ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

No-Reference Image Quality Assessment Based on Human Visual System and Dual-Branch Multi-Level Residual Network

Xianwei Qiu[1,a,*], Jun Li[1,b], Wuyang Shan[1,c], Wuyang Fan[2,d]

[1]College of Computer Science and Cybersecurity, Chengdu University of Technology, Chengdu, China
[2]College of Architecture and Civil Engineering, Chengdu University, Chengdu, China
[a]870631008@qq.com, [b]junli@cdut.edu.cn, [c]shanwuyang@cdut.edu.cn, [d]fzhf213@163.com
**Corresponding author*

Abstract: *This paper presents a No-Reference Image Quality Assessment (NR-IQA) method inspired by human visual perception. It integrates a dual-branch multi-level residual network that combines two complementary streams: one processing HSV images to capture content features aligned with human perception, and another utilizing contrast-sensitive weighted gradient (CSG) images to extract structural and texture features. This dual-branch architecture provides a comprehensive assessment of image quality from both content and structural perspectives. An advanced feature fusion strategy is introduced, employing a dedicated weight module to assign varying importance to content and structural features, ensuring an accurate final quality score. Experimental results on benchmark datasets, including LIVE, CSIQ, TID2013, LIVEC, and KonIQ-10k, showcase the superiority of our method over existing state-of-the-art NR-IQA techniques in key metrics like PLCC and SROCC. Our work holds significant practical value in advancing image quality assessment and has potential applications in multimedia compression, image restoration, enhancement, and related image processing technologies.*

Keywords: *NR-IQA, Dual-Branch Network, Feature Fusion Strategy*

1. Introduction

Images are crucial for information acquisition, but distortion during acquisition, transmission, processing, and storage can hinder information extraction. In Image Quality Assessment (IQA), methodologies are divided into subjective and objective IQA. Subjective IQA relies on human observation and scoring, suitable for small image sets but inefficient for large datasets[1]. Objective IQA uses algorithms to mimic human perception, overcoming manual scoring limitations and boosting efficiency for large-scale evaluations. Objective IQA is further categorized into Full-Reference (FR-IQA), Reduced-Reference (RR-IQA), and No-Reference (NR-IQA) methods. NR-IQA, which doesn't require reference images, holds greater value and practical significance due to challenges in acquiring high-quality references.

Traditional NR-IQA (No-Reference Image Quality Assessment) methodologies rely on straightforward regression models to convert handcrafted low-level attributes, including Natural Scene Statistics (NSS) attributes, into numerical assessments of image quality[2][3][4][5][6]. However, these manually extracted features have inherent limitations. Specifically, the manual feature design process often hinges on researchers' experiences and intuitions, potentially leading to the oversight of comprehensive factors that impact image quality. The inherent subjectivity in this design approach can result in discrepancies and unpredictability in evaluation outcomes. Furthermore, manually devised features are frequently tailored for specific distortion types or image contents, potentially compromising their effectiveness when faced with diverse distortion types or image contents.

With the rapid advancements in deep learning, numerous Blind Image Quality Assessment (BIQA) methods leveraging this technology have emerged, such as IQA-CNN[7], DIQaM-NR[8], DIQA[9], HyperIQA[10], DB-CNN[11], and TS-CNN[12]. These methods possess formidable learning capabilities, enabling them to automatically extract high-level features from distorted images. Compared to traditional methods, deep learning-based approaches offer an end-to-end training paradigm, facilitating a more efficient mapping of extracted features to image quality scores.

While current deep learning-driven IQA methods strive to enhance their performance through

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

optimizations of network architectures or the integration of additional network modules to bolster feature extraction capabilities, they tend to overlook the pivotal influence and potential guiding role of Human Visual System (HVS) characteristics. Notably, the contrast sensitivity feature of the HVS highlights the varying sensitivity of the human eye to different spatial frequencies[13], a factor that warrants further consideration in the development of BIQA methods. Campbell et al[14]proposed a contrast sensitivity function to explicitly calculate the sensitivity of the HVS to different spatial frequencies. Several traditional IQA methods[15][16] have employed the contrast sensitivity function to assign weights to extracted features, aiming to achieve superior performance.

The goal of Image Quality Assessment (IQA) is to measure image distortion based on human visual perception. This research explores the integration of Human Visual System (HVS) characteristics with deep learning techniques. Although existing methods have shown progress, there's potential for deeper exploration. This article introduces HVPIQA, a no-reference IQA method based on human visual perception and a dual-branch multi-level residual network. Contributions include:

(1)Converting RGB images to HSV and CSG formats to extract content features from HSV and texture/structural features from CSG.

(2)Introducing a dual-branch network to extract multi-level features from both HSV and CSG images, which are then fused and mapped to a quality score.

(3)Incorporating a weighting mechanism to prioritize significant features from each branch.

(4)Demonstrating superior performance through extensive experimentation and comparisons with state-of-the-art methods across five databases.

The structure of the rest of this document is as follows: Section 2 reviews traditional, handcrafted feature-based IQA methods and those based on deep learning. Section 3 elaborates on the components of the HVPIQA methodology discussed in this paper. Section 4 presents a series of experiments validating the improvements of our proposed HVPIQA. Finally, Section 5 summarizes the key findings and conclusions.

2. Related works

2.1. NR-IQA Based on Hand-crafted Features

Natural Scene Statistics (NSS) models are widely used for reliable feature extraction in Image Quality Assessment (IQA). NSS assumes that natural images occupy a constrained subspace defined by specific statistical properties[17]. Distortions affect these properties, and capturing deviations from them in distorted images indicates image quality. Researchers used the Generalized Gaussian Distribution (GGD) to model wavelet decomposition coefficients, identifying distortion types and predicting quality scores. They further developed DIIVINE[2] for a comprehensive portrayal of scene statistics. Saad et al[3] applied the Discrete Cosine Transform (DCT) to derive contrast and structural features, correlating them with quality scores through a probabilistic model. However, these transform-domain methods are computationally intensive.

