Image Quality Assessment Based on Feature Fusion and Local Adaptation

Minjuan GAO[1]*, Yankang LI[2], Xuande ZHANG[3]
School of Computer Science and Technology, Taiyuan Normal University, Shanxi 030600, China[1, 2]
School of Electrical Information and Artificial Intelligence, Shaanxi University of Science & Technology, Xian 710021, China[3]

***Abstract*—No-reference image quality assessment (NR-IQA) aims to evaluate the perceptual quality of images without access to corresponding reference images and has broad applications in real-world image processing scenarios. However, existing NR-IQA methods often suffer from limited accuracy and generalization, especially under complex and diverse distortion types. To address this challenge, we propose Inc-LAENet, a novel NR-IQA framework that leverages multi-scale deep residual representations, integrates feature fusion mechanisms, and incorporates a local adaptive perception module to achieve improved assessment accuracy and generalization. Specifically, ResNet50 is employed to extract hierarchical residual features, an enhanced Inception-style module (Inc-s) strengthens sensitivity to various distortion patterns, and a lightweight local adaptive extraction module efficiently captures fine-grained structural information. Extensive experiments demonstrate the effectiveness of the proposed method, achieving SROCC values of 0.967 and 0.935 on the synthetic distortion datasets LIVE and CSIQ, and 0.852 and 0.898 on the authentic distortion datasets LIVEC and KonIQ-10k, respectively. These results confirm that Inc-LAENet provides a robust and efficient solution for NR-IQA tasks across both synthetic and real-world scenarios.**

***Keywords*—No-reference image quality assessment; deep learning; multi-scale; feature fusion; local adaptation**

I. INTRODUCTION

Millions of digital images are shared on social media platforms every day. Images may be distorted to a certain extent during the acquisition and transmission process. In order to ensure an acceptable level of visual experience, it is crucial to conduct reliable and accurate image quality assessments (IQA) that have a good correlation with human perception. IQA can be divided into two categories: subjective evaluation and objective evaluation [1]. Subjective evaluation is the observer's own judgment of the image quality based on his or her visual experience. The objective evaluation is to design an algorithm to simulate the human visual system to analyze the relevant features of the image and give a corresponding quality score. Objective evaluation can be divided into three basic types according to whether a reference image is used: full-reference IQA [2] [3] (FR-IQA), reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA). NR-IQA [4] does not rely on the original reference image, but can predict its quality score based on the image's own features. This feature makes it more flexible and practical in practical applications, and therefore has broad application prospects in visual tasks such as image quality assessment [5], [6].

Although deep learning has made significant progress in the field of no-reference image quality assessment (NR-IQA) in recent years. However, this field still faces many challenges. First, image quality often depends on the comprehensive analysis of multi-scale features, while most existing methods are limited to feature extraction at a single scale, making it difficult to simultaneously capture local details and global structural information of the image. Second, different regions in the image have different effects on quality assessment, but existing methods usually fail to effectively model this non-uniformity, making it difficult for the model to focus on key areas that are critical to quality perception.

In order to solve the problem of insufficient accuracy and generalization ability of existing NR-IQA methods in complex distortion scenes, this study proposes an efficient NR-IQA method "Inc-LAENet" that integrates multiple modules. Its main contributions are as follows:

1) A NR-IQA framework (Inc-LAENet) that integrates deep residual features and local adaptive perception is proposed to achieve efficient modeling of global and local features.

2) The feature enhancement expression module (Inc-s) is designed, combining the dilated convolution and multi-scale convolution structure to improve the perception of local details and global semantic information of the image.

3) Introduce the Lightweight Adaptive Extraction (LAE) module, which focuses on the more quality-sensitive areas in the image through an adaptive weight mechanism, effectively improving the accuracy and robustness of the prediction.

