Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

BPGI: A Brain-Perception Guided Interactive Network for Stereoscopic Omnidirectional Image Quality Assessment

Yun Liu[1], Member, IEEE, Sifan Li[1], Zihan Liu[1], Haiyuan Wang[1], and Daoxin Fan[1]

[1]Faculty of information, Liaoning University, Shenyang 110036, China

Corresponding author: Yun Liu (email: yunliu@lnu.edu.cn).

ABSTRACT Stereoscopic omnidirectional image quality assessment is a combination task of stereoscopic image quality assessment and omnidirectional image quality assessment, which is more challenge than traditional three-dimensional image. Previous works fail to present a satisfying performance due to neglecting human brain perception mechanism. To solve the above problem, we proposed an effective brain-perception guided interactive network for stereoscopic omnidirectional image quality assessment (BPGI), which is built following three perception step: visual information processing, feature fusion cognition, and quality evaluation. Considering the stereoscopic perception characteristics, binocular and monocular visual feature are both extracted. Following human complexity cognition mechanism, a Bi-LSTM module is introduced to dig the deeply inherent relationship between monocular and binocular visual feature and improve the feature representation ability of the proposed model. Then a visual feature fusion module is built to obtain effective interactive fusion for quality prediction. Experimental results prove that the proposed model outperforms many stat-of-the-art models, and can be effectively applied to predict the quality of stereoscopic omnidirectional image.

INDEX TERMS Human brain perception mechanism, quality assessment, stereoscopic omnidirectional image, viewport feature extraction

I. INTRODUCTION

WITH the development of information and technology, omnidirectional content has developed rapidly, among which stereoscopic omnidirectional image (SOI) arouses much attention [1], [2]. SOI not only has the wider view field, but also has a stereoscopic perception, which can provide higher immersive quality of experience (QoE). Unfortunately, it is inevitably introduced noise during capturing, transmission and display, which leads poor QoE. It is needed to build effective quality assessment model to help suppliers select high quality of SOI. To deal with the above problem, researchers focus on designing effective SOI quality assessment (SOIQA) models, which can be divided into two categories: subjective model and objective model. Subjective model is time-consumer, inconvenient and unreal-time, while the objective model can automatically provide real-time prediction results, which becomes the main-stream method of SOIQA.

SOIQA is a more challenge problem than traditional two-dimensional (2D) image quality assessment, and it needs to take both the panoramic characteristics and stereoscopic perception into account, which is corresponding to omnidirectional image quality assessment (OIQA) task and stereoscopic image quality assessment (SIQA) task. Like traditional 2D image quality assessment task, three types of quality assessment models are built for the SIQA task and OIQA task: full-reference (FR), reduce-reference (RR) and no-reference (NR) models. Considering that the reference image is not always available, NR models has more practical significance than FR and RR models. For the SIQA task, the earlier NR models are proposed based on 2D image quality metrics, such as SSIM [3], PSNR [4], and so on [5], [6], [7], which leads to poor performance without considering the unique characteristics of stereoscopic image (SI). Considering the disparity information in SI, SIQA models are built based on human stereoscopic visual characteristics, which achieves a improve performance [8], [9], [10]. Among

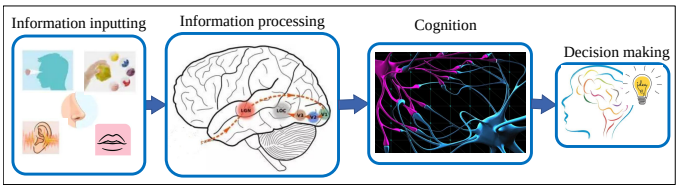

FIGURE 1. The Diagram of Human Making Decision.

them, the combination of binocular and monocular information presents a promising performance, which identifies the direction of stereoscopic quality assessment task. For the OIQA task, at the beginning, researchers take omnidirectional image (OI) as an expanding 2D image with different projection format to design OIQA model, and neglect the unique characteristics of OI, which limits the overall performance. Considering that consumer usually watches one viewport at a time by wearing the Head Mounted Displays (HMDs), the viewport-based OIQA models are proposed, which improves the overall performance [11]. Inspired by human visual system (HVS), recent OIQA models, combing the local viewport visual features and global visual features of the entire OI, present a further performance improvement, which provides a novel way to solve the OIQA task.