To avoid image transformation, researchers proposed methods for direct NSS feature extraction in the spatial domain. Mittal et al[5] introduced BRISQUE, an NR-IQA method using locally normalized luminance coefficients to quantify naturalness loss. They also presented NIQE[6], a simple yet effective spatial domain NSS model comparing pristine and distorted image features. Yue et al[18] proposed an NR-IQA method for contrast-distorted images, combining entropy information and Kullback-Leibler divergence to derive quality scores.

2.2. NR-IQA Based on Deep Learning

TAIn recent years, learning-based methods have proven effective in No-Reference Image Quality Assessment (NR-IQA). Kang et al[7] introduced CNNs for NR-IQA, segmenting distorted images into 32×32 pixel patches and assigning quality scores. Shen et al[19] designed a saliency-based filtering model, incorporating an upsampling layer subnetwork and dual-stream feature fusion. Liu et al[20] proposed a multi-level feature enhancement method, with an Attention Enhancement Module (AEM) and Residual Feature Augmentation (RFA) module. Ye et al[21] introduced DRIQA-NR, using disentangled representation to separate content and distortion features. Zhu et al[22] presented an end-to-end Multi-task Efficient Transformer (METER) framework, with modules for distortion type

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

identification and adaptive quality prediction. Chen et al[23] introduced DFAN-IQA, integrating ResNet50 and Vision Transformer (ViT) through multi-attention and multi-semantic feature aggregation modules. Zhang et al[24] proposed an Attention-Driven Residual Dense Network, using multi-scale feature extraction and cascaded residual dense channel attention blocks.

Despite advancements, there is room for improvement in deep learning-based NR-IQA. Integrating Human Visual System (HVS) features can align models closer to human perception. HSV images facilitate color perception and manipulation, while CSG images excel in capturing edge details and localized quality fluctuations. To enhance multi-scale distortion feature representation, we introduce a dual-branch, multi-level residual network, learning content features from HSV images and structural/textural features from CSG images. This approach mimics HVS perception and comprehensively captures distortion information, yielding more precise and reliable IQA results.

3. Proposed Method

Despite advancements, there is room for improvement in deep learning-based NR-IQA. Integrating Human Visual System (HVS) features can align models closer to human perception. HSV images facilitate color perception and manipulation, while CSG images excel in capturing edge details and localized quality fluctuations. To enhance multi-scale distortion feature representation, we introduce a dual-branch, multi-level residual network, learning content features from HSV images and structural/textural features from CSG images. This approach mimics HVS perception and comprehensively captures distortion information, yielding more precise and reliable IQA results.

To effectively integrate human visual perception attributes with deep learning for No-Reference Image Quality Assessment (NR-IQA), we propose HVPIQA. The overall process is illustrated in Figure 1:

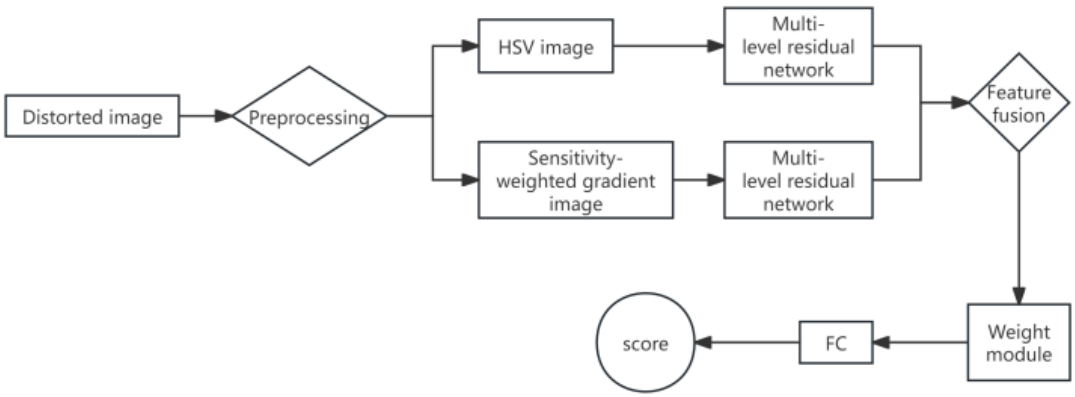

Figure 1: Overall flowchart.

Images are processed into HSV and contrast-sensitive weighted gradient images aligned with human perception, then fed into two identical multi-level residual networks: one for content features from HSV images, the other for structural and texture features from gradient images. These features are fused and weighted by a module to emphasize significant ones, finally translated into a quality score by a fully connected layer.

3.1. HSV Color Space Images

The HSV color space aligns with human perception, comprising hue (H) for color, saturation (S) for purity, and value (V) for brightness, enabling precise color variation visualization. The formula for converting RGB to HSV is outlined below[25]

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

$$H=\begin{cases}0^{\circ}, & \text{if } \max=\min \\ 60^{\circ}\times\dfrac{g-b}{\max-\min}+0^{\circ}, & \text{if } \max=r \text{ and } g\geq b \\ 60^{\circ}\times\dfrac{g-b}{\max-\min}+360^{\circ}, & \text{if } \max=r \text{ and } g<b \\ 60^{\circ}\times\dfrac{b-r}{\max-\min}+120^{\circ}, & \text{if } \max=g \\ 60^{\circ}\times\dfrac{r-g}{\max-\min}+240^{\circ}, & \text{if } \max=b \end{cases}$$

$$S=\begin{cases}0, & \text{if } \max=0 \\ \dfrac{\max-\min}{\max}, & \text{otherwise}\end{cases}$$

$$V=\frac{\max}{255} \tag{1}$$

An example of converting an RGB image to an HSV image is illustrated in Figure 2.

