectively capturing both local textures and global structures. Its heatmap shows a sharper focus on critical areas, such as the edges of the tower and power lines, which are precisely the regions that field workers monitor during inspections.

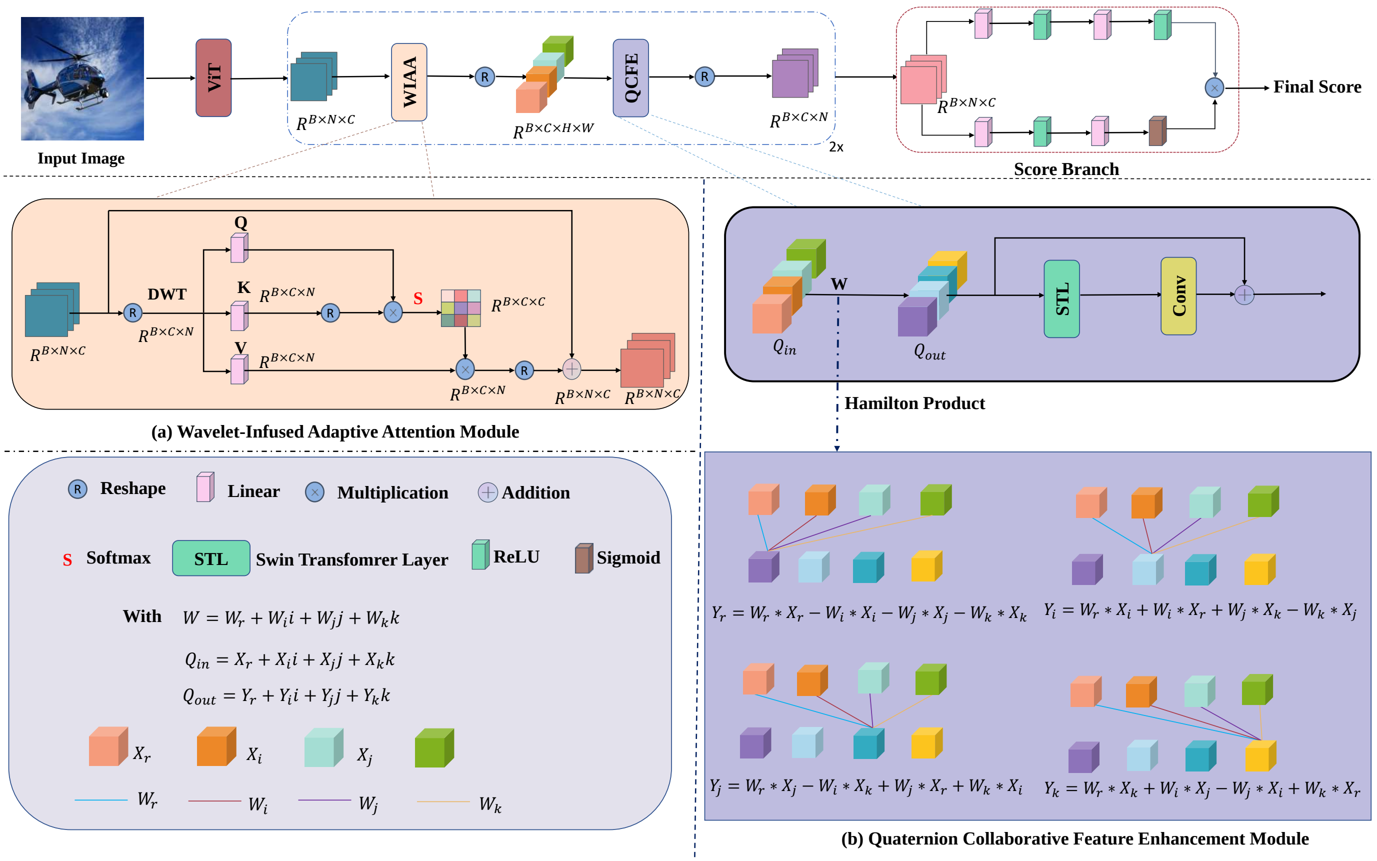
deep learning, CNN-based methods have become the focus of significant research in image quality assessment. CNN-based approaches primarily aim to extract features from distorted images and predict subjective quality scores. For example, DB-CNN [18] uses a deep bilinear structure, where separate CNN models process synthetic and real distorted images, followed by bilinear pooling to combine the feature maps for quality score prediction. Another approach, BIECON [19], trains CNN models on patches of distorted images, leveraging quality scores from FR-IQA models as labels for training.

Despite the significant progress made with CNN-based methods, their reliance on local feature extraction has hindered their ability to effectively capture global dependencies, which are critical for understanding complex image structures and distortions. To address these limitations, researchers have begun exploring Transformer models. While Transformers [20] can model non-local dependencies [21] through multi-head attention, they exhibit two key limitations: 1) Their computational complexity grows quadratically with spatial dimensions, hindering efficient multi-scale analysis; 2) Their channel-agnostic attention mechanisms fail to capture critical inter-channel relationships essential for texture-rich images. To tackle these issues, MANIQA [22] proposed an innovative approach by employing attention mechanisms in both channel and spatial dimensions, using a multi-scale strategy to enhance global and local interactions across different regions of an image. While this method improves performance to some extent, it still has limitations. Firstly, although Transformer models can capture global dependencies, their high computational complexity and primary focus on spatial dimensions lead to inefficiency in dealing with multi-scale features and fail to fully utilize the information in the channel dimensions. Secondly, existing models lack the ability to adapt to high-noise environments and complex textures, resulting in limited assessment accuracy in real-world applications such as power grids (as shown in Fig. 1). Moreover, current methods have limited generalization across datasets and different application scenarios, and lack adaptability to unknown distortions and new scenarios. These challenges highlight the need for further innovation in the field of image quality assessment to develop more robust, efficient, and widely applicable models.

To address the challenges in NR-IQA, we propose WAQNIQA, a wavelet-enhanced Quaternion network. WAQNIQA is built upon two complementary modules: the Wavelet-Infused Adaptive Attention (WIAA) module and the Quaternion Collaborative Feature Enhancement (QCFE) module. The WIAA module combines wavelet transforms and attention mechanisms to handle complex textures and structural details effectively. The wavelet transform captures both spatial and frequency information, enabling robust multi-scale analysis that helps the model adapt to varying feature complexities. At the same time, the attention mechanism selectively focuses on the most critical features, ensuring the network prioritizes relevant information for accurate quality prediction. Building on the refined feature representations produced by the WIAA module, the QCFE module utilizes Quaternion representation learning to process image channels as a unified entity. This approach captures intricate inter-channel dependencies, resulting in richer and more structured feature representations. By modeling both local/global and inter-channel dependencies, the WIAA and QCFE modules work in tandem, enhancing the model's ability to learn subtle variations and complex patterns, which significantly improves generalization across diverse scenarios. Furthermore, we introduce PowerGridIQ, the first dataset specifically designed for NR-IQA tasks in power grid environments. This unique dataset comprises 100 UAV-captured images of operational sites at different heights, simulating nine real-world distortion types commonly encountered in power grid monitoring. To ensure label reliability, 50 industry professionals independently scored the 900 degraded images. Extensive experiments on PowerGridIQ and six public datasets demonstrate that WAQNIQA consistently outperforms state-of-the-art methods. Furthermore, to rigorously assess WAQNIQA's generalization capability, we compared it against state-of-the-art AI-Generated Content (AIGC) Image Quality Assessment (IQA) models. Notably, WAQNIQA demonstrated competitive performance on the AGIQA-1K dataset , which is specifically designed for AIGC image quality assessment , showcasing its robustness and generalization capabilities even without explicit training on text-to-image semantic alignment. This underscores its strong generalization and adaptability to challenging image quality assessment tasks.

In summary, our contributions are as follows:

- We propose WAQNIQA, a novel wavelet-augmented Quaternion network for NR-IQA. This network uniquely combines wavelet transforms and Quaternion learning to synergistically capture both multi-scale spatial-frequency details and intricate inter-channel dependencies, which are crucial for robustly assessing complex textures and structural information in degraded images.


Input Image
ViT
$R^{B\times N\times C}$
WIAA
R
$R^{B\times C\times H\times W}$
QCFE
$R^{B\times C\times N}$
2x
$R^{B\times N\times C}$
Final Score
Score Branch
Q
K
V
DWT
S
$R^{B\times C\times N}$
$R^{B\times C\times C}$
$R^{B\times N\times C}$
W
Q_{in}
Q_{out}
STL
Conv
Hamilton Product
Reshape
Linear
Multiplication
Addition
Softmax
Swin Transfomrer Layer
ReLU
Sigmoid
With $W = W_r + W_i i + W_j j + W_k k$
$Q_{in} = X_r + X_i i + X_j j + X_k k$
$Q_{out} = Y_r + Y_i i + Y_j j + Y_k k$
X_r
X_i
X_j
W_r
W_i
W_j
W_k
$Y_r = W_r * X_r - W_i * X_i - W_j * X_j - W_k * X_k$
$Y_i = W_r * X_i + W_i * X_r + W_j * X_k - W_k * X_j$
$Y_j = W_r * X_j - W_i * X_k + W_j * X_r + W_k * X_i$
$Y_k = W_r * X_k + W_i * X_j - W_j * X_i + W_k * X_r$

(a) Wavelet-Infused Adaptive Attention Module

(b) Quaternion Collaborative Feature Enhancement Module

Fig. 2. The overall of our WAQNIQA. The distorted image is divided into 8×8 image blocks, which are then embedded and fed into a Vision Transformer (ViT) for feature extraction. The features extracted by ViT are initially represented as $R^{B\times N\times C}$, where B, N, and C denote the batch size, number of elements in each feature map, and number of channels, respectively. These features are reshaped to $R^{B\times C\times N}$ and undergo a wavelet transform to capture spatial and frequency information, facilitating robust multi-scale analysis. The wavelet-transformed features are then processed by an attention mechanism, and the output is obtained through a residual connection with the original input feature. This integration of wavelet transform and attention mechanisms enables the WIAA module to effectively capture complex textures and distinct structural information. Subsequently, the enhanced features are further processed by the Quaternion Collaborative Feature Enhancement (QCFE) module, which leverages Quaternion representation learning to model image channels as a unified entity. This allows QCFE to capture intricate inter-channel dependencies and generate richer, more structured feature representations. Then, the enhanced features are processed in the score-branch structure to obtain the corresponding perceived quality scores.