As the emerging combination task, both the stereoscopic visual characteristics and omnidirectional visual characteristics should be considered in SOIQA model [12], [13]. Reviewing the models of SIQA and OIQA, human visual system plays an important role, which provides us an inspiration to design an effective SOIQA model. Recent research found that during human make the decision, a series of activities occur in human brain cerebral cortex, which involves in information processing, cognition and decision making [14]. shown in Fig.1. Inspired by this, we propose a brain-perception guided interactive network (BPGI) for SOIQA by combing binocular and monocular visual information processing, feature interactive fusion and quality evaluation. Firstly, following human stereoscopic perception mechanism, binocular and monocular visual information are both captured, and global and local visual features are fused together to dig the rich sematic information. Then a cognition interactive module is built to fuse the binocular and monocular visual features, which can well complement with each other and help people to make the final decision. Finally, the fusion features are mapped to get the quality score.

Our contributions can be concluded as follows:

- The proposed model is designed by following the whole procedure of human brain decision making, which is consistent with human visual perception.
- Multi-scale strengthen (MSS) module is designed to strengthen the feature representation ability and the Bi-LSTM is introduced to effectively fuse local and global features, which can dig the deeply inherent relationship and improve the learning ability of the proposed model.
- The interactive fusion of monocular and binocular visual features can well complement with each other, and assist the model to capture rich stereoscopic information, which can further improve the overall performance.
- The viewport characteristics of OI and stereoscopic visual perception of SI are both considered, which can improve the accuracy of the model.

II. Related Works

A. A RELATED QUALITY ASSESSMENT MODELS

SOIQA is the combination of SIQA and OIQA, so the existed SIQA and OIQA models can point the way to build the SOIQA model. For the SIQA models, they can be classified into two types: 2D metric-based model and HVS-based model. The 2D metric-based model is designed by directly applying 2D metrics on SI without considering disparity information, and which leads to poor results [15], [16]. According to the physiological research, HVS-based SIQA models are proposed [17], [18], [19]. Shen et al. [20] established a novel SIQA model by combining global features and Local features. Shao et al. [21] and Chen et al. [22] focus on human binocular perception mechanism, and proposed the effective SIQA models, which promotes the development of SIQA model. Based on the binocular fusion and competition in human brain, Li et al. [23] applied the frequency information to simulate human binocular perception, and built an effective SIQA network. By simulating the process of human stereo visual perception, Zhou et al. [17] extract the visual features of each view and built a multi-layer interactive network, which proves the importance of monocular and binocular visual features in SIQA. Liu et al. [24] then proposed a two-stream interactive network to dig the binocular visual features and the monocular visual features, which yields good performance. The above works give us a hint that binocular and monocular visual information are the key information in SIQA model, which should be extracted following human stereoscopic perception mechanism.

For the OIQA models, they can be classified into three types: 2D metric-based models, projection-based models, and viewport-based models [25], [26]. Considering the wide view information, 2D metric-based models take the OI as a traditional 2D image and directly calculate the 2D metrics to predict the OI quality score [27], [28]. Focus on the characteristics of projection methods, projection-based models are built by extracting the visual features from many image patches, which yields a better performance than 2D metric-based models [29], [30], [31]. However, the above models fail to accurately simulate the real visual perception procedure during people watching an OI, which limits its overall performance. Motivated by the viewports' characteristics, viewport-based models are then proposed [32], [33]. Sun et al. [34] extracted visual features from six viewports and applied a multi-channel network to conduct the quality

evaluation, which further improves the performance of OIQA models compare to projection-based models. Yang et al. [35] proposed a Transformer-based model by generating the viewport sequence and predicted the quality of OI. Xu et al. [36] proposed a novel viewport-oriented graph neural network by combing the viewport local visual information and the global information, which proves the reasonable of fusing local and global features. Later, Yan et al. [37] proposed a blind OIQA models by capturing both global and local visual information, which further verifies the important role of local and global features in OIQA. The successful of viewport-based model that considering local and global visual features simultaneously provides us a way to solve the OIQA task.

B. SOIQA WORKS

Compared with SIQA and OIQA tasks, SOIQA task is more challenge due to the complex depth perception mechanisms in a whole 180×360° range surrounding the viewer. Yang et al. [38] built a free available dataset and designed an end-to-end three-dimensional (3D) convolutional neural network to predict the 3D panoramic virtual reality video quality. However, the above model neglected the importance of viewports during the training process. Tian et al. [11] focus on the stitching distortions introduced in the panoramic content, and proposed a Viewport-Sphere-Branch Network (VSBNet) via dual-branch quality estimation, which can well reflect the characteristics of panorama perception. Considering the actual viewing experience and binocular characteristics, Chen et al. [12] proposed a multi-viewport based full-reference SOIQA model, which can provide an overall quality score, including the depth perception, visual comfort and image quality. Later, considering the binocular rivalry and multi-view characteristics, they [12] built an effective predictive coding theory-based SOIQA method, which proves the importance of viewport and binocular visual features in SOIQA task. Considering the particularities of imaging and display of SOI, Qi et al. [39] took the binocular perception and omnidirectional perception into consideration, and proposed a viewport perception-based blind SOIA method. To eliminate 2D-to-Sphere intrinsic sampling distortions via deformable convolutions, Chai et al. [9] adopted a deformable convolutional neural network to extract spherical features, and built a three-channel network for feature encoding, including left view, right view and binocular difference channels, which proves that the monocular should be considered like in SIQA models. Overall, the above models were all considered two aspects: the viewport perception and stereoscopic perception, and how to accurately simulate the human viewing process and perception mechanism is the key to build an effective SOIQA model.