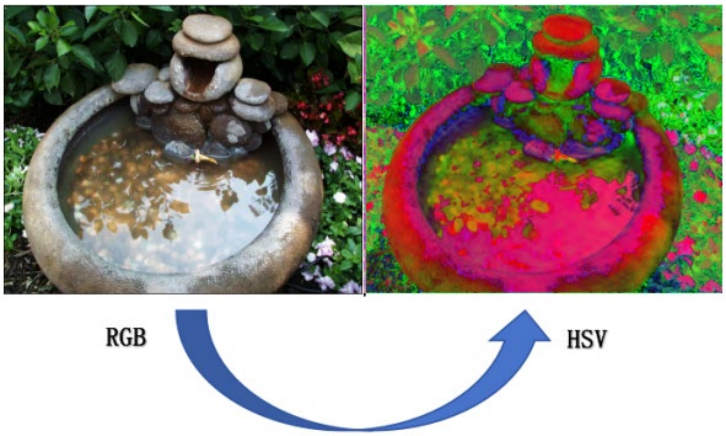

Figure 2: RGB to HSV.

3.2. CSG Images

Gradient images capture critical structural details sensitive to the human visual system (HVS). The Scharr operator[26], a popular edge detection filter, excels in precision and low computational complexity, ideal for real-time processing. Its unique 3x3 kernel uses second-order differences for accurate edge extraction, ensuring uniform sensitivity across directions. This is crucial for image quality assessment, capturing edge information in distorted images to assess structural integrity and clarity. The Scharr operator's structure is shown in Figure 3.

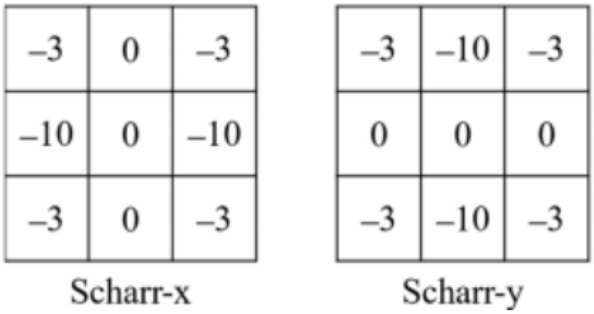

Figure 3: Scharr operator.

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

The HVS exhibits contrast sensitivity, varying with spatial frequencies, akin to spatial attention and image saliency[27]. Campbell et al[14] introduced a contrast sensitivity function to quantify HVS sensitivity to different spatial frequencies:

$$A(f) = 2.6(0.192 + 0.114f)e^{-(0.114f)^{1.1}} \tag{2}$$

Where f represents the spatial frequency of a point. For point I (i, j), its spatial frequency can be calculated as:

$$\begin{aligned} f &= \sqrt{f_x^2 + f_y^2} \\ f_x &= I(i,j) - I(i-1,j) \\ f_y &= I(i,j) - I(i,j-1) \end{aligned} \tag{3}$$

This method applies contrast sensitivity weighting to the gradient image to enhance frequency information that is sensitive to the HVS, thus aligning the model closely with the perception of the HVS. Specifically, the contrast sensitivity function is used to calculate the contrast sensitivity of each pixel in the distorted image. This results in a contrast sensitivity image, which is then combined with the gradient image to obtain a CSG image:

$$I_{CWG} = \alpha I_C + \beta I_G + \gamma \tag{4}$$

In the equation, I_{CWG} represents the gradient image, I_C represents the contrast sensitivity image, and I_G represents the CSG image. α, β and γ are constants. We have set α = β = 0.5 and γ = 0.

Representative gradient images and the corresponding CSG images are shown in the figure 4:

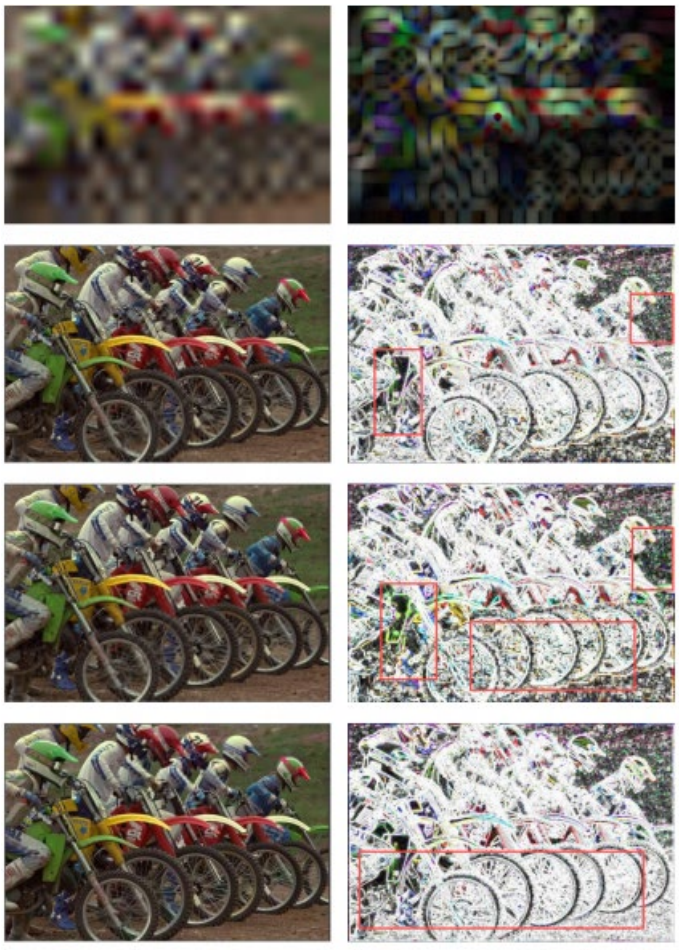

Figure 4: Comparison between distorted images and CSG images.