- We design the Wavelet-Infused Adaptive Attention (WIAA) module, which integrates wavelet transforms with attention mechanisms to enhance local and global feature interactions, allowing the model to adapt to varying feature complexities.
- We introduce the Quaternion Collaborative Feature Enhancement (QCFE) module, which leverages Quaternion representation learning to model channels as a unified entity, leading to richer and more structured feature representations.
- We introduce the PowerGridIQ dataset, the first specifically designed for NR-IQA in power grid environments. It comprises 100 UAV-captured images of operational sites, each affected by nine types of realistic degradations. Evaluated by 50 industry experts, PowerGridIQ provides a valuable resource for advancing NR-IQA research in industrial information.
- We conducted extensive experiments on the PowerGridIQ dataset as well as six public datasets. The results demonstrate that our method outperforms all existing methods available today. Furthermore, our model exhibits competitive performance on AI-Generated Content (AIGC) image quality assessment, showcasing its strong generalization capabilities even without explicit training on text-to-image semantic alignment.

II. Related Work

A. No-Reference Image Quality Assessment

No-Reference Image Quality Assessment (NR-IQA) methods [23] evaluate image quality without relying on reference images. These approaches are typically divided into natural scene statistics (NSS)-based methods [13], [14], [24] and learning-based methods [16], [17]. NSS methods assume that handcrafted features from natural images follow regular patterns, enabling detection of distortions such as spatial irregularities [25], [26], gradient changes [25], and DCT artifacts [27]. However, their effectiveness declines when dealing with complex real-world distortions. With the rise of deep learning, CNN-based models [28], [29] have emerged, typically predicting quality from uniformly sampled patches. HyperIQA [12] separates low- and high-level features for better perception, while MEON [30] integrates distortion identification and quality prediction in an end-to-end manner. Despite their local feature modeling ability, CNNs suffer from limited receptive fields and poor non-local representation, restricting performance on complex image structures. To address this, Transformer-based methods [20], [22], [31] have been introduced, leveraging self-attention to capture global dependencies. MUSIQ [21] encodes multi-scale features to address input size issues, MANIQA [22] applies multi-dimensional attention, and SaTQA [32] incorporates supervised contrastive learning for enhanced feature extraction. While effective globally, these models often overlook fine-grained local details and struggle in complex real-world settings—such as power grid environments with high noise—resulting in suboptimal quality prediction.

B. Wavelet Transform in Computer Vision

Wavelet transform (WT) is highly effective for time-frequency analysis due to its reversibility and ability to retain complete signal information. It has been widely adopted in computer vision tasks [33]–[35]. In image denoising [36], WT separates signal and noise across different scales, effectively reducing noise while preserving fine details. WAEN [37] integrates attention and wavelet embeddings for improved multi-scale reconstruction in video super-resolution. AWNet [38] combines non-local attention with discrete WT to enhance image quality and signal processing. In image classification, WaveNet [39] applies WT for channel compression, reducing redundancy while maintaining essential features to improve accuracy and efficiency. Building on these strengths, we integrate WT into our WIAA module. Its multi-scale analysis enables the detection of distortions at different frequency bands, such as compression artifacts in low frequencies and noise in high frequencies, thereby enhancing prediction accuracy. WT also preserves critical image structures, separates noise from meaningful content, and captures frequency characteristics necessary for evaluating image sharpness. These advantages are central to the performance of our WIAA module in distortion detection and quality assessment.

C. Quaternion Networks

A Quaternion is a hyper-complex number first introduced by Hamilton [40], and has since been applied across diverse fields. In deep learning, Quaternions have been used in 3D audio processing [41], few-shot image segmentation [42], and human pose estimation [43], [44]. For example, QuaterNet [43] employs Quaternions to represent joint rotations in RNNs and CNNs, achieving strong performance in long-term motion prediction. Tay et al. [45] further proposed Quaternion-based attention and transformer models, offering lightweight and memory-efficient alternatives. Motivated by these advantages, we incorporate Quaternion representation into our QCFE module for image quality assessment. Quaternions enable unified modeling of multi-channel image data by capturing complex inter-channel dependencies, which is critical for accurate quality evaluation. Their structured mathematical form enhances the model's ability to learn intricate image patterns. Moreover, their inherent rotational invariance ensures consistent quality predictions regardless of image orientation. Together, these properties improve the robustness and accuracy of our QCFE module across diverse image conditions.

III. Method

A. Overview

Given a distorted image $\boldsymbol{I} \in \mathbb{R}^{H \times W \times 3}$, where H and W represent the height and width, our goal is to estimate its perceptual quality score. In the following, we will use N=H*W to represent the number of elements in each feature map. Let g_θ represent the Vision Transformer (ViT) [46] with learnable parameters θ, and let $\mathbf{E}_l \in \mathbb{R}^{n \times d_l \times H_l W_l}$ denote the features extracted from the l^{th} layer of ViT, where $l \in \{1, 2, \ldots, 12\}$. Here, n represents the batch size, and d_l, H_l, and W_l denote the channel size, width, and height of the feature map from the l^{th} layer, respectively. To extract features from different semantic levels, we select the feature maps from four layers $\mathbf{E}_l$, where $l \in \{7, 8, 9, 10\}$, and concatenate them to obtain a combined feature representation $\mathcal{G} \in \mathbb{R}^{n \times \sum_l d_l \times H_l W_l}$. Next, we apply the Wavelet-Infused Adaptive Attention (WIAA) module, which enhances the interaction between local and global features by leveraging wavelet decomposition and attention mechanisms to focus on important multi-scale features. Following this, the enriched feature map $\mathcal{G}$ is processed by the Quaternion Collaborative Feature Enhancement (QCFE) module. This module captures intricate inter-channel dependencies using Quaternion operations to improve the representation of complex patterns within the features. After the feature enhancement via WIAA and QCFE, the final score can be computed by the score branch (it consists of a sequence of linear layers and activation functions):

$$\tilde{Score} = \frac{\sum_{i=1}^{M} \omega_i \times s_i}{\sum_{i=1}^{M} \omega_i} \tag{1}$$

where M represents the total number of patches extracted from each image, ω_i represents the importance weight of each image patch, s_i denotes the quality score of each image patch.

B. Preliminary

1) Wavelet transform: Wavelet Transform (WT) is a signal processing tool with the ability of localized analysis in both time domain and frequency domain. Unlike the Fourier transform, which can only provide information about the overall frequency of the signal, the wavelet transform is able to reveal the characteristics of the signal at different times and at different scales (corresponding to the frequency) simultaneously by using "wavelet" basis functions that are limited in time and frequency. This allows it to show superior performance when dealing with non-stationary signals, such as transient phenomena, abrupt changes, or multi-scale structures.

The basic idea of wavelet transform is to decompose a signal into a series of wavelet coefficients with different scales and locations. These coefficients reflect the local information of the signal at the corresponding scale and position. By adjusting the scale parameter (dilation) and the displacement parameter (translation), the wavelet function can move freely in the time-frequency plane, so as to realize the fine analysis of the signal.

The Continuous Wavelet Transform (CWT) is defined as follows:

$$W_x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{2}$$

where a is the scaling parameter, which controls the scaling of the wavelet function as well as the corresponding frequency. When $|a|$ is large, the wavelet is stretched, corresponding to low frequency information. When $|a|$ is small, the wavelet is compressed, which corresponds to high-frequency information. b is the displacement information, which controls the position of the wavelet function on the time axis. $\psi(t)$ is the mother wavelet function, which is an oscillatory decay function satisfying certain mathematical conditions, usually with zero mean and finite energy. $\psi^*(t)$ is the complex conjugate of $\psi(t)$. $\frac{1}{\sqrt{|a|}}$ is the normalization factor used to preserve energy conservation.

Discrete Wavelet Transform (DWT) is the discrete form of CWT, which makes the computation more efficient by discretizing the scale and displacement parameters. DWT is usually implemented by Multi-Resolution Analysis (MRA), which decomposes the signal into a series of low-frequency (approximate) and high-frequency (detailed) components.

2) Quaternion: We begin by introducing essential Quaternion concepts. Quaternions are rank-4 hyper-complex numbers that extend complex numbers in a non-commutative manner. In our methods, the relationship between Hamilton products and Quaternion algebra is fundamental, underpinning the approaches developed in this study. This overview sets the foundation for subsequent discussions and applications.

A Quaternion $\mathcal{Q}$ in the Quaternion domain D, where $\mathcal{Q} \in D$, can be represented in the form:

$$\mathcal{Q} = e + f\mathbf{i} + g\mathbf{j} + h\mathbf{k} \tag{3}$$

where $e, f, g,$ and h are real numbers, and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the Quaternion unit basis elements. In this representation, e is referred to as the real part of the Quaternion, while $f\mathbf{i}+g\mathbf{j}+h\mathbf{k}$ is the imaginary part. The imaginary units $\mathbf{i}, \mathbf{j}, \mathbf{k}$ follow specific multiplication rules: $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$. These rules result in a non-commutative algebra, which provides Quaternions their unique properties in representing rotations and orientations in three-dimensional space.

A pure Quaternion is a special type of Quaternion where the real part is zero, i.e., $e = 0$. Consequently, a pure Quaternion takes the form:

$$\mathcal{Q} = f\mathbf{i} + g\mathbf{j} + h\mathbf{k} \tag{4}$$

which can be interpreted as a three-dimensional vector.