III. PROPOSED MODEL

Incorporating the SIQA task and OIQA task, we propose an effective OSIQA model based on human brain perceptual mechanism, shown in Fig.2. Specifically, each view of SOI is stitched into several viewports, and global binocular visual features and local binocular visual features are firstly extracted for stereoscopic perception. Then global monocular visual features and local monocular visual features are also obtained to complement the binocular visual features for quality perception. To simulate human brain perception mechanism, global binocular visual features and local binocular visual feature, and global monocular visual features and local monocular visual feature are fused together, respectively, to form effective cognition. Finally, an interactive fusion module is built to achieve smoothly fusion to dig the interactive semantic information, which can provide an accurate predict score. The details are described as follows.

A. BINOCULAR VISUAL FEATURES

For the give SOI, we apply the difference map between the left view and right view to capture the binocular visual features, which has been proved its effects on disparity information representation and computing complexity reducing [19]. Take the left view and right view of SOI as I_L and I_R respectively, the difference map can be got as follows:

$$d_g = |I_L - I_R| \tag{1}$$

The difference map is fed into a VMamba module, and last three stages' outputs of VMamba, show in Fig.3, are sent into the MSS module to strengthen the global binocular feature. Specifically, the last three stages' outputs of VMamba are concatenated together and then further strengthened its semantic representation ability based on the MSS module. The process can be formulated as follows:

$$C_3^g = \text{LN}(\text{Concat}(f_2^g, f_2^g, f_2^g)) \tag{2}$$

$$B_g = (\text{softmax}(Li(C_3^g) \otimes Li(C_3^g)) \otimes Li(C_3^g)) \otimes C_3^g \tag{3}$$

where f_i^g denotes the output of the i-th (i=2,3,4) stage of VMamba. LN is Layer Normalization, Li denotes the linear layer and Concat is the concatenation operation. B_g is the global binocular feature.

Considering that local visual feature can well complement with the global visual features, we split the left view and the right view of the SOI into several viewports following work [11], respectively, which can provide rich local visual information. Like the global binocular features extraction procedure, the difference map of each viewport is captured and sent into VMamba followed by MSS module to capture the local binocular features of each viewport. Then the local binocular features of all viewports are concatenated together to obtain the local binocular visual features, denoted as B_l.

B. MONOCULAR VISUAL FEATURES

It has been proved that monocular visual features play an important role in SIQA task, so we also extract the