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

The case images show 'fastfading' distortion, decreasing from level 4 at the top to 1 at the bottom. The left column displays distorted images, and the right column shows their CSG counterparts. Comparison reveals that CSG images emphasize distortion changes due to their focus on human visual perception-sensitive areas, enhancing the human visual system's ability to discern these changes.

3.3. Multi-level Residual Network

When evaluating image quality, the HVS considers both global high-level and local low-level features. This is vital for IQA, as real-world distortions are unevenly distributed. Relying solely on global features may miss local distortions. Hence, we propose a multi-level residual network using nine residual blocks grouped into three levels to extract and fuse multi-scale distortion features, enabling simultaneous focus on both global and local features. Each block contains two convolutional layers, two batch normalization layers, and a PReLU. Detailed structures are shown in figures 5 and 6:

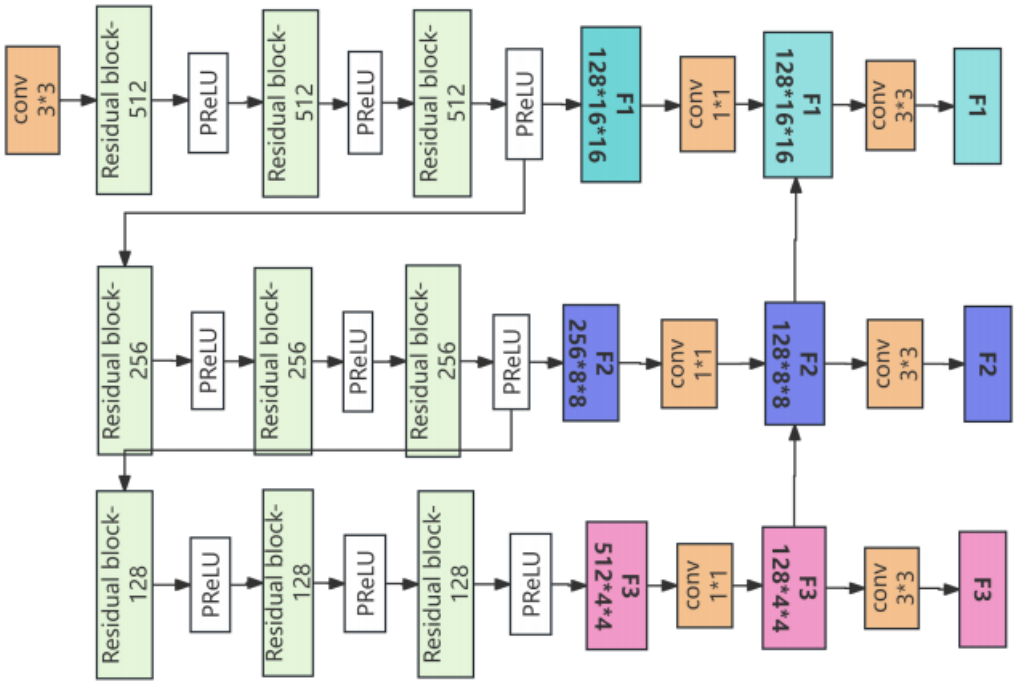

Figure 5: Diagram of The Multi-level Residual Network Structure.

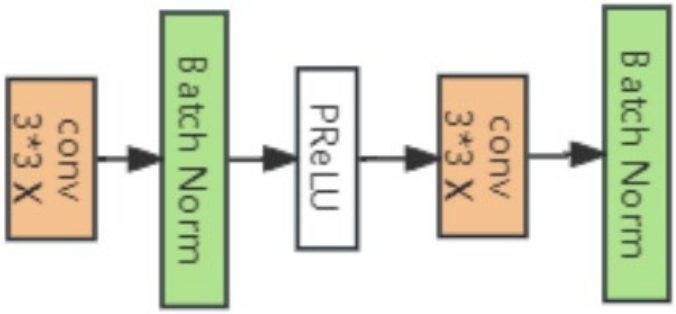

Figure 6: Residual Block-x.

After each convolutional layer, PReLU is used to avoid undesirable initialization. A 3×3 convolutional layer precedes the residual blocks to increase the number of feature map channels. Additionally, when the feature map size changes, a 1×1 convolutional layer is used before the residual blocks to adjust the size to match the input. The outputs after the PReLU of the third, sixth, and ninth residual blocks are designated as the first-level feature F1, second-level feature F2, and third-level feature F3, respectively. Subsequently, to reduce the number of channels, F1, F2, and F3 are processed through 1×1 convolutional layers to generate FF1, FF2, and FF3. When FF1, FF2, and FF3 differ in size, upsampling is applied to increase the spatial resolution of the coarser feature maps by a factor of 2. Specifically, FF3 is obtained by applying a 1×1 convolution to F3; FF2 is obtained by merging the result

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

of a 1×1 convolution on F2 with the upsampled result of FF3; and FF1 is obtained by merging the result of a 1×1 convolution on F1 with the upsampled result of FF2. To mitigate aliasing effects introduced by upsampling on each merged feature map, three 3×3 convolutional layers are used to obtain the final feature maps FFF1, FFF2, and FFF3. This process can be expressed as follows:

$$\begin{cases} FFF_3 = g\left(f\left(F_3\right)\right) \\ FFF_2 = g\left(f\left(F_2\right) \oplus h\left(f\left(F_3\right)\right)\right) \\ FFF_1 = g\left(f\left(F_1\right) \oplus h\left(f\left(F_2\right) \oplus h\left(f\left(F_3\right)\right)\right)\right) \end{cases} \tag{5}$$

In the formulas, $\oplus$ represents element-wise addition, $f()$ represents the 1×1 convolution operation, $g()$ represents the 3×3 convolution operation, and $h()$ represents the upsampling operation.