Various operations can be defined on Quaternions, including addition, scalar multiplication, conjugation, and Quaternion multiplication. These operations are fundamental for manipulating and combining Quaternions in applications such as computer graphics, robotics, and physics.

1. Addition: The addition of two Quaternions $\mathcal{Q}$ and R is performed element-wise:

$$\mathcal{Q} + R = (e + e') + (f + f')\mathbf{i} + (g + g')\mathbf{j} + (h + h')\mathbf{k} \tag{5}$$

where $\mathcal{Q} = e + f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$ and $R = e' + f'\mathbf{i} + g'\mathbf{j} + h'\mathbf{k}$. This operation is commutative and associative, similar to vector addition in real space.

2. Scalar Multiplication: The multiplication of a Quaternion Q by a scalar γ scales each component of the Quaternion:

$$\gamma\mathcal{Q} = \gamma e + \gamma f\mathbf{i} + \gamma g\mathbf{j} + \gamma h\mathbf{k} \tag{6}$$

where $\gamma \in \mathbb{R}$. This operation stretches or compresses the Quaternion without changing its orientation.

3. Conjugation: The conjugate of a Quaternion $\mathcal{Q}$, denoted as $\mathcal{Q}^\star$, is obtained by negating its imaginary components:

$$\mathcal{Q}^\star = e - f\mathbf{i} - g\mathbf{j} - h\mathbf{k} \tag{7}$$

Conjugation is an important operation in Quaternion algebra as it reflects the Quaternion across the real axis and is used to compute the inverse of a Quaternion.

4. Norm: The unit Quaternion $\mathcal{Q}^\triangleleft$ is defined as:

$$\mathcal{Q}^\triangleleft = \frac{\mathcal{Q}}{\sqrt{r^2 + x^2 + y^2 + z^2}} \tag{8}$$

5. Multiplication: The multiplication of two Quaternions $\mathcal{Q}$ and R is a more complex operation that involves both the real and imaginary parts. The Quaternion product $\mathcal{Q} \otimes R$ is defined as:

$$\begin{aligned} \mathcal{Q} \otimes R = &(ee' - ff' - gg' - hh') \\ &+ (fe' + ef' - hg' + gh')\mathbf{i} \\ &+ (ge' + hf' + eg' - fh')\mathbf{j} \\ &+ (he' - gf' + fg' + eh')\mathbf{k} \end{aligned} \tag{9}$$

where $\mathcal{Q} = e + f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$ and $R = e' + f'\mathbf{i} + g'\mathbf{j} + h'\mathbf{k}$. Unlike addition, Quaternion multiplication is non-commutative (i.e., $\mathcal{Q} \otimes R \neq R \otimes \mathcal{Q}$), which means the order of multiplication matters. This property is particularly useful for representing 3D rotations, as the multiplication of Quaternions allows for the combination of rotations in a concise and computationally efficient manner.

The equation above clearly illustrates the interaction between Quaternions $\mathcal{Q}$ and R, highlighting the importance of the Hamilton product in Quaternion neural networks.

C. Wavelet-Infused Adaptive Attention Module

Self-attention is a foundational mechanism in Transformers [20], which calculates attention scores by computing dot products between queries, keys, and values. While effective, traditional self-attention does not fully utilize the channel information and often lacks sensitivity to fine-grained features, which are crucial for detailed image analysis and understanding. Inspired by multi-scale feature extraction techniques [33], [39], we propose the Wavelet-Infused Adaptive Attention

(WIAA) module (as shown in Fig. 2 (a)) to overcome these limitations and enhance the model's capacity to capture both local and global information.

Given input features $X \in \mathbb{R}^{B\times C\times N}$, where B, C, and $N = H \times W$ denote the batch size, number of channels, and spatial elements respectively. We perform a haar wavelet transform to extract a multi-scale representation of the features. The wavelet transform is particularly useful in decomposing input features into various frequency components, which allows it to retain both local and global information from the signal. This process is represented mathematically as:

$$\begin{aligned} X_{LL}^{(b,c)} &= \mathrm{approx}\left(X_{b,c}\right) \\ X_{LH}^{(b,c)} &= \mathrm{horiz\text{-}detail}\left(X_{b,c}\right) \\ X_{HL}^{(b,c)} &= \mathrm{vert\text{-}detail}\left(X_{b,c}\right) \\ X_{HH}^{(b,c)} &= \mathrm{diag\text{-}detail}\left(X_{b,c}\right) \end{aligned} \tag{10}$$

where $b \in 1, \ldots, B$ and $c \in 1, \ldots, C$. Here, $X_{LL}^{(b,c)}$ denotes the low-frequency approximation component that preserves the primary structural information of the image. $X_{LH}^{(b,c)}$ represents the horizontal detail component, which captures vertical edge features. $X_{HL}^{(b,c)}$ corresponds to the vertical detail component that extracts horizontal edge information. Finally, $X_{HH}^{(b,c)}$ is the diagonal detail component responsible for capturing diagonal texture patterns.

The complete wavelet-transformed representation is denoted as:

$$\begin{aligned} X_{b,c}^{(w)} = \{X_{LL}^{(b,c)}, X_{LH}^{(b,c)}, X_{HL}^{(b,c)}, X_{HH}^{(b,c)}\} \\ \forall b \in \{1, \ldots, B\},\ c \in \{1, \ldots, C\} \end{aligned} \tag{11}$$

After obtaining the wavelet-transformed features $X^{(w)}$, the WIAA module generates the matrices of queries (Q'), keys (K'), and values (V') through linear transformations. Then computed using scaled dot-product attention, which measures the similarity between queries and keys to produce attention weights that are applied to the values These transformations can be expressed as:

$$\alpha^{(b,c)} = \sigma\left(\frac{\langle W_Q' X^{(w)}, W_K' X^{(w)}\rangle}{\sqrt{d_k}}\right) \tag{12}$$

where σ denotes softmax, $W_{Q'}$ and $W_{K'}$ are learnable projection matrices for queries and keys respectively, d_k is the key dimension, and $\langle\cdot,\cdot\rangle$ represents the inner product operation.

This will simultaneously optimizes both local detail preservation captured by the high-frequency subbands (X_{LH}, X_{HL}, X_{HH}) and global structural representation encoded in the low-frequency component (X_{LL}).

The resulting attention-weighted features provide a context-aware representation that emphasizes informative multi-scale patterns while suppressing irrelevant information, effectively capturing both spatial and frequency-domain characteristics essential for image quality assessment. By integrating wavelet transforms with adaptive attention mechanisms, the WIAA module enables an effective balance between global structural information and local texture details. The spatially-aware attention further operates on representations enriched with frequency-domain cues, allowing the model to natively attend to salient features across spatial-frequency domains.

This dual-domain representation not only enhances the model's adaptability to heterogeneous feature distributions but also mitigates the loss of high-frequency information, a common drawback in standard Transformer-based models. As a result, the WIAA module significantly improves performance in image quality assessment tasks by preserving perceptual fidelity across multiple scales and capturing subtle variations in both spatial and spectral domains.

D. Quaternion Collaborative Feature Enhancement Module

While standard CNNs and Transformers have advanced feature extraction, they fundamentally rely on real-valued scalar computations. In these architectures, inter-channel relationships are modeled by summing up independent scalar convolutions, which treats color channels or feature maps as separate entities before aggregation. This process often discards the latent structural coupling inherent in color images—specifically the phase relationship between the red, green, and blue channels.

In contrast, Image Quality Assessment (IQA) relies heavily on preserving these holistic structural discrepancies caused by distortions. Quaternion algebra offers a theoretically grounded solution by treating multiple channels as a single hyper-complex vector entity rather than independent scalars. Through the Hamilton product, Quaternions enforce a rigorous interaction mechanism where each output component is a linear combination of all input components, modulated by rotation and scaling in a high-dimensional space. This allows the network to capture nonlinear cross-channel dependencies and phase information naturally, which are critical for detecting complex distortions (such as color noise and chromatic aberration) that disrupt inter-channel coherence, without requiring the excessive parameter depth typical of scalar networks. We propose the Quaternion Collaborative Feature Enhancement (QCFE) module, which uses quaternion convolution to process multiple feature channels holistically, learning richer representations in multidimensional space. As shown in Fig. 2 (b), QCFE enhances network adaptability to complex input conditions while efficiently modeling both channel-wise and feature-wise interactions.

The QCFE leverages Quaternion convolutions to enhance the representation learning of deep networks by modeling complex inter-channel dependencies. Traditional convolutional operations handle each channel independently, often missing intricate patterns and interactions across channels. In contrast, Quaternion convolutions treat groups of four channels as a unified Quaternion entity, allowing for richer and more structured feature representations.

Given an input feature map $\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels (which should be divisible by 4 for Quaternion operations), and H and W are the spatial dimensions, we represent this feature map in Quaternion form:

$$\mathbf{X} = \mathbf{X}_r + \mathbf{X}_i\mathbf{i} + \mathbf{X}_j\mathbf{j} + \mathbf{X}_k\mathbf{k} \tag{13}$$

where $\mathbf{X}_r, \mathbf{X}_i, \mathbf{X}_j$, and $\mathbf{X}_k \in \mathbb{R}^{B\times(C/4)\times H\times W}$ correspond to the real and imaginary parts of the Quaternion representation. A Quaternion convolution kernel $\mathbf{W}$ is similarly defined as:

$$\mathbf{W} = \mathbf{W}_r + \mathbf{W}_i\mathbf{i} + \mathbf{W}_j\mathbf{j} + \mathbf{W}_k\mathbf{k} \tag{14}$$

where $\mathbf{W}_r, \mathbf{W}_i, \mathbf{W}_j$, and $\mathbf{W}_k$ are the real and imaginary parts of the kernel.