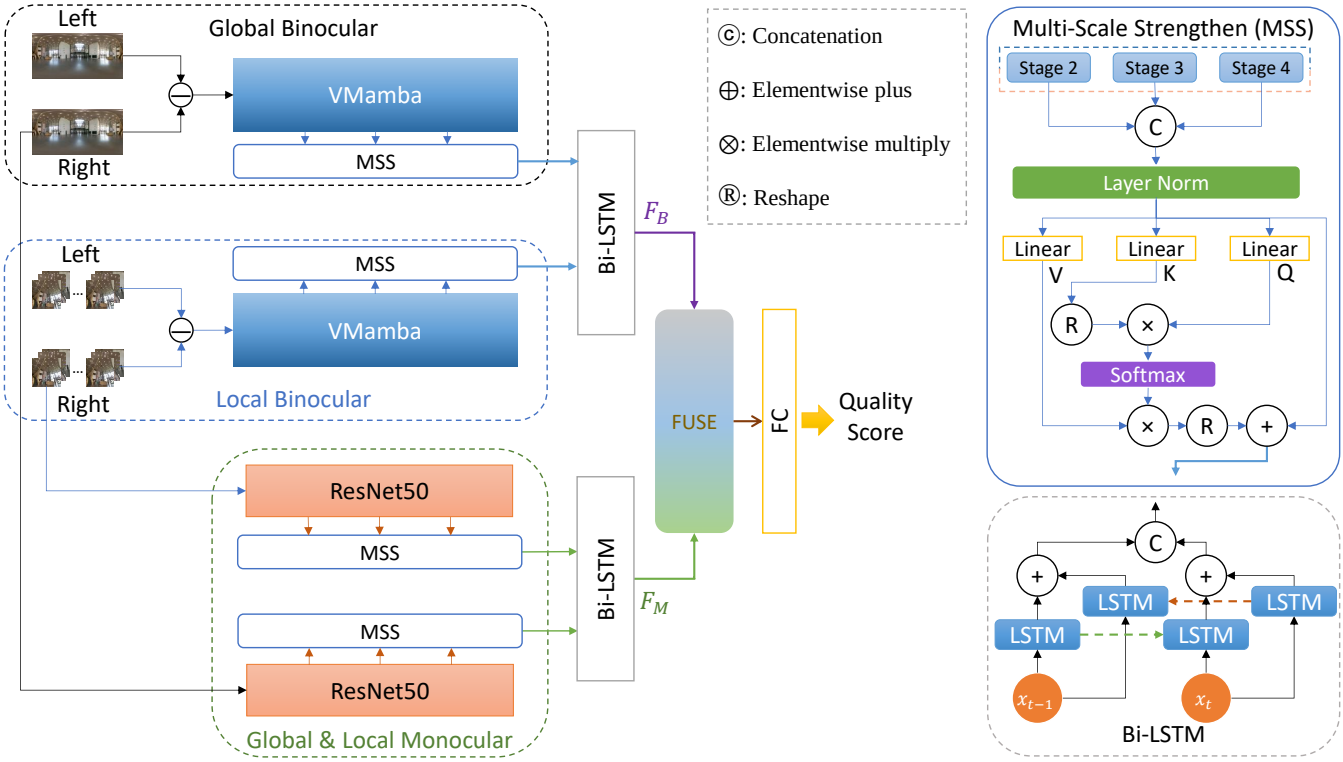

FIGURE 2. **The Flowchart of the Proposed Model.**

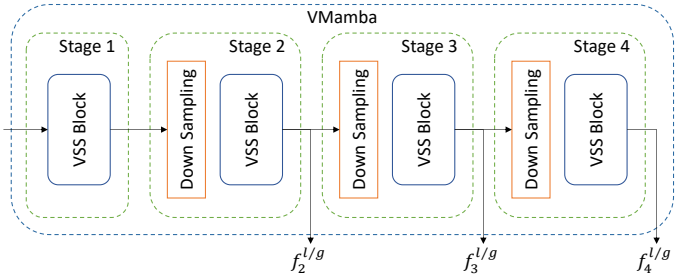

FIGURE 3. **The diagram of VMamba.**

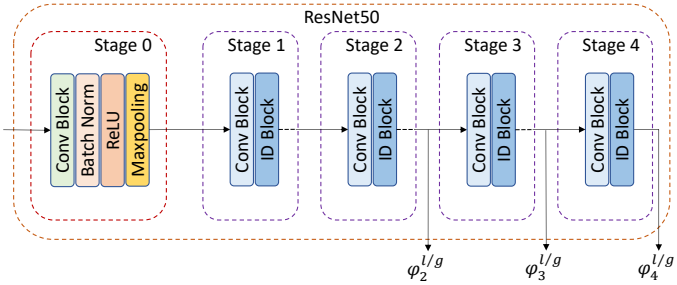

FIGURE 4. **The Diagram of ResNet50.**

monocular visual features to improve the performance of our model, including global and local monocular visual features. Considering that the left view and the right view of a SOI is similar, we only take the right view to extract the monocular visual features. ResNet50 is taken as the backbone, and the outputs of the last three stage, shown in Fig.4, are extracted to capture the monocular visual features, including the global and local monocular visual features. Take the global monocular visual feature extraction as an example, given the right view of the distorted SOI, the multi-scale monocular features are firstly obtained, and then sent into the MSS module to strengthen and extract the global monocular visual features M_g, shown as follows:

$$K_3^g = \mathrm{LN}(\mathrm{Concat}(\phi_2^g, \phi_3^g, \phi_4^g)) \tag{4}$$

$$M_g = (\mathrm{softmax}(Li(K_3^g) \otimes Li(K_3^g)) \otimes Li(K_3^g)) \otimes K_3^g \tag{5}$$

Similar to the above procedure, each viewport is sent into the ResNet50 followed by a MSS module to capture the monocular features of each viewport. Then the monocular features of all viewports are concatenated together to obtain the local monocular visual features M_l.