3.4. Weighting Mechanism

The method proposed in this paper posits that content and structural features contribute unequally to quality scores. Thus, a training-based weighting mechanism is introduced to assign weights to these features, as illustrated in Figure 7.

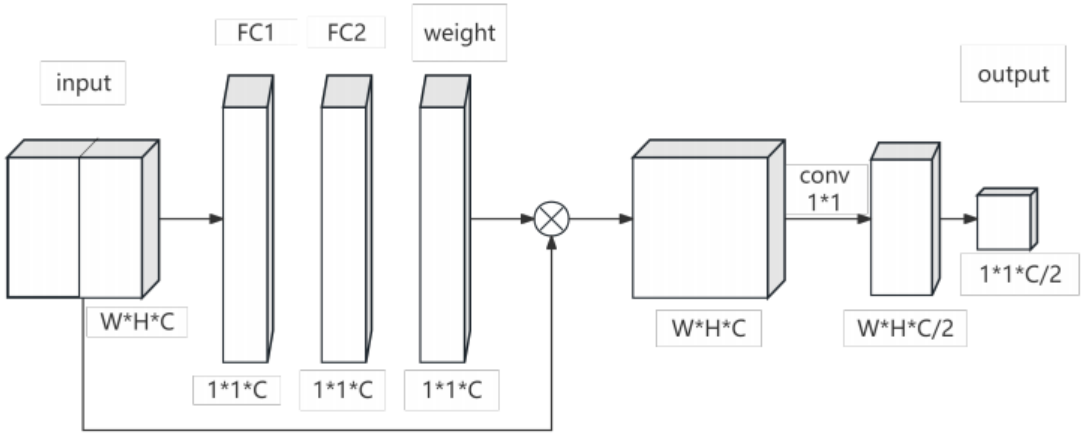

Figure 7: Weighting module.

Firstly, the fused content and structural features are concatenated to form final fused features. After passing through two fully connected layers, weight values for each channel are derived. These weights are then multiplied with the fused features to enhance their representational ability and robustness. A 1×1 convolution halves the number of channels, reducing computational costs. A global average pooling method is applied to obtain a multi-scale feature vector. The process can be summarized as follows:

$$\begin{cases} F = F_{\mathrm{i}} \otimes F_j \\ W = \sigma\left(W_2 \delta\left(W_1 GAP(F)\right)\right) \\ F' = WF \\ F_m = GAP\left(\ \mathrm{Conv}_{1\times1}\left(F'\right)\right) \end{cases} \tag{6}$$

In this formula, F_{i} and F_j represent the feature maps of the first and second branches, respectively. W_1 and W_2 are the parameters of the two fully connected layers. σ() and δ() denote the sigmoid function and the ReLU function, respectively. GAP() stands for global average pooling, and $\mathrm{Conv}_{1\times1}()$ represents a 1×1 convolution operation.

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

4. Experiments

4.1. Database

To evaluate our HVPIQA, we conducted extensive experiments using synthetic (LIVE[28], CSIQ[29], TID2013[30]) and authentic (LIVEC[31], KonIQ-10k[32]) distortion databases. These experiments compared our method against state-of-the-art approaches. Relevant details of the databases are summarized in Table 1.

Table 1: Details of the database.

Database	*Ref*	*Dis*	*Type*	*Score*
LIVE	*29*	*779*	*5*	*MOS*
CSIQ	*30*	*886*	*6*	*MOS*
TID2013	*25*	*3000*	*24*	*MOS*
LIVEC		*1162*		*MOS*
KonIQ-10k	*81*	*10.1k*	*25*	*MOS*

The LIVE dataset, established by the Image and Video Engineering Lab at the University of Texas at Austin in 2006, is recognized as one of the largest annotated image quality datasets. It comprises 779 distorted images generated by applying five types of computer-induced distortions, such as Gaussian blur, JPEG compression, and white noise, at 5 to 6 varying levels.

The CSIQ dataset, created by the Computational Perception and Image Quality Lab at Oklahoma State University in 2009, includes six distortion types with 4 to 5 levels each. These include Gaussian blur, JPEG compression, and additive white Gaussian noise among others.

The TID2013 dataset, developed by Tampere University of Technology as an extension of TID2008 and released in 2013, encompasses a wide range of distortion types, including newly introduced ones like saturation changes and comfort noise, in addition to traditional types like Gaussian blur and JPEG compression.

The LIVEC dataset, initiated in 2016, is known as the "wild image quality challenge dataset." It features images captured from real-world scenes by various photographers using different camera equipment, embodying complex and authentic distortions.

Established by Hosu et al. in 2020, the KonIQ-10k dataset is a large-scale collection for assessing real-world image quality. It meticulously selects 10,073 high-quality images from the YFCC100M dataset, ensuring diverse image content and quality distribution.

4.2. Experimental Scheme and Evaluation Metrics

To ensure no overlap between training and testing image content, we selected 80% of the synthetic distortion database for training using reference images, reserving the remaining 20% for testing. This approach was also applied to databases containing real distortions. To enhance reliability and generalization, we conducted ten random splits of each database following these guidelines, using the average of these ten experimental results as the final evaluation basis. When assessing IQA method effectiveness, we employed two metrics: Spearman's Rank Order Correlation Coefficient (SROCC) and Pearson's Linear Correlation Coefficient (PLCC). SROCC quantifies the monotonic relationship between predicted and actual scores, while PLCC evaluates their linear correlation. Both metrics range from -1 to 1, with higher absolute values indicating better model performance. By utilizing these two metrics, we could comprehensively and accurately assess IQA method performance.