The output feature map $\mathbf{Y}$ after applying Quaternion convolution can be written as:

$$\mathbf{Y} = \mathbf{Y}_r + \mathbf{Y}_i\mathbf{i} + \mathbf{Y}_j\mathbf{j} + \mathbf{Y}_k\mathbf{k} \tag{15}$$

where each component of $\mathbf{Y}$ is computed using a combination of convolutions between the components of $\mathbf{X}$ and $\mathbf{W}$. Specifically, the real part $\mathbf{Y}_r$ and the imaginary parts $\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_k$ are calculated as follows:

$$\mathbf{Y}_r = \mathbf{W}_r * \mathbf{X}_r - \mathbf{W}_i * \mathbf{X}_i - \mathbf{W}_j * \mathbf{X}_j - \mathbf{W}_k * \mathbf{X}_k \tag{16}$$

$$\mathbf{Y}_i = \mathbf{W}_r * \mathbf{X}_i + \mathbf{W}_i * \mathbf{X}_r + \mathbf{W}_j * \mathbf{X}_k - \mathbf{W}_k * \mathbf{X}_j \tag{17}$$

$$\mathbf{Y}_j = \mathbf{W}_r * \mathbf{X}_j - \mathbf{W}_i * \mathbf{X}_k + \mathbf{W}_j * \mathbf{X}_r + \mathbf{W}_k * \mathbf{X}_i \tag{18}$$

$$\mathbf{Y}_k = \mathbf{W}_r * \mathbf{X}_k + \mathbf{W}_i * \mathbf{X}_j - \mathbf{W}_j * \mathbf{X}_i + \mathbf{W}_k * \mathbf{X}_r \tag{19}$$

here, $*$ denotes the standard 2D convolution operation. The Quaternion convolution effectively captures inter-channel dependencies by computing weighted combinations of input feature maps across four channels, incorporating both their spatial and cross-channel relationships.

The core innovation of the QCFE module lies in its utilization of the Hamilton product to induce 'collaborative' feature refinement. Unlike standard convolutions that perform independent sliding windows, the Quaternion convolution in QCFE forces a coupled interaction among the four feature components (r, i, j, k). Mathematically, as shown in (16)–(19), the calculation of the real part Y_r is not solely dependent on the input real part X_r, but is negatively modulated by the interaction of imaginary parts ($W_i * X_i$, etc.). This mechanism acts as a holistic regularization, ensuring that features are not just extracted but 'calibrated' against each other. For IQA, this is crucial: finding a distortion in one channel (e.g., luminance noise in the real part) allows the network to immediately adjust the weights for the chrominance channels (imaginary parts), effectively modeling how distortions propagate across perception. The QCFE thus transforms the feature map from a loose collection of responses into a tightly coupled structural representation, significantly enhancing the model's sensitivity to complex, mixed-type distortions.

The resulting feature map $\mathbf{Y}$ after Quaternion convolution is inherently more expressive, as it encapsulates complex relationships among channels. This can be critical in scenarios involving fine-grained classification, segmentation, or any task requiring robust feature extraction under challenging conditions, such as image quality assessment.

$$\mathbf{Y} = \mathrm{QC}(\mathbf{X}, \mathbf{W}) = (\mathbf{W}_r, \mathbf{W}_i, \mathbf{W}_j, \mathbf{W}_k) \otimes (\mathbf{X}_r, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) \tag{20}$$

where $\otimes$ denotes the Hamilton product applied through convolution. This operation serves as the core mechanism for synergistic feature modeling in the proposed QCFE module.

Following the Quaternion convolution, the feature map Y undergoes further refinement through two sequential internal Transformer layers. This process captures both local and global feature interactions:

$$F_{i,1} = O_{stl_{i,1}}(Y) \quad i \in \{1, 2\} \tag{21}$$

where $O_{stl_{i,j}}$ denotes the j-th Swin Transformer layer of the i-th block. This refinement primarily utilizes multi-head attention to model these interactions, as described by the following formulas (the operation for $F_{i,2}$ is analogous):

$$\begin{aligned} Q_i &= W_Q \cdot Y \\ K_i &= W_K \cdot Y \\ V_i &= W_V \cdot Y \end{aligned} \tag{22}$$

After computing the queries, keys, and values, the attention scores are calculated as follows:

$$\mathrm{AttnScore}_i = \mathrm{softmax}\left(\frac{Q_i(K_i)^\top}{\sqrt{d_k}} + B_{rel}\right) \tag{23}$$

where B_{rel} is the relative position matrix and d_k is the dimension of the key vector. The attention scores are then used to weight the value vectors, resulting in the output of the multi-head attention mechanism:

$$H_i = \mathrm{AttnScore}_i V_i \tag{24}$$

Subsequently, we refine the features through a linear transformation:

$$F_{i,1} = W_O^{(1)}\mathrm{Concat}(H_i) + Y \tag{25}$$

where W_O represents the matrix of the linear transformation.

The second Transformer layer is then applied to $F_{i,1}$:

$$F_{i,2} = O_{stl_{i,2}}(F_{i,1}) \quad i \in \{1, 2\} \tag{26}$$

Next, a convolutional layer processes $F_{i,2}$:

$$F'_{i,2} = O_{conv}(F_{i,2}) \tag{27}$$

Finally, the output of the convolutional layer is scaled and adjusted by adding back the initial input Y, yielding the enhanced features from the QCFE module:

$$Z = 0.1 \cdot F'_{i,2} + Y \tag{28}$$

The integration of Quaternion convolutions with residual structure enables QCFE module to effectively capture complex inter-channel dependencies and subtle patterns in images. This is particularly crucial in image quality assessment, as issues in one channel (such as brightness or color distortion) can

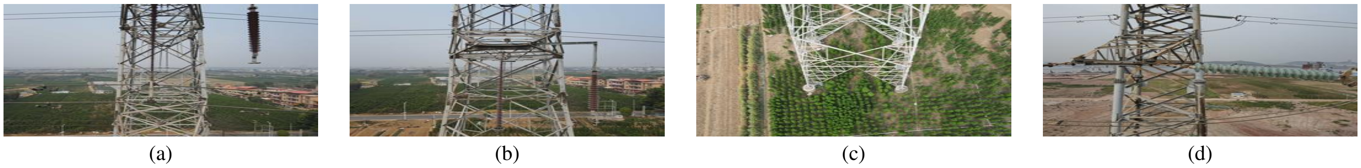

(a) (b) (c) (d)

Fig. 3. Some images of power towers at different heights and angles taken by drones.

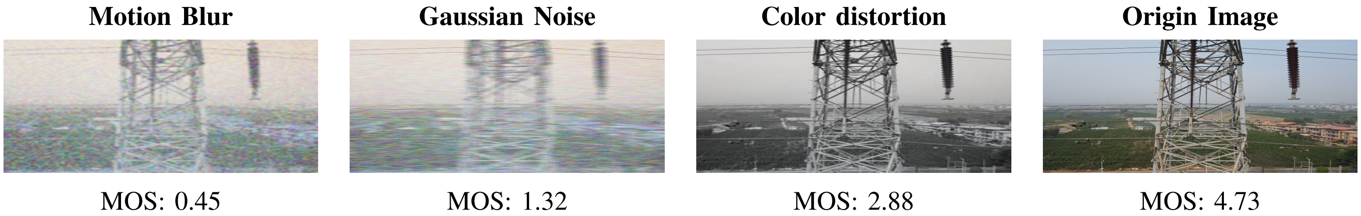

Motion Blur | Gaussian Noise | Color distortion | Origin Image

Fig. 4. Subjective MOS scores under different distortion types.

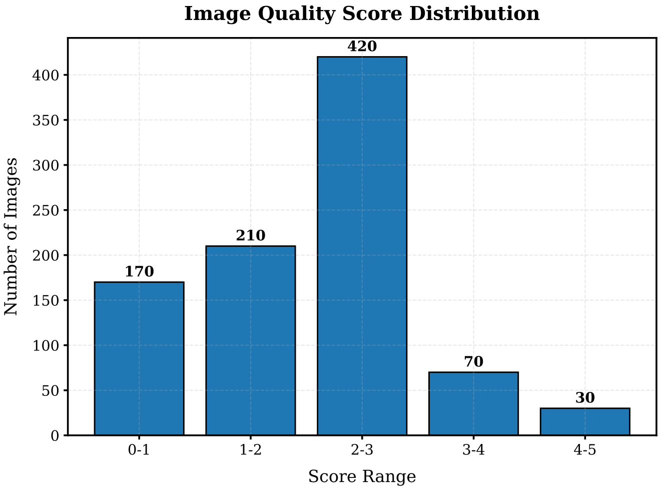


Fig. 5. MOS histograms of PowerGridIQ datasets.

significantly impact other channels and overall visual perception. For instance, noise or blur in the brightness channel may lead to unnatural color representation, thereby degrading the overall image quality. The QCFE module's unique architecture enhances the network's capability to extract nuanced features by modeling these channel dependencies more comprehensively. This results in improved generalization across diverse IQA datasets and better discrimination of fine-grained quality variations.

IV. PowerGridIQ Dataset

Data Collection The dataset is provided by a power grid organization and consists of high-resolution images specifically designed for image quality assessment in complex power grid environments. Skilled personnel utilized drones to capture 100 high-resolution images (8000 × 6000 pixels) of transmission line towers from various angles. These images (examples shown in Fig. 3) offer rich detail of the towers and their surroundings, which is essential for accurately evaluating image quality in real-world, high-complexity environments.

Dataset Description The PowerGridIQ dataset is the first IQA dataset dedicated to the power grid domain, simulating the real-world conditions encountered in operational power grid monitoring. The dataset includes nine common distortions selected in consultation with power grid staff, which reflect frequent image quality issues encountered in this field. These distortions include Gaussian noise, motion blur, color distortion, and both horizontal and vertical flips. Although these distortions are synthetic, they were carefully chosen to simulate real-world image degradation scenarios that can be observed in practical power grid monitoring. For example, the simulated Gaussian noise and motion blur are intended to represent the effects of environmental factors such as weather conditions, camera shake, and atmospheric interference. The flipping operations were applied to images already affected by noise, blur, or color changes to simulate the compounding effects of multiple distortions. This approach draws inspiration from Zhao [47], who demonstrated that flipping can significantly influence perceived image quality. Additionally, the real-world variability of drone-captured imagery—resulting from different angles, distances, and environmental factors further increases the complexity of image quality assessment, as the model must account for spatial inconsistencies across the images.