C. FEATURE COGNITION FUSION

Following human brain cognition mechanism, we take the Bi-LSTM module to fuse the local and global visual features, including binocular and monocular visual features, which can well dig the inherent relationship between them and strengthen the feature representation of our model. Take the feature fusion of local binocular features and global binocu-

lar features as an example, the cognition fusion process can be formulated as follows.

$$\overrightarrow{h_t} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}}) \tag{6}$$

$$\overleftarrow{h_t} = \text{LSTM}(x_t, \overleftarrow{h_{t-1}}) \tag{7}$$

$$y_t = \vec{W}\vec{h_t} + \overleftarrow{W}\overleftarrow{h_t} + b_y \tag{8}$$

$$F_B = \text{Concat}(\text{Concat}_i^n(y_i^g), \text{Concat}_t^n(y_i^l)) \tag{9}$$

where $\overrightarrow{h_t}$ denotes the forward hidden layer value of the t-th input, and $\overleftarrow{h_t}$ denotes the backward hidden layer value of the t-th input. $\overrightarrow{W}$ denotes the forward propagation weights and $\overleftarrow{W}$ denotes the backward weights. y_i^g denotes the i-th element in global binocular visual features B_g, and y_i^l denotes the i-th element in local binocular visual features B_l. F_B is the binocular feature.

Similar to the binocular feature cognition fusion, the monocular features F_M are then obtained.

D. FEATURE INTERACTIVE FUSION AND QUALITY AND QUALITY PREDICTION

Following human stereoscopic perception mechanism, we built an interactive fusion module, shown in Fig.5, to fuse the binocular features F_B and monocular features F_M, which can be formulated as follows:

$$F'(x) = \text{MLP}(\text{softmax}(x)) \tag{10}$$

$$F = \text{Conv}(\text{ReLU}(F'(F_B) \oplus (F'(F_B) \otimes F_M))) \tag{11}$$

where Conv means a convolutional layer and MLP denotes a multilayer perception.

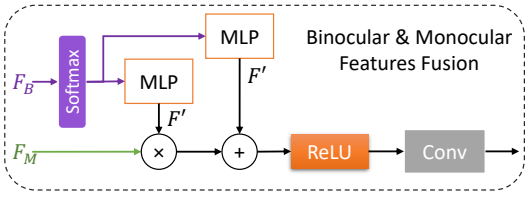

FIGURE 5. Feature Interactive Fusion Module.

Then the fused feature F is sent into a fully-connected layer to map into the quality score. We calculate the Euclidean loss to update the parameters of our model, which can be formulated as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} ||s_i - p_i||_2^2 \tag{12}$$

where N is the number of the training samples, s_i denotes the ground truth quality score of the i-th sample, and p_i denotes the i-th sample's quality score predicted by the model.

IV. EXPERIMENTAL RESULTS AND COMPARISON

A. DATASET

SOLOD dataset [40]. It is constructed 276 distorted images based on 6 reference images with JPEG and BPG compression, which consists 84 symmetrically and 192 asymmetrically distorted images. Each of them is assigned with the subjective mean opinion score (MOS) value, which is range from 1 to 5. The higher the value is, the better quality is.

B. CRITERIA

Three commonly used criteria are adopted to prove the performance of the model, including Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC) and Root Mean Squared Error (RMSE). PLCC and SRCC are range from 0 to 1, and higher value means better performance, while lower RMSE means better performance.

C. RESULTS AND COMPARISON

1) Performance on Benchmark Dataset

To prove the effectiveness of our model, many state-of-the-art models are applied to make the comparison, which can be classified into 4 categories: 2D IQA models, 2D OIQA models, 3D IQA models and 3D OIQA models. The comparison results are listed in Table 1. It can be seen that 2D OIQA has the worst performance among all the models, which indicates that simply apply 2D OIQA to evaluate 3D OI quality does not work. The probably reason is that they either considered the viewport features, or 3D visual features. 3D IQA models present a promising result, even outperform one 3D OIQA model, which proves the contribution of stereoscopic characteristics in SOIQA. 3D OIQA models achieve better performance than the above three types of models, especially SOIQE model, which further proves the necessary of combining SIQA and OIQA. While our model, not only considering stereoscopic binocular and monocular visual features, but also taking the local viewport stereoscopic perception into consideration, yields the best performance among all the four types of models, which proves the reasonable and effectiveness of our model. The above results indicate that our model can effectively applied to predict the quality of SOI, and is consistent with human visual system.