4.3. Performance on a Single Database

We selected nine additional IQA methods for comparison, including two handcrafted feature-based methods: DIIVINE[2] and BRISQUE[5], and seven advanced deep learning-based methods: BIECON[33], HyperIQA[10], TS-CNN[12], DB-CNN[11], MEON[34], METER[22], and DFAN[23]. For some methods, we

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

adopted results reported in their original papers. To evaluate their performance, we used two common metrics: Spearman's Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC). SROCC measures the monotonicity of IQA metrics, while PLCC assesses the linear relationship between objective and subjective quality scores. Following convention, the top two scores in each category are bolded. Table 2 shows that all methods performed satisfactorily on LIVE and CSIQ databases with limited distortion types. However, performance significantly declined on TID2013 and KADID-10k databases with more complex distortions, as well as on LIVEC and KonIQ-10k databases with authentic distortions.

Table 2: Performance on a single database.

Method	LIVE		CSIQ		TID2013		LIVEC		KonIQ-10k	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
DIIVINE	0.923	0.913	0.743	0.777	0.664	0.535				
BRISQUE	0.942	0.940	0.829	0.746	0.694	0.604	0.585	0.607	0.692	0.673
BIECON	0.962	0.961	0.838	0.825	0.762	0.717	0.613	0.595	0.651	0.618
HyperIQA	0.966	0.962	0.942	0.923	0.858	0.840	0.882	0.859	0.917	0.906
TS-CNN	0.965	0.969	0.905	0.892	0.784	0.779	0.667	0.655	0.729	0.722
DB-CNN	0.972	0.960	0.908	0.901	0.855	0.835	0.601	0.606	0.736	0.722
MEON	0.954	0.943	0.890	0.839	0.811	0.828	0.688	0.693	0.760	0.754
METER	0.953	0.956	0.905	0.876	0.752	0.701	0.536	0.516	0.671	0.655
DFAN	0.971	0.968	0.959	0.946	0.865	0.816	0.869	0.851	0.884	0.875
Ours	0.975	0.972	0.952	0.943	0.860	0.846	0.873	0.856	0.914	0.909

In summary, based on SROCC and PLCC evaluations, the proposed HVPIQA exhibits commendable performance across five widely-used databases. Compared to other dual-path architectures, including DB-CNN[11] and TS-CNN[12], our method holds a superior position on most databases. Notably, it demonstrates significant performance differences compared to TS-CNN[12] on databases with authentic distortions. This superiority is mainly attributed to the incorporation of human visual system characteristics in HVPIQA. Specifically, HSV space images enable more effective extraction of content information from distorted images, while CSG images facilitate better extraction of frequency information relevant to human vision. Furthermore, the introduced multi-scale feature fusion module allows the model to focus on both global content and local details, prioritizing more important aspects of content and structural features through a weighting mechanism.

4.4. Performance on Individual Distortion Types

To evaluate model performance on individual distortion types, we conducted additional experiments using the LIVE and CSIQ databases. Our model was trained on all distortion types present in these databases and then tested on each individual distortion type. For performance assessment, HVPIQA was compared with nine other IQA algorithms: DIIVINE[2], BRISQUE[5], BIECON[33], HyperIQA[10], TS-CNN[12], DB-CNN[11], MEON[34], IQA-CNN[7], and METER[22]. Table 3 presents the SROCC and PLCC results for individual distortion types on the LIVE and CSIQ databases. From Table 3, it is evident that HVPIQA consistently achieved the top two SROCC values for single distortion types on both LIVE and CSIQ datasets, highlighting its significant advantage in assessing image quality for specific distortion types.

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

Table 3: SROCC comparisons of various distortion types on the LIVE and CSIQ datasets.

Method	PLCC					SROCC				
	JP2K	JPEG	WN	BLUR	FF	JP2K	JPEG	WN	BLUR	FF
DIIVINE	0.901	0.887	**0.987**	0.787	0.879	0.925	0.913	**0.985**	0.789	0.873
BRISQUE	0.923	0.973	0.985	0.951	0.903	0.914	0.965	0.978	0.951	0.877
BIECON	0.949	0.971	0.977	0.951	0.916	0.952	**0.974**	0.980	0.956	0.923
HyperIQA	0.954	0.966	0.979	0.940	**0.945**	0.949	0.961	0.972	0.926	**0.934**
TS-CNN	**0.971**	0.967	0.984	**0.970**	0.934	**0.966**	0.950	0.979	**0.963**	0.911
DB-CNN	0.920	0.964	0.980	0.951	0.933	0.914	0.951	0.972	0.944	0.926
MEON	0.950	0.959	0.971	0.960	0.899	0.953	0.964	0.981	0.958	0.904
IQA-CNN	0.941	0.970	0.979	0.966	0.914	0.936	0.965	0.974	0.952	0.906
METER	0.964	**0.979**	0.981	0.944	0.939	0.955	0.972	0.980	0.935	0.930
Ours	**0.977**	**0.984**	**0.988**	**0.972**	**0.944**	**0.972**	**0.981**	**0.986**	**0.967**	**0.948**

4.5. Performance Across Different Databases

Consistency between prediction results and human subjective evaluations is crucial in Image Quality Assessment (IQA) methods. Therefore, an excellent IQA method must demonstrate high accuracy across various databases. To determine the generalization capability of IQA methods, we conducted cross-database evaluation experiments. Specifically, we trained the IQA methods on the LIVE database and evaluated their performance on the CSIQ and KonIQ-10k databases. In these comparative evaluations, we benchmarked our HVPIQA against DIIVINE[2], BRISQUE[5], BIECON[33], HyperIQA[10], TS-CNN[12], DB-CNN[11], MEON[34], IQA-CNN[7], and METER[22]. The SROCC results on the CSIQ and KonIQ-10k databases are summarized in Table 4, where A/B indicates training on A and testing on B.