Challenges of PowerGridIQ Compared to Other Datasets PowerGridIQ introduces several novel challenges that distinguish it from traditional NR-IQA datasets. Unlike traditional IQA datasets focusing on synthetic distortions in controlled environments, PowerGridIQ captures realistic operational conditions by simulating common real-world distortions. The ultra-high resolution of the images (8000 × 6000 pixels) intensifies these distortions, demanding that models capture fine-grained details across large spatial areas. Additionally, the variability in image perspectives and angles inherent in drone-based capture introduces further complexity. Different sections of an image may exhibit varying degrees of blur, noise, or other distortions, which presents a significant challenge for models. Moreover, the intricate textures of transmission towers, cables, and pylons, along with high-frequency details typical of power grid infrastructure, require a model capable of dynamically handling both global context and local features for reliable image quality assessment.

Evaluation Process To validate the dataset, 900 distorted images were synthesized from the original drone images. Fifty domain experts with extensive experience in power grid scenarios were invited to assess the quality of these images. Each expert rated the images on a scale from 0 to 5, where 0 represented extremely poor quality and 5 indicated excellent quality (some examples are as shown in Fig. 4). The evaluation process was standardized: all images were reviewed on calibrated high-resolution monitors, at fixed viewing distances, and under neutral lighting conditions. Experts rated various attributes such as clarity, contrast, and distortion, resulting in highly consistent and reliable subjective assessments. The distribution of MOS data is shown in Fig. 5.

Score Aggregation and Reliability After collecting individual scores, we calculated the mean opinion score (MOS) for each image, which served as the ground truth for model training and evaluation. To ensure the reliability of these scores, we conducted consistency checks using metrics such as Cohen's kappa coefficient to identify and exclude any outliers, ensuring that the final MOS values were both robust and reflective of expert consensus. This rigorous evaluation process provides a solid foundation for training high-performance NR-IQA models, specifically designed for complex power grid environments.

V. Experiment

A. Datasets

To comprehensively evaluate our approach, we conducted experiments on seven publicly available datasets (detailed in Table I), and PowerGridIQ. Notably, our evaluation includes both traditional image quality datasets and AI-generated content (AIGC) datasets, addressing the growing need for synthetic media quality assessment. The synthetic distortion database was partitioned into training and test sets based on reference images to ensure robust evaluation. Detailed experimental configurations are presented below:

LIVE [48]: This benchmark dataset contains 29 reference and 779 distorted images with five distortion types: JPEG, JP2K, Gaussian noise, blur, and fast fading. DMOS scores (0–100, higher is worse) were collected using a single-stimulus methodology.

CSIQ [49]: CSIQ includes 30 reference and 866 distorted images across six distortion types, each at 4–5 severity levels. Images were quality-ranked horizontally by subjects; resulting DMOS scores are normalized to [0,1], where lower values indicate better quality.

TID2013 [50]: Comprising 3,000 images derived from 25 references, TID2013 features 24 distortion types at five levels. Rated via a double-stimulus method, it provides MOS scores (0–9), with higher values denoting better quality.

KADID-10K [51]: Consisting of 81 reference and 10,125 distorted images (25 distortion types × 5 levels), KADID-10K covers a broad spectrum of traditional and modern distortions, offering a rich dataset for comprehensive IQA evaluation.

KonIQ10K [52]: With 10,073 authentically distorted images selected from a large multimedia source, KonIQ10K reflects real-world content diversity. Quality scores were crowdsourced, making it well-suited for in-the-wild IQA tasks.

LIVE-C [53]: LIVE-C captures real images from mobile devices, affected by multiple authentic distortions such as blur, noise, overexposure, underexposure, and compression artifacts. Over 350,000 quality ratings from 8,100+ AMT users were collected using a single-stimulus slider from 1 to 100.

AGIQA-1K [54]: AGIQA-1K is the first IQA dataset dedicated to AI-generated images, containing 1,080 samples with MOS scores assessed across five quality dimensions: technical flaws, AI artifacts, unnatural appearance, content disparity, and overall aesthetics. It provides a valuable benchmark for evaluating the perceptual quality of outputs from generative models.

TABLE I
SUMMARY OF PUBLIC DATASETS USED IN OUR EXPERIMENTS.

Databases	Dist. Images	Dist. Types	Dataset Types
LIVE	799	5	synthetic
CSIQ	866	6	synthetic
TID2013	3,000	24	synthetic
KADID-10K	10,125	25	synthetic
KonIQ10K	10073	-	authentic
PowerGridIQ	900	9	synthetic
LIVE-C	1162	-	authentic
AGIQA-1K	1080	31	authentic

B. Implementation Details

Our experiments were conducted on an NVIDIA GeForce RTX 4090 using PyTorch 1.11.0 with CUDA 11.3. We employed a ViT [46] as the pre-trained model, with the patch size set to 8. The models were trained and evaluated on eight different datasets. During training, images were randomly cropped to a size of 224×224. Each dataset was split into training and testing sets in an 80:20 ratio. For databases containing synthetic distortions, the split was performed with respect to reference images to avoid semantic overlap between training and test sets. We utilized five distinct random seeds to enhance robustness. The model was trained using a learning rate of 1×10^{-5} and a batch size of 16, optimized with the Adam optimizer and a weight decay of 1×10^{-5}. A cosine annealing schedule was applied for learning rate adjustment. The Mean Squared Error (MSE) loss function was used to compute the training loss. During the testing phase, we randomly cropped the original images into 224×224 patches 20 times, and the final score was computed by averaging the predictions across these 20 crops. To ensure reliability, the results were further averaged over 10 different dataset splits. Each experiment was repeated five times with different random seeds, and the average performance metrics across all repetitions are reported. All comparison methods were evaluated under the same experimental setup.

C. Evaluation Criteria

For performance evaluation, we used two adopted criteria, namely Spearman's rank-order correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC). The specific details of these two metrics are given below:

TABLE II
COMPARISON OF THE PERFORMANCE OF OUR METHOD WITH OTHER METHODS. BOLD AND UNDERLINE ENTRIES ARE THE BEST AND SECOND-BEST PERFORMANCES, RESPECTIVELY.

Method	LIVE		CSIQ		TID2013		KADID-10K		KonIQ10K		PowerGridIQ	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
DIIVINE [27]	0.908	0.892	0.776	0.804	0.567	0.643	0.435	0.413	0.558	0.546	0.402	0.422
BRISQUE [25]	0.944	0.929	0.748	0.812	0.571	0.626	0.567	0.528	0.685	0.681	0.489	0.501
ILNIQE [24]	0.906	0.902	0.865	0.822	0.648	0.521	0.558	0.528	0.537	0.523	0.468	0.487
BIECON [19]	0.961	0.950	0.823	0.815	0.762	0.717	0.648	0.623	0.654	0.651	0.521	0.533
MEON [30]	0.955	0.951	0.864	0.852	0.824	0.808	0.691	0.604	0.628	0.611	0.524	0.579
WaDIQaM [55]	0.955	0.942	0.844	0.852	0.855	0.835	0.752	0.739	0.807	0.804	0.574	0.571
DBCNN [18]	0.971	0.952	0.959	0.946	0.865	0.816	0.856	0.851	0.884	0.875	0.588	0.591
TIQA [31]	0.965	0.947	0.838	0.825	0.858	0.846	0.855	0.850	0.903	0.892	0.691	0.708
MetaIQA [56]	0.959	0.954	0.908	0.899	0.868	0.856	0.775	0.762	0.856	0.887	0.623	0.602
P2P-BM [57]	0.958	0.949	0.902	0.899	0.856	0.862	0.849	0.840	0.885	0.872	0.654	0.681
HyperIQA [12]	0.966	0.949	0.942	0.923	0.858	0.840	0.845	0.852	0.917	0.906	0.665	0.693
TReS [14]	0.968	0.944	0.942	0.922	0.883	0.863	0.858	0.915	0.928	0.915	0.625	0.652
SaTQA [32]	0.983	0.946	0.974	0.965	0.948	0.938	0.929	0.926	0.941	0.930	0.725	0.683
MANIQA [22]	0.987	0.953	0.972	0.966	0.946	0.940	0.939	0.936	0.945	0.925	0.855	0.840
TempQT [58]	0.982	0.947	0.960	0.950	0.908	0.892	0.887	0.889	0.930	0.921	0.816	0.793
$DiffV^2IQA$ [11]	0.978	**0.955**	0.982	0.975	0.954	0.946	0.930	0.927	0.936	0.917	0.862	0.835
AGAIQA [59]	0.973	0.909	0.978	0.973	0.948	0.944	0.941	0.935	0.939	0.936	0.859	0.836
AKD-IQA [60]	0.968	0.942	0.959	0.952	0.944	0.942	0.928	0.913	0.904	0.896	0.822	0.801
VISGA [61]	0.987	0.951	0.971	0.960	0.914	0.901	0.940	0.932	0.940	0.931	0.850	0.832
WAQNIQA (Ours)	**0.990**	**0.955**	**0.991**	**0.988**	**0.955**	**0.952**	**0.943**	**0.939**	**0.953**	**0.947**	**0.877**	**0.844**

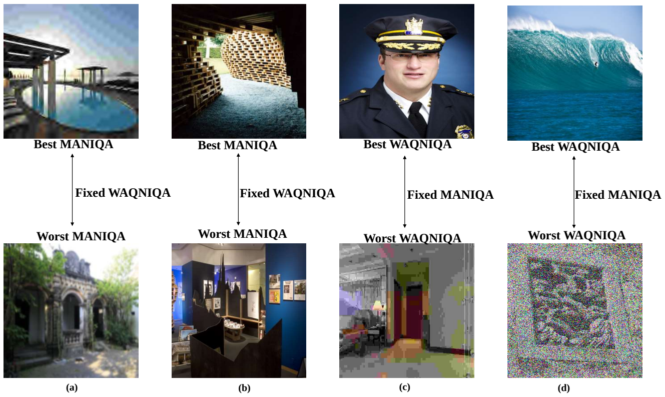


Fig. 6. Results of the gMAD competition [67] between WAQNIQA and MANIQA. (a) and (b) show the performance of WAQNIQA fixed at low-quality and high-quality levels, respectively. Meanwhile, (c) and (d) illustrate the performance of MANIQA fixed at low-quality and high-quality levels, respectively.

$$\mathrm{PLCC} = \frac{\sum_{j=1}^{N}(S_j-\mu_{S_j})(\hat{S}_j-\mu_{\hat{S}_j})}{\sqrt{\sum_{j=1}^{N}(S_j-\mu_{S_j})^2}\sqrt{\sum_{j=1}^{N}(\hat{S}_j-\mu_{\hat{S}_j})^2}} \tag{29}$$

where S_j and $\hat{S}_j$ denote the ground truth and predicted quality scores of the i-th image, respectively, and μ_{S_j} and $\mu_{\hat{S}_j}$ denote their averages. N denotes the tested image.