2) Performance on Different Distortion Types

To verify the performance of our model on each distortion type, we conduct the comparison experiment, and the results are listed in Table 2 and Table 3. It can be seen that the overall trend is similar with that on the whole dataset. 2D OIQA has the worst performance among all the models, and 3D IQA models present a promising result. Our model holds the best performance 3 times under different types of distortions, and achieves the best performance on all criteria on symmetrically and asymmetrically distortions, which shows

TABLE 1. Performance Comparison on SOLID Database

Type	Method	PLCC ↑	SRCC ↑	RMSE ↓
2D IQA	PSNR [3]	0.629	0.603	0.789
	SSIM [4]	0.882	0.888	0.478
	MS-SSIM [5]	0.773	0.755	0.643
	FSIM [6]	0.889	0.883	0.465
	VSI [7]	0.881	0.873	0.479
2D OIQA	S-PSNR [25]	0.593	0.567	0.816
	WS-PSNR [27]	0.585	0.559	0.823
	CPP-PSNR [26]	0.593	0.566	0.817
3D IQA	Chen *et al.* [22]	0.853	0.827	0.530
	W-SSIM [8]	0.893	0.891	0.457
	W-FSIM [8]	0.889	0.885	0.464
	SINQ [10]	0.810	0.779	0.460
3D OIQA	VP-BSOIQA [39]	0.853	0.842	0.411
	Chai-SOIQE [9]	0.879	0.872	0.372
	SOIQE [12]	0.927	0.924	0.383
	BPGI (**ours**)	**0.931**	**0.929**	**0.366**

TABLE 2. Performance Comparison for Different Distortion Types on SOLID Database

Type	Method	PLCC ↑		SRCC ↑		RMSE ↓	
		JPEG	BPG	JPEG	BPG	JPEG	BPG
2D IQA	PSNR [3]	0.564	0.740	0.538	0.673	0.901	0.624
	SSIM [4]	0.907	0.857	0.893	0.879	0.460	0.477
	MS-SSIM [5]	0.841	0.730	0.833	0.687	0.591	0.633
	FSIM [6]	0.894	0.896	0.880	0.902	0.490	0.411
	VSI [7]	0.898	0.888	0.885	0.886	0.480	0.426
2D OIQA	S-PSNR [25]	0.515	0.736	0.477	0.660	0.936	0.627
	WS-PSNR [27]	0.505	0.732	0.464	0.658	0.949	0.631
	CPP-PSNR [26]	0.517	0.735	0.475	0.660	0.934	0.628
3D IQA	Chen *et al.* [22]	0.909	0.797	0.904	0.736	0.454	0.559
	W-SSIM [8]	0.905	0.887	0.888	0.879	0.464	0.428
	W-FSIM [8]	0.893	0.933	0.885	0.933	0.492	0.333
	SINQ [10]	0.819	0.821	0.790	0.789	0.499	0.377
3D OIQA	VP-BSOIQA [39]	0.883	0.861	0.852	0.856	0.551	0.373
	Chai-SOIQE [9]	0.898	0.891	0.882	0.885	0.310	0.378
	SOIQE [12]	0.933	**0.955**	0.928	**0.939**	0.393	**0.275**
	BPGI (**ours**)	**0.940**	0.948	**0.931**	0.937	**0.362**	0.347

the stability and effectiveness of our model in all distortion types. Although SOIQE model presents a competition results on different distortion types, it has worse performance than our model on symmetrically and asymmetrically distortions. It needs to mention that in real world stereoscopic image usually has the asymmetrically distortion, which makes our model has wider practical application value than SOIQE. Overall. Our model has the best overall performance on different distortion types, which further proves the effectiveness and practicability of our model.

3) Ablation Experiments

To validate the contribution of each component, we also conduct the ablation experiment, and the results are listed in Table 4. Experiment (1) (the second row) removes all MSS modules in our model, and only send the output of the VMamba and ResNet50 to the Bi-LSTM. Experiment (2) and (3) (the third and fourth row) remove the entire binocular branch and monocular branch. Experiment (4) replace the fusion block with concatenation operation, and Experiment (5) replace the Bi-LSTM with concatenation

TABLE 3. Performance Comparison for Symmetrically and Asymmetrically Distorted Images on SOLID Database