Table 4: Generalizability comparison of cross-dataset testing on SROCC.

Method	LIVE/CSIQ	LIVE/KonIQ-10k	LIVE/KonIQ-10k
DIIVINE	0.602	0.342	0.791
BRISQUE	0.573	0.354	0.735
BIECON	0.732	0.365	0.749
HyperIQA	0.744	0.579	0.857
TS-CNN	0.621	0.431	0.823
DB-CNN	0.702	0.415	0.815
MEON	0.605	0.387	0.788
IQA-CNN	0.698	0.407	0.803
METER	0.769	0.581	0.901
Ours	0.755	0.586	0.910

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

Our analysis of the experimental results from the first two groups, conducted on different databases, reveals that HVPIQA consistently ranks among the top two. However, when comparing training and testing performance between different databases versus within the same database, a noticeable performance decline is evident. This decline can be attributed to significant differences in distortion types between the LIVE database used for training and the other two databases. Specifically, the LIVE and CSIQ datasets share only four common distortion types: JP2K, JPEG, WN, and BLUR. Furthermore, the CSIQ dataset includes two additional distortion types not present in the LIVE dataset, indicating that our model has not been exposed to these extra distortion types during training, which may contribute to the lower scores. To validate this observation, we conducted a third group of experiments, training on the CSIQ dataset and testing on the LIVE dataset, which confirmed our initial findings.

These experimental results demonstrate that while HVPIQA performs well across multiple databases, its performance is affected when faced with distortion types not included in the training set. This highlights the challenges faced by IQA methods in terms of generalization, particularly when dealing with multiple distortion types. To improve the model's generalization capability, future research could explore using a wider range of distortion types for training or developing IQA methods that can adaptively learn new distortion types.

4.6. Ablation Experiments

To validate the effectiveness of each module in HVPIQA, we conducted ablation studies on the LIVE, TID2013, and CSIQ datasets. The results of these ablation studies are presented in Table 5.

Table 5: Ablation experiment.

baseline	*1*	*1*	*1*	*1*
RGB image	*1*	*0*	*0*	*0*
HSV image	*0*	*1*	*1*	*1*
CSG image	*0*	*1*	*1*	*1*
MFEF	*0*	*0*	*1*	*1*
WM	*0*	*0*	*0*	*1*
LIVE	*0.947/0.952*	*0.956/0.961*	*0.964/0.968*	*0.972/0.975*
CSIQ	*0.909/0.921*	*0.924/0.937*	*0.933/0.946*	*0.943/0.952*
TID2013	*0.765/0.802*	*0.806/0.825*	*0.821/0.840*	*0.846/0.860*

The first column represents the use of the base network architecture with RGB images as training data for two different branches. The second column employs the base network structure but uses HSV and CSG images for the two branches, noting improvements in SROCC/PLCC results by 0.9%/0.3%, 1.5%/1.6%, and 4.1%/2.3% across the three datasets, respectively. The third column integrates the Multi-Feature Enhanced Fusion (MFEF) module, leading to further improvements in SROCC/PLCC results by 0.8%/0.7%, 0.9%/0.9%, and 1.5%/1.5% on the three datasets. Finally, the fourth column represents the comprehensive method proposed in this paper, which achieves additional improvements in SROCC/PLCC results by 0.8%/0.7%, 1.0%/0.6%, and 2.5%/2.0% on the three datasets, respectively.

These results demonstrate that changing the type of training data (from RGB to HSV and CSG) and integrating the MFEF module can significantly enhance the performance of Image Quality Assessment (IQA) methods. Furthermore, the comprehensive method proposed in this paper builds on these improvements, indicating that the synergy between the modules is crucial for improving overall performance. The results of these ablation studies provide strong evidence for the effectiveness of the HVPIQA method.

5. Conclusion

This paper introduces a no-reference image quality assessment (NR-IQA) method that integrates human visual perception with a dual-branch multi-level residual network. The effectiveness of this

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

method is demonstrated through three key modules: dual-branch feature extraction, multi-level feature extraction and fusion, and a weighting mechanism. The method converts input images into HSV and CSG images that align with human perception, while simultaneously extracting content and structure-texture features through parallel residual networks. The fusion of these complementary multi-level features enhances the overall feature representation, and the weighting mechanism fine-tunes their respective contributions to the final quality score.

Upon reviewing the development and evaluation of this method, several key insights are gained. Firstly, incorporating diverse image representations and features that align with human perception is crucial for achieving accurate and reliable NR-IQA. Secondly, the design of the multi-level residual network facilitates the effective extraction of hierarchical information, which is essential for capturing subtle variations in image quality.

Compared to synthetically distorted images, real-world distorted images often exhibit complex and varied conditions, often with randomness and unpredictability. Current state-of-the-art IQA methods face numerous challenges when assessing the quality of real-world distorted images, such as limited generalization ability, high computational complexity, resource constraints, and inconsistency in subjective evaluations. Therefore, future research must delve deeper into overcoming these challenges and enhancing the accuracy and practicality of IQA methods.

To achieve this goal, future research can focus on several areas: firstly, developing more generalized models that can adapt to a wider range of distortion types and image content; secondly, optimizing algorithms to reduce computational complexity and increase evaluation speed; thirdly, exploring more effective feature representation and extraction methods to more accurately capture subtle changes in image quality; and fourthly, leveraging deep learning and other technologies to enhance the adaptive capabilities and robustness of models. Through these efforts, we can drive the continuous development of IQA methods and provide more reliable tools for accurate image quality assessment.