The SROCC can be defined as:

$$\mathrm{SROCC} = 1 - \frac{6\sum_{i=1}^{N}x_i^2}{N(N^2-1)} \tag{30}$$

where x_i denotes the difference between the grades in ground-truth and predicted quality scores for the i-th test image.

These two criteria range from 0 to 1, with higher values indicating better performance.

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED WAQNIQA VERSUS STATE-OF-THE-ART MODELS ON THE LIVE-C DATASET.

Method	LIVE-C	
	PLCC	SROCC
DIIVINE [27]	0.591	0.588
BRISQUE [25]	0.629	0.629
ILNIQE [24]	0.508	0.508
BIECON [19]	0.613	0.613
MEON [30]	0.710	0.697
WaDiQaM [55]	0.739	0.671
DBCNN [18]	0.869	0.869
TIQA [31]	0.861	0.845
Meta-IQA [56]	0.802	0.835
P2P-BM [57]	0.842	0.844
HyperIQA [12]	0.872	0.872
TRes [14]	0.877	0.846
SaTQA [32]	0.903	0.877
TempQT [58]	0.886	0.870
MANIQA [22]	0.887	0.871
$DiffV^2IQA$ [11]	0.897	0.882
AGAIQA [59]	0.904	0.892
AKD-IQA [60]	0.881	0.850
VISGA [61]	0.893	0.882
WAQNIQA (Ours)	**0.905**	**0.894**

TABLE IV
QUANTITATIVE COMPARISON ON THE AGIQA-1K DATASET. THE BEST RESULTS ARE BOLDED.

Method	AGIQA-1K	
	PLCC	SROCC
ResNet50 [62]	0.653	0.636
StairIQA [63]	0.608	0.530
MGQA [64]	0.676	0.601
IPCE [65]	0.879	0.853
MANIQA [22]	0.674	0.632
$DiffV^2IQA$ [11]	0.693	0.679
T2I [66]	**0.889**	**0.855**
WAQNIQA (Ours)	0.716	0.692

TABLE V
RESULTS OF CROSS-DATASET VALIDATION.

Train on	KonIQ10K		LIVE-C		LIVE		LIVE	
Test on	LIVE-C		KONIQ10K		CSIQ		TID2013	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
WaDIQaM [55]	0.682	0.676	0.711	0.701	0.704	0.691	0.462	0.451
P2P-BM [57]	0.770	0.759	0.740	0.735	0.712	0.702	0.488	0.472
HyperIQA [12]	0.785	0.762	0.772	0.761	0.744	0.731	0.551	0.544
TReS [14]	0.786	0.775	0.733	0.725	0.761	0.753	0.562	0.553
SaTQA [32]	0.791	0.785	0.788	0.782	0.831	0.824	0.617	0.608
MANIQA [22]	0.773	0.769	0.759	0.748	0.813	0.806	0.581	0.570
TempQT [58]	0.794	0.789	0.766	0.750	0.816	0.792	0.586	0.575
$DiffV^2IQA$ [11]	0.782	0.766	0.778	0.767	0.833	0.821	0.602	0.581
AGAIQA [59]	0.788	0.765	0.789	0.778	0.840	0.826	0.612	0.594
AKD-IQA [60]	0.762	0.755	0.726	0.718	0.692	0.685	0.538	0.531
VISGA [61]	0.796	0.782	0.785	0.779	0.772	0.766	0.606	0.588
WAQNIQA (Ours)	**0.801**	**0.791**	**0.792**	**0.788**	**0.852**	**0.830**	**0.620**	**0.612**

D. Quantitative Evaluation

1) Comparison with general-purpose datasets: As evidenced in Tables II and III, our proposed method, WAQNIQA, consistently outperforms state-of-the-art approaches across multiple datasets, achieving superior PLCC and SROCC values. This exceptional performance can be attributed primarily to the complementary strengths of the WIAA and QCFE modules. The WIAA module ingeniously integrates wavelet transforms with adaptive attention mechanisms, enabling robust multi-scale analysis and significantly enhancing the interaction between local and global features. While the wavelet transform precisely captures detailed spatial and frequency information, the attention mechanism emphasizes key features, jointly improving prediction accuracy. Concurrently, the QCFE module employs Quaternion representation learning to model image channels as a unified entity. This innovative approach not only effectively captures inter-channel dependencies but also generates richer feature representations. The synergistic effect of these modules empowers WAQNIQA to generalize effectively across diverse NR-IQA scenarios, consistently achieving superior performance on all evaluated datasets. Our results highlight WAQNIQA's exceptional capability in handling various types of image distortions, from synthetic noise to complex textures, demonstrating its robustness and versatility in diverse and challenging environments. While WAQNIQA achieves state-of-the-art performance on standard synthetic IQA datasets such as LIVE and CSIQ, its performance on the PowerGridIQ dataset is relatively less pronounced. This is likely due to the dataset's inherent complexity, characterized by high-resolution imagery, intricate textures, and diverse environmental conditions. Future improvements could include designing domain-specific pre-training strategies or integrating additional context-aware features to better handle such real-world complexities.

2) Comparison with AIGC datasets: To evaluate WAQNIQA's generalization capability, we compared it against state-of-the-art AI-Generated Content Image Quality Assessment (AIGC-IQA) models. Unlike traditional IQA methods that focus primarily on visual fidelity, AIGC-IQA models also assess semantic alignment between input text and generated images. As shown in Table IV, WAQNIQA demonstrates competitive performance on the AGIQA-1K dataset, achieving PLCC of 0.716 and SROCC of 0.692. Notably, WAQNIQA was not explicitly trained on text-to-image semantic alignment, yet it outperforms several AIGC-specialized models like MGQA [64] and shows significant improvement over general models such as ResNet50 and MANIQA. This strong performance without semantic alignment training demonstrates the robustness and generalization potential of our method. While specialized AIGC models like T2I [66] and IPCE [65] achieve superior performance due to their targeted design, WAQNIQA's competitive standing highlights its effectiveness as a general-purpose IQA model capable of evaluating both traditional and AI-generated content.

E. Cross-dataset Evaluations

To rigorously assess the generalization ability of the proposed WAQNIQA model, we conducted comprehensive cross-dataset validation experiments. In this evaluation, the model was trained on the KonIQ-10K, LIVE-C, and LIVE datasets and tested on LIVE-C, KonIQ-10K, CSIQ, and TID2013 without any fine-tuning or parameter adaptation, ensuring an unbiased assessment of its out-of-domain performance. We compared WAQNIQA against the best-performing deep learning–based NR-IQA models, including WaDIQaM, P2P-BM, HyperIQA, TReS, SaTQA, MANIQA, TempQT, $DiffV^2IQA$, , AGAIQA, AKD-IQA and VISGA. As shown in Table V, WAQNIQA consistently achieves the highest performance across all test datasets in terms of both PLCC and SROCC. Notably, it outperforms the strongest baseline model by a clear margin, demonstrating superior robustness and adaptability to diverse distortion types and image content. These results confirm that WAQNIQA exhibits significantly stronger generalization capability than existing methods under strict cross-dataset evaluation protocols.

F. The Measurement of Model Stability

To evaluate the robustness of the models, we conducted a group Maximum Differentiation (gMAD) competition [67] using WAQNIQA and MANIQA on the Waterloo Explore

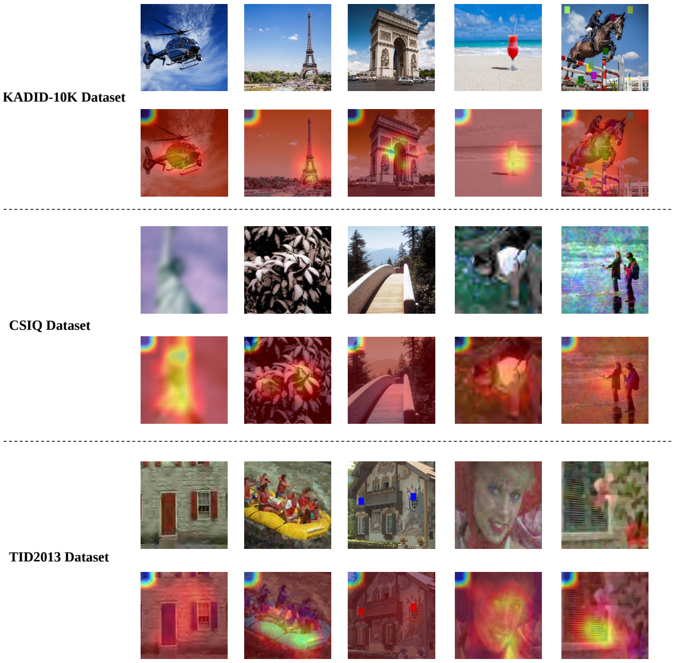


Fig. 7. Grad-CAM activation maps of WAQNIQA on KADID-10K, CSIQ and TID2013 dataset.