Type	Method	PLCC ↑		SRCC ↑		RMSE ↓	
		Sym	Asym	Sym	Asym	Sym	Asym
2D IQA	PSNR [3]	0.791	0.394	0.789	0.354	0.758	0.756
	SSIM [4]	0.944	0.821	0.902	0.814	0.409	0.470
	MS-SSIM [5]	0.869	0.631	0.836	0.615	0.613	0.638
	FSIM [6]	0.930	0.853	0.890	0.847	0.456	0.430
	VSI [7]	0.931	0.834	0.887	0.807	0.454	0.454
2D OIQA	S-PSNR [25]	0.805	0.364	0.766	0.313	0.735	0.766
	WS-PSNR [27]	0.807	0.325	0.762	0.302	0.732	0.778
	CPP-PSNR [26]	0.806	0.334	0.766	0.310	0.734	0.775
3D IQA	Chen *et al.* [22]	0.944	0.767	0.890	0.700	0.411	0.528
	W-SSIM [8]	0.944	0.834	0.902	0.832	0.409	0.454
	W-FSIM [8]	0.930	0.845	0.890	0.842	0.456	0.440
	SINQ [10]	0.847	0.766	0.792	0.745	0.365	0.493
3D OIQA	VP-BSOIQA [39]	0.894	0.831	0.853	0.776	0.396	0.414
	Chai-SOIQE [9]	0.887	0.848	0.885	0.832	0.337	0.433
	SOIQE [12]	0.970	0.867	0.931	0.866	0.301	0.411
	BPGI (**ours**)	**0.972**	**0.898**	**0.934**	**0.887**	**0.296**	**0.395**

TABLE 4. The Results of the Ablation Experiments

Method	PLCC ↑	SRCC ↑	RMSE ↓
BPGI (**ours**)	**0.931**	**0.929**	**0.366**
w/o Trans. Attn.	0.929	0.924	0.367
w/o Summations	0.920	0.918	0.370
Fuse replaced by Concat	0.868	0.872	0.537
LSTM replaced by Concat	0.896	0.885	0.462

operation. Experiment (1) proves that the MSS module well strengthens the feature representation ability of our model and improve the overall performance, which is one of our contributions. Experiment (2) proves that the binocular visual features play more important role than the monocular visual features. Experiment (3) indicates that the global binocular visual features are well complement with the local binocular visual features, and so does in monocular visual features, which is also one of our contributions. Besides, Experiment (4) indicates that the model simply concatenating binocular and monocular visual features instead of applying the fusion module we designed leads to poor performance, which proves that the fusion model can further improves the performance of our model. What's more, Experiment (5) leads to poor performance, which proves the reasonable of applying Bi-LSTM to obtain refined feature fusion. Overall, the above results prove the reasonable of our model, and can be applied to predict the quality of SOI.

V. CONCLUSION

In this paper, we propose an effective SOIQA model guided by human brain perceptual mechanism. Four channels visual features are extracted, including global binocular visual features, local binocular visual features, global monocular visual features, and local monocular visual features. Considering the human brain cognition procedure, an effective fusion module (Bi-LSTM) is adopted to capture the inherent relationship. All the features are then interactive fused together to map the quality score, and experimental results proves the effectiveness of our model. Overall, our model can be applied to evaluate SOI quality, and is consistent with human visual perception. In the future, we will pay our attention on more effective SOIQA models, and extent to study the quality assessment method of stereoscopic omnidirectional video.

REFERENCES

[1] Z. L. Wan, X. Yan, and Z. Y. L. et al., "Blind stereoscopic omnidirectional image quality assessment via a binocular viewport hypergraph convolutional network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2025, 2025.

[2] Z. L. Wan, Q. S. Yang, and Z. Y. L. et al., "Dual-stream perception-driven blind quality assessment for stereoscopic omnidirectional images," in *ACM Multimedia*, 2024.

[3] A. Hor'e and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *Proceedings of 2010 20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369.

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, 2003, pp. 1398–1402.

[6] 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, no. 8, pp. 2378–2386, 2011.

[7] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.

[8] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3d images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[9] X. Chai, F. Shao, Q. Jiang, X. Meng, and Y.-S. Ho, "Monocular and binocular interactions oriented deformable convolutional networks for blind quality assessment of stereoscopic omnidirectional images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3407–3421, 2022.

[10] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Processing: Image Communication*, vol. 58, pp. 287–299, 2017.

[11] C. Tian, F. Shao, X. Chai, Q. Jiang, L. Xu, and Y. Ho, "Viewport sphere-branch network for blind quality assessment of stitched 360° omnidirectional images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2546–2560, 2023.

[12] Z. Chen, J. Xu, C. Lin, and W. Zhou, "Stereoscopic omnidirectional image quality assessment based on predictive coding theory," *IEEE Journal of Selected Topics in Signal Processing*, 2019.

[13] J. Xu, Z. Luo, W. Zhou, W. Zhang, and Z. Chen, "Quality assessment of stereoscopic 360-degree images from multi-viewports," in *The 34th Picture Coding Symposium (PCS): 2019*, 2019.