References

[1] Li, F., Shuang, F., Liu, Z., Qian, X.: A cost-constrained video quality satisfaction study on mobile devices. IEEE Transactions on Multimedia 20(5), 1154–1168(2017)
[2] Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE transactions on Image Processing 20(12), 3350–3364 (2011)
[3] Saad, M.A., Bovik, A.C., Charrier, C.: A dct statistics-based blind image quality index. IEEE Signal Processing Letters 17(6), 583–586 (2010)
[4] Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: A natural scene statistics approach in the dct domain. IEEE transactions on Image Processing 21(8), 3339–3352 (2012)
[5] Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing 21(12), 4695–4708(2012)
[6] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters 20(3), 209–212 (2012)
[7] Kang, L., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733–1740 (2014)
[8] Bosse, S., Maniry, D., M¨uller, K.-R., Wiegand, T., Samek, W.: Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on image processing 27(1), 206–219 (2017)
[9] Kim, J., Nguyen, A.-D., Lee, S.: Deep cnn-based blind image quality predictor. IEEE transactions on neural networks and learning systems 30(1), 11–24 (2018)
[10] Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.3667–3676 (2020)
[11] Zhang, W., Ma, K., Yan, J., Deng, D., Wang, Z.: Blind image quality assessment using a deep bilinear convolutional neural network. IEEE Transactions on Circuits and Systems for Video Technology 30(1), 36–47 (2020) https://doi.org/10.1109/TCSVT.2018.2886771
[12] Yan, Q., Gong, D., Zhang, Y.: Two-stream convolutional networks for blind image quality assessment. IEEE Transactions on Image Processing 28(5), 2200–2211(2018)
[13] Mannos, J., Sakrison, D.: The effects of a visual fidelity criterion of the encoding of images. IEEE transactions on Information Theory 20(4), 525–536 (1974)
[14] Campbell, F.W., Robson, J.G.: Application of fourier analysis to the visibility of gratings. The

ISSN 2616-5775 Vol. 8, Issue 2: 7-19, DOI: 10.25236/AJCIS.2025.080202

Journal of physiology 197(3), 551 (1968)
[15] Gao, X., Lu, W., Tao, D., Li, X.: Image quality assessment based on multiscale geometric analysis. IEEE Transactions on Image Processing 18(7), 1409–1423(2009)
[16] Saha, A., Wu, Q.M.J.: Utilizing image scales towards totally training free blind image quality assessment. IEEE Transactions on Image Processing 24(6), 1879–1892 (2015)
[17] Sheikh, H.R., Bovik, A.C., De Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Transactions on image processing 14(12), 2117–2128 (2005)
[18] Yue, G., Hou, C., Zhou, T., Zhang, X.: Effective and efficient blind quality evaluator for contrast distorted images. IEEE Transactions on Instrumentation and measurement 68(8), 2733–2741 (2018)
[19] Shen, L., Zhang, C., Hou, C.: Saliency-based feature fusion convolutional network for blind image quality assessment. Signal, Image and Video Processing 16(2),419–427 (2022)
[20] Liu, C., Zheng, Y., Liao, K., Chen, B., Wang, K., Zhong, C., Xie, B., Miao, Y.: No-reference image quality assessment of multi-level residual feature augmentation.Signal, Image and Video Processing 17(4), 1275–1283 (2022)
[21] Ye, Z., Wu, Y., Liao, D., Yu, T., Yang, J., Hu, J.: Driqa-nr: no-reference image quality assessment based on disentangled representation. Signal, Image and Video Processing, 1–9 (2022)
[22] Zhu, P., Liu, S., Liu, Y., Yap, P.-T.: Meter: Multi-task efficient transformer for no-reference image quality assessment. Applied Intelligence 53, 29974–29990 (2023)
[23] Chen, Y., Chen, Z., Yu, M., Tang, Z.: Dual-feature aggregation network for no-reference image quality assessment. MultiMedia Modeling, 149–161 (2023)
[24] Zhang, Y., Wang, C., Lv, X., Song, Y.: Attention-driven residual-dense network for no-reference image quality assessment. Signal, Image and Video Processing 18(Suppl 1), 537–551 (2024)
[25] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
[26] He, Y., Xin, M., Wang, Y., Xu, C.: Automatic edge detection method of power chip packaging defect image based on improved canny algorithm (2024).IEEEXplore
[27] Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009). IEEE
[28] Jayaraman, D., Mittal, A., Moorthy, A.K., Bovik, A.C.: Objective quality assessment of multiply distorted images. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp.1693–1697 (2012). IEEE
[29] Eric, C., Larson, Damon, M.:Most apparent distortion: full-reference image quality assessment and the role of strategy[J].Journal of Electronic Imaging 19(1) (2009).
[30] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al.: Image database tid2013: Peculiarities, results and perspectives. Signal processing: Image communication 30, 57–77 (2015)
[31] Ghadiyaram, D., Bovik, A.C.: Massive online crowdsourced study of subjective and objective picture quality. IEEE Transactions on Image Processing 25(1),372–387 (2015)
[32] Hosu, V., Lin, H., Sziranyi, T., Saupe, D.: Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing 29, 4041–4056 (2020)
[33] Kim, J., Lee, S.: Fully deep blind image quality predictor. IEEE Journal of selected topics in signal processing 11(1), 206–220 (2016)
[34] Yan, Q., Gong, D., Zhang, Y.: Two-stream convolutional networks for blind image quality assessment. IEEE Transactions on Image Processing 28(5), 2200–2211(2019)