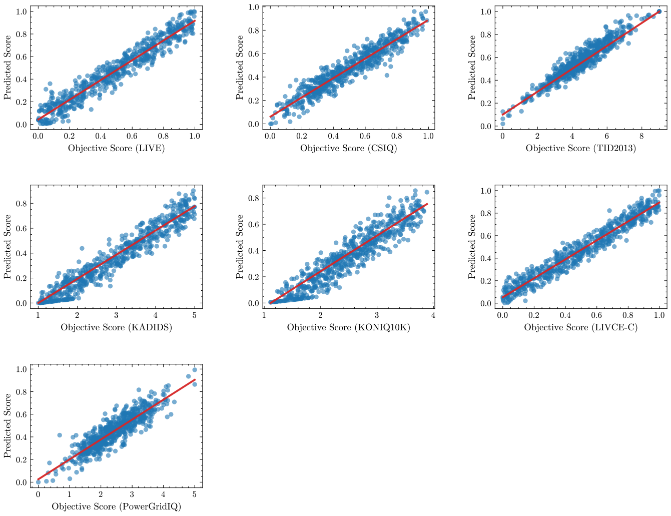


Fig. 8. Visualization of our model's prediction results on seven datasets using scatter plots.

database [68]. The Waterloo Exploration Database is a large Image Quality Assessment (IQA) database of 4,744 original natural images and 94,880 distorted images generated from them, designed to address the limited diversity of content in existing IQA databases and to better test the generalization capabilities of IQA models. The gMAD approach is grounded in the concept of attempting to disprove a model rather than merely validating its effectiveness; models that are difficult to disprove are deemed relatively superior. In the gMAD contests, one participant adopts the offensive role while the other assumes a defensive stance. When the defending party believes the images are of equal quality, the attacker systematically seeks the maximum quality difference between image pairs. If a human observer can easily differentiate between the qualities of the pairs, the attacker wins; conversely, if the pairs are perceived to have the same quality, the defender wins. Fig. 6 illustrates the results of this competition. When WAQNIQA acts as the defender, MANIQA struggles to identify image pairs with clear quality differences. In contrast, when WAQNIQA takes on the role of the attacker, it effectively reveals MANIQA's deficiencies by uncovering pairs with varying quality levels. This demonstrates that the proposed method is more robust than MANIQA.

G. Visualization

1) Visualization of Scatter Plot: We evaluated the performance of our model on seven datasets by plotting scatter diagrams of the predicted scores against the corresponding ground truth scores (MOS or DMOS), as illustrated in Figure Fig. 8. A linear regression line was fitted to each plot to capture the underlying relationship. The approximately homoscedastic distribution of data points along the regression line suggests low residual variance, indicating a statistically significant and consistent linear correlation between the predicted and reference scores.

2) Visualization of Grad-CAM: To evaluate the effectiveness of our proposed WAQNIQA model in capturing perceptually relevant features, we used Grad-CAM visualization on distorted images from three widely used datasets. KADID-10K, CSIQ and TID2013. Figure 7 presents the Grad-CAM activation maps, where brighter regions indicate areas of higher attention by the model. The distorted images are shown at the top, with their corresponding activation maps below. Grad-CAM visualizations reveal that WAQNIQA consistently directs its attention to perceptually relevant regions, accurately identifying distorted areas while disregarding irrelevant details. This alignment with human visual perception further validates the model's capacity for nuanced quality assessment.

H. Ablation Study

To validate the effectiveness of our proposed WIAA module and QCFE module, we conducted comprehensive ablation experiments. Our results, as presented in Table VI, demonstrate the significant contributions of both modules to the overall performance of our WAQNIQA model.

Effectiveness of WIAA The WIAA module plays a critical role in enhancing the model's ability to capture complex textures and structural details across diverse datasets. By integrating wavelet transforms with adaptive attention mechanisms, the WIAA module enables the model to conduct multi-scale analysis, capturing both local and global features in spatial and frequency domains. This allows for more accurate processing of different types of distortions, from synthetic artifacts to natural image degradations. As shown in Table VI, adding the WIAA module consistently improves the model's performance across datasets: for example, on the TID2013 dataset, PLCC increases from 0.946 to 0.955, and SROCC rises from 0.940 to 0.952. On the KonIQ10K dataset, PLCC

TABLE VI
ABLATION STUDY ON EACH MODULE.

		LIVE		CSIQ		TID2013		KADID-10K		KonIQ10K		PowerGridIQ		LIVE-C	
WIAA	QCFE	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
✗	✗	0.987	0.953	0.972	0.966	0.946	0.940	0.939	0.936	0.945	0.925	0.855	0.840	0.887	0.871
✓	✗	0.988	0.954	0.978	0.968	0.950	0.945	0.941	0.935	0.951	0.938	0.864	0.842	0.891	0.875
✗	✓	0.986	0.954	0.980	0.976	0.947	0.942	0.942	0.937	0.948	0.930	0.859	0.843	0.894	0.880
✓	✓	**0.990**	**0.955**	**0.991**	**0.988**	**0.955**	**0.952**	**0.943**	**0.939**	**0.953**	**0.947**	**0.877**	**0.844**	**0.905**	**0.894**

improves from 0.945 to 0.951, and SROCC goes up from 0.925 to 0.938. Similarly, for the PowerGridIQ dataset, the inclusion of WIAA results in a performance boost, with PLCC rising from 0.855 to 0.864, and SROCC improving from 0.840 to 0.842. These improvements highlight WIAA's ability to adapt to different types of image quality challenges, making it especially effective for tasks requiring detailed analysis of texture and structure.

Effectiveness of QCFE The QCFE module provides a critical layer of enhancement by leveraging Quaternion representation learning. Unlike standard scalar-based convolutions that treat channel aggregation as simple linear summation, the QCFE module utilizes the Hamilton product to enforce a rigorous, coupled interaction among feature channels. This mechanism ensures that the feature components (representing Real and Imaginary parts) actively modulate each other, thereby transforming loose feature collections into a structurally unified entity. This allows the model to capture nonlinear cross-channel dependencies and subtle phase coherence, which are essential for identifying complex distortions like chromatic aberrations and mixed noise.

Quantitative results validate this theoretical advantage. As shown in Table VI, incorporating the QCFE module consistently boosts performance across all datasets. On TID2013, for instance, adding QCFE increases PLCC from 0.950 to 0.955 and improves SROCC from 0.945 to 0.952. On KonIQ10K, the improvements are even more pronounced, with PLCC rising from 0.951 to 0.953 and SROCC jumping from 0.938 to 0.947. For PowerGridIQ, which presents a variety of real-world image distortions, QCFE offers significant gains, raising PLCC from 0.864 to 0.877 and improving SROCC from 0.842 to 0.844. The specific effectiveness on authentically distorted datasets like KonIQ10K and PowerGridIQ demonstrates QCFE's strong capability in disentangling complex, inter-dependent quality variations that are often overlooked by traditional feature enhancement techniques.

TABLE VII
EVALUATION OF TRAINING TIME, INFERENCE TIME, PARAMETER COUNTS, AND FLOPS FOR DIFFERENT MODELS.

Model	Training Time (per epoch)	Inference Time	Params	FLOPs
HyperIQA [12]	18 mins	98.31 ms	27.37M	54.73G
TReS [14]	40 mins	330.45 ms	152.45M	199.64G
SaTQA [32]	28 mins	142.38 ms	84.21M	89.40G
MANIQA [22]	34 mins	309.66 ms	135.76M	138.50G
AGAIQA [59]	37 mins	300.95 ms	148.77M	160.52G
WAQNIQA (ours)	32 mins	267.72 ms	133.75M	108.13G

I. Discussion and Limitation

Runtime efficiency and computational complexity are critical for practical IQA applications. As shown in Table VII, WAQNIQA achieves superior performance on the KADID-10k dataset compared to lightweight baselines like HyperIQA, while maintaining moderate computational demands. With 133.75M parameters, it is more compact than MANIQA (135.76M), TReS (152.45M), and AGAIQA (148.77M), and its FLOPs (108.13G) remain significantly lower than most high-performing competitors. Although HyperIQA is more efficient in training and inference (267.72 ms vs. faster), WAQNIQA delivers higher accuracy, making it better suited for scenarios where precision is prioritized over extreme speed. This balance stems from WAQNIQA's effective use of diverse image features for quality assessment. In practice, model selection should weigh application constraints: HyperIQA excels in resource-limited settings, whereas WAQNIQA offers a favorable accuracy-efficiency trade-off for tasks demanding robust performance.

VI. CONCLUSION

WAQNIQA is a novel No-Reference Image Quality Assessment (NR-IQA) network with two core modules: the Wavelet-Infused Adaptive Attention (WIAA) and the Quaternion Collaborative Feature Enhancement (QCFE). WIAA integrates wavelet transforms with attention for robust multi-scale spatial and frequency-domain analysis. QCFE uses Quaternion representation to holistically model image channels, enhancing inter-channel dependencies. We introduce PowerGridIQ, the first IQA dataset for the power grid domain, featuring nine common distortions. WAQNIQA outperforms state-of-the-art methods on PowerGridIQ, public datasets, and AIGC datasets. Notably, WAQNIQA performs competitively against specialized AIGC-IQA models, demonstrating its general-purpose effectiveness. Future work will focus on improving WAQNIQA's efficiency via advanced architectures, model compression, and self-supervised learning to reduce dependence on labeled data.

REFERENCES

[1] A. Alsaeedi and M. Zubair, "A study on sentiment analysis techniques of twitter data," *International Journal of Advanced Computer Science and Applications*, 2023.

[2] T.-Y. Chiu, Y. Zhao, and D. Gurari, "Assessing image quality issues for real-world problems," *Cornell University - arXiv,Cornell University - arXiv*, Mar 2020.

[3] T.-C. Chang, S.-D. Xu, and S. Su, "Ssim-based quality-on-demand energy-saving schemes for oled displays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, pp. 623–635, 2016.