[14] M. N. Shadlen and D. Shohamy, "Decision making and sequential sampling from memory," *Neuron*, vol. 90, no. 5, pp. 927–939, 2016.

[15] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. Akar, "Towards compound stereo-video quality metric: a specific encoder-based framework," in *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, 2006, pp. 218–222.

[16] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *Eur. J. Image Video Process.*, pp. 1–13, 2008.

[17] W. Zhou, Z. Chen, and W. Li, "Dual-stream interactive networks for no-reference stereoscopic image quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3946–3958, 2019.

[18] Y. Chang, S. Li, J. Jin, A. Liu, and W. Xiang, "Stereo image quality assessment considering the difference of statistical feature in early visual pathway," *Journal of Visual Communication and Image Representation*, vol. 89, p. 103643, 2022.

[19] Y. Liu, B. Huang, H. Yu, and Z. Zheng, "No-reference stereoscopic image quality evaluator based on human visual characteristics and relative gradient orientation," *Journal of Visual Communication and Image Representation*, vol. 81, Nov. 2021.

[20] L. Shen, X. Chen, Z. Pan, K. Fan, F. Li, and J. Lei, "No-reference stereoscopic image quality assessment based on global and local content characteristics," *Neurocomputing*, vol. 424, pp. 132–142, 2021.

[21] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1940–1953, 2013.

[22] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[23] S. Li and M. Wang, "No-reference stereoscopic image quality assessment based on convolutional neural network with a long-term feature fusion," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, pp. 318–321.

[24] Y. Liu, B. Huang, G. Yue, J. Wu, X. Wang, and Z. Zheng, "Two-stream interactive network based on local and global information for no-reference stereoscopic image quality assessment," *Journal of Visual Communication and Image Representation*, vol. 87, 2022.

[25] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proceedings of 2015 IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 31–36.

[26] V. Zakharchenko, K. P. Choi, and J. Park, "Quality metric for spherical panoramic video," in *Proceedings of Optical Engineering and Applications*, 2016.

[27] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1408–1412, 2017.

[28] S. Chen, Y. Zhang, Y. Li, Z. Chen, and Z. Wang, "Spherical structural similarity index for objective omnidirectional video quality assessment," in *IEEE International Conference on Multimedia and Expo*, 2018, pp. 1–6.

[29] H. G. Kim, H.-T. Lim, and Y. M. Ro, "Deep virtual reality image quality assessment with human perception guider for omnidirectional image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 917–928, 2020.

[30] H. Jiang, G. Jiang, M. Yu, Y. Zhang, Y. Yang, Z. Peng, F. Chen, and Q. Zhang, "Cubemap-based perception-driven blind quality assessment for 360 - degree images," *IEEE Transactions on Image Processing*, vol. 30, pp. 2364–2377, 2021.

[31] Y. Liu, X. Yin, Y. Wang, Z. Yin, and Z. Zheng, "Hvs-based perception-driven no-reference omnidirectional image quality assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[32] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 5, pp. 1724–1737, 2021.

[33] Y. Zhou, Y. Sun, L. Li, K. Gu, and Y. Fang, "Omnidirectional image quality assessment by distortion discrimination assisted multi-stream network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1767–1777, 2022.

[34] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan, and S. Ma, "Mc360iqa: A multi-channel cnn for blind 360-degree image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 64–77, 2020.

[35] L. Yang, M. Xu, T. Liu, L. Huo, and X. Gao, "Tvformer: Trajectory-guided visual quality assessment on 360° images with transformers," in *ACM International Conference on Multimedia*, 2022, pp. 799–808.

[36] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 5, pp. 1724–1737, 2020.

[37] J. Yan, Z. Tan, Y. Fang, J. Rao, and Y. Zuo, "Max360iq: Blind omnidirectional image quality assessment with multi-axis attention," *Pattern Recognition*, vol. 162, 2025.

[38] J.Yang, T. Liu, B. Jiang, H. Song, and W. Lu, "3d panoramic virtual reality video quality assessment based on 3d convolutional neural networks," *IEEE Access*, vol. 6, pp. 38 669–38 682, 2018.

[39] Y. Qi, G. Jiang, M. Yu, Y. Zhang, and Y.-S. Ho, "Viewport perception based blind stereoscopic omnidirectional image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3926–3941, 2021.

[40] J. Xu, C. Lin, W. Zhou, and Z. Chen, "Subjective quality assessment of stereoscopic omnidirectional image," in *Proceedings of Advances in Multimedia Information Processing, PCM 2018*, 2018, pp. 589–599.