[4] C. Deng, S. Wang, Z. Li, G. Huang, and W. Lin, “Content-insensitive blind image blurriness assessment using weibull statistics and sparse extreme learning machine,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, pp. 516–527, 2019.

[5] F. Shao, Z. Fu, Q. Jiang, G. Jiang, and Y.-S. Ho, “Transformation-aware similarity measurement for image retargeting quality assessment via bidirectional rewarping,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, pp. 3053–3067, 2021.

[6] F. Shao, Y. Gao, F. Li, and G. Jiang, “Toward a blind quality predictor for screen content images,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 1521–1530, 2018.

[7] K. Y. Chan, H. K. Lam, C. K. fai Yiu, and T. S. Dillon, “A flexible fuzzy regression method for addressing nonlinear uncertainty on aesthetic quality assessments,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 2363–2377, 2017.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[9] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, pp. 2378–2386, 2011.

[10] J. Wu, Y. Liu, G. Shi, and W. Lin, “Saliency change based reduced reference image quality assessment,” *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, 2017.

[11] Z. Wang, B. Hu, M. Zhang, J. Li, L. Li, M. Gong, and X. Gao, “Diffusion model-based visual compensation guidance and visual difference analysis for no-reference image quality assessment,” *IEEE Transactions on Image Processing*, vol. 34, pp. 263–278, 2025.

[12] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, “Blindly assess image quality in the wild guided by a self-adaptive hyper network,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020.

[13] D. Ghadiyaram and A. C. Bovik, “Perceptual quality prediction on authentically distorted images using a bag of features approach,” *Journal of Vision*, vol. 17, 2016.

[14] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, “No-reference image quality assessment via transformers, relative ranking, and self-consistency,” *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3989–3999, 2021.

[15] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE Transactions on Image Processing*, vol. 20, pp. 3350–3364, 2011.

[16] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, “dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs,” *IEEE Transactions on Image Processing*, vol. 26, pp. 3951–3964, 2017.

[17] P. Ye, J. Kumar, L. Kang, and D. S. Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1098–1105, 2012.

[18] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind image quality assessment using a deep bilinear convolutional neural network,” *IEEE Transactions on Circuits and Systems for Video Technology*, p. 36–47, Jan 2020.

[19] J. Kim and S. Lee, “Fully deep blind image quality predictor,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 206–220, 2017.

[20] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Neural Information Processing Systems*, 2017.

[21] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, “Musiq: Multi-scale image quality transformer,” *International Conference on Computer Vision,International Conference on Computer Vision*, Aug 2021.

[22] S. Yang, T. Wu, S. Shi, S. Gong, M. Cao, J. Wang, and Y. Yang, “Maniqa: Multi-dimension attention network for no-reference image quality assessment,” *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1190–1199, 2022.

[23] Y. Zhang, X. Min, and G. Zhai, “Split-conv: A resource-efficient compression method for image quality assessment models,” *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, 2023.

[24] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” *IEEE Transactions on Image Processing*, vol. 24, pp. 2579–2591, 2015.

[25] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, 2012.

[26] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Processing Letters*, vol. 20, pp. 209–212, 2013.

[27] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the dct domain,” *IEEE Transactions on Image Processing*, vol. 21, pp. 3339–3352, 2012.

[28] Z. Tu, X. Yu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. Bovik, “Rapique: Rapid and accurate video quality prediction of user generated content,” *Cornell University - arXiv,Cornell University - arXiv*, Jan 2021.

[29] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, “On the use of deep learning for blind image quality assessment,” *Signal, Image and Video Processing*, p. 355–362, Feb 2018.

[30] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, “End-to-end blind image quality assessment using deep neural networks,” *IEEE Transactions on Image Processing*, vol. 27, pp. 1202–1213, 2018.

[31] J. You and J. Korhonen, “Transformer for image quality assessment,” in *2021 IEEE International Conference on Image Processing (ICIP)*, Sep 2021.

[32] J. Shi, P. Gao, and J. Qin, “Transformer-based no-reference image quality assessment via supervised contrastive learning,” *ArXiv*, vol. abs/2312.06995, 2023.

[33] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, “Deep wavelet prediction for image super-resolution,” *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1100–1109, 2017.

[34] W. Zou, L. Chen, Y. Wu, Y. Zhang, Y. Xu, and J. Shao, “Joint wavelet sub-bands guided network for single image super-resolution,” *IEEE Transactions on Multimedia*, vol. 25, pp. 4623–4637, 2023.

[35] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, “Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun 2021.

[36] Z. Li, X. Chen, C.-M. Pun, and S. Wang, “Wavenhancer: Unifying wavelet and transformer for image enhancement,” *ArXiv*, vol. abs/2212.08327, 2022.

[37] Y.-J. Choi, Y.-W. Lee, and B.-G. Kim, “Wavelet attention embedding networks for video super-resolution,” in *2020 25th International Conference on Pattern Recognition (ICPR)*, Jan 2021.

[38] L. Dai, L. Xiao-hong, C. Li, and J. Chen, “Awnet: Attentive wavelet network for image isp,” *Cornell University - arXiv,Cornell University - arXiv*, Aug 2020.

[39] H. Salman, C. Parks, S. Hong, and J. Z. Zhan, “Wavenets: Wavelet channel attention networks,” *2022 IEEE International Conference on Big Data (Big Data)*, pp. 1107–1113, 2022.

[40] W. R. S. Hamilton, “Xi. on quaternions; or on a new system of imaginaries in algebra,” *Philosophical Magazine Series 1*, vol. 33, pp. 58–60, 1848.

[41] M. R. Celsi, S. Scardapane, and D. Comminiello, “Quaternion neural networks for 3d sound source localization in reverberant environments,” *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2020.

[42] Z. Zheng, G. Huang, X. Yuan, C.-M. Pun, H. Liu, and W.-K. Ling, “Quaternion-valued correlation learning for few-shot semantic segmentation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, pp. 2102–2115, 2023.

[43] D. Pavllo, D. Grangier, and M. Auli, “Quaternet: A quaternion-based recurrent model for human motion,” *ArXiv*, vol. abs/1805.06485, 2018.

[44] Z. Zhou, Y. Huo, G. Huang, A. Zeng, X. Chen, L. Huang, and Z. Li, “Qean: quaternion-enhanced attention network for visual dance generation,” *The Visual Computer*, Apr 2024.

[45] Y. Tay, A. Zhang, A. T. Luu, J. Rao, S. Zhang, S. Wang, J. Fu, and S. C. Hui, “Lightweight and efficient neural natural language processing with quaternion networks,” in *Annual Meeting of the Association for Computational Linguistics*, 2019.

[46] L. Yuan, Y. Chen, T. Wang, W. Yu, Y. Shi, F. E. H. Tay, J. Feng, and S. Yan, “Tokens-to-token vit: Training vision transformers from scratch on imagenet,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 538–547, 2021.

[47] K. Zhao, K. Yuan, M.-T. Sun, M. Li, and X. Wen, “Quality-aware pretrained models for blind image quality assessment,” *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22302–22313, 2023.

[48] H. Sheikh, M. Sabir, and A. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, p. 3440–3451, Nov 2006.

[49] D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, p. 011006, Jan 2010.

[50] N. N. Ponomarenko, L. Jin, O. Ieremeiev, V. V. Lukin, K. O. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database tid2013: Peculiarities, results and perspectives," *Signal Process. Image Commun.*, vol. 30, pp. 57–77, 2015.

[51] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, Jun 2019.

[52] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, p. 4041–4056, Jan 2020.

[53] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, pp. 372–387, 2015.

[54] Z. Zhang, C. Li, W. Sun, X. Liu, X. Min, and G. Zhai, "A perceptual quality assessment exploration for aigc images," *2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 440–445, 2023.

[55] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, pp. 206–219, 2016.

[56] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "Metaiqa: Deep meta-learning for no-reference image quality assessment," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14131–14140, 2020.

[57] Z. Ying, H. Niu, P. Gupta, D. K. Mahajan, D. Ghadiyaram, and A. C. Bovik, "From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3572–3582, 2019.

[58] J. Shi, P. Gao, and A. Smolic, "Blind image quality assessment via transformer predicted error map and perceptual quality token," *IEEE Transactions on Multimedia*, vol. 26, pp. 4641–4651, 2024.

[59] H. Wang, J. Liu, H. Tan, J. Lou, X. Liu, W. Zhou, and H. Liu, "Blind image quality assessment via adaptive graph attention," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, pp. 10299–10309, 2024.

[60] B. Hu, W. Chen, J. Zheng, L. Li, W. Lu, and X. Gao, "No-reference image quality assessment via inter-level adaptive knowledge distillation," *IEEE Transactions on Broadcasting*, vol. 71, pp. 581–592, 2025.

[61] H. Shi, W. Xie, H. Qin, Y. Li, and L. Fang, "Visual state space model with graph-based feature aggregation for no-reference image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, pp. 5589–5601, 2025.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[63] W. Sun, H. Duan, X. Min, L. Chen, and G. Zhai, "Blind quality assessment for in-the-wild images via hierarchical feature fusion strategy," *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 01–06, 2022.

[64] T. Wang, W. Sun, X. Min, W. Lu, Z. Zhang, and G. Zhai, "A multi-dimensional aesthetic quality assessment model for mobile game images," *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, 2021.

[65] F. Peng, H. Fu, A. Ming, C. Wang, H. Ma, S. He, Z. Dou, and S. Chen, "Aigc image quality assessment via image-prompt correspondence," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 6432–6441, 2024.

[66] J. Xia, L. He, B. Hu, B. Han, and X. Gao, "A two-stage aigc image quality assessment with t2i correspondence and visual perception," *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[67] K. Ma, Q. Wu, Z. Wang, Z. Duanmu, H. Yong, H. Li, and L. Zhang, "Group mad competition? a new methodology to compare objective image quality models," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1664–1673, 2016.

[68] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, pp. 1004–1016, 2017.