The remainder of this study is organized as follows: Section II reviews related work in the field of NR-IQA. Section III introduces the overall architecture of Inc-LAENet, along with detailed descriptions of the Semantic Feature Extraction Network, the Feature Enhancement Module, and the Lightweight Local Adaptive Extraction Module. Section IV describes the experimental setup, including datasets, evaluation metrics, implementation details, and presents both comparative and ablation experiments. Section V summarizes the key findings. Section VI provides further discussions on the performance advantages, comparisons with state-of-the-art methods, generalization capabilities, and potential directions for future work.

Fundamental Research Programme of Shanxi Province (202203021222237). Supported by Scientific and Technologial Innovation Programs of Higher Education Institutions in Shanxi (2022L401).

II. Related Work

Traditional no-reference image quality assessment relies on the statistical characteristics and visual properties of the image itself, and requires manual extraction of image features, which is difficult and complex to implement, such as natural scene statistics (NSS) features, gradient features, texture features, etc. Mittal et al. proposed BRISQUE [7], which extracts the local normalized brightness coefficient (MSCN) of the image and its statistical characteristics, and uses the support vector machine (SVM) for quality prediction. Saad et al. proposed a NR-IQA method called BLIINDS [8], which uses the statistical information of the local discrete cosine transform (DCT) coefficients to extract features and then uses a multivariate Gaussian model to evaluate image quality. Furthermore, Saad et al. proposed BLIINDS-II [9] to evaluate image quality by optimizing the statistical characteristics of DCT. Zhang et al. proposed a general IQA method ILNIQE [10], which uses the mean minus, the contrast normalization coefficient, and the gradient statistic to extract features. However, hand-designed features often rely on prior knowledge, and have difficulty in fully capturing images when faced with complex feature distortions, and have limited generalization capabilities when faced with some natural distortions.

In recent years, deep learning has become popular [11] due to the powerful feature extraction capabilities of convolutional neural networks (CNN) [12]. Convolutional neural networks combine feature extraction with the learning process, avoiding the limitations of manually extracted features and showing good evaluation results. They have also been applied to NR-IQA [13-16]. For example, Kang et al. [17] proposed an NR-IQA method based on convolutional neural networks, which extracts features from images and predicts quality scores by training the network. Kang et al. [18] proposed a multi-task CNN for simultaneous image quality estimation and distortion type identification. Liu et al. [19] proposed an NR-IQA method called RankIQA, which solves the problem of the Siamese network's limitation on the size of the IQA dataset and evaluates quality by ranking image quality.

III. Proposed Method

This study proposes a no-reference image quality assessment method based on multi-scale feature fusion, as shown in Fig. 1. The method consists of four core modules: a backbone network (used to extract multi-level semantic information), a feature enhancement module, a multi-level semantic feature fusion module, and a quality prediction network (used to output the final image quality score).

During the training process, the pre-trained deep residual network ResNet50 [20] is first used as the feature extraction network to extract semantic features x1, x2, x3, and x4 at different levels from the distorted RGB image. Subsequently, x1, x2, and x3 are downsampled by the LAE module to enhance the local information perception capability. At the same time, the features of the four scales are deeply optimized by the Inc-s module. Next, the processed features are input into the global pooling layer and spliced in the feature dimension to fully integrate multi-scale information. Finally, the fused features are input into the quality regression network to generate the final image quality score.

A. Semantic Feature Extraction Network

ResNet50 is a variant of the residual network proposed by He et al. to solve the problems of gradient vanishing and gradient exploding as the network depth increases. ResNet50 consists of 50 layers of convolutional networks, and its structure mainly includes convolutional layers, batch normalization layers, ReLU activation functions, and fully connected layers. The performance of each residual block is shown in Formula (1):

$$F(x) = H(x) - x \tag{1}$$

where, $H(x)$ represents the original mapping function and $F(x)$ represents the residual mapping. During the forward propagation process, the output of the residual block is Formula (2). Its structure diagram is shown in Fig. 2.

$$Y = F(x) + x \tag{2}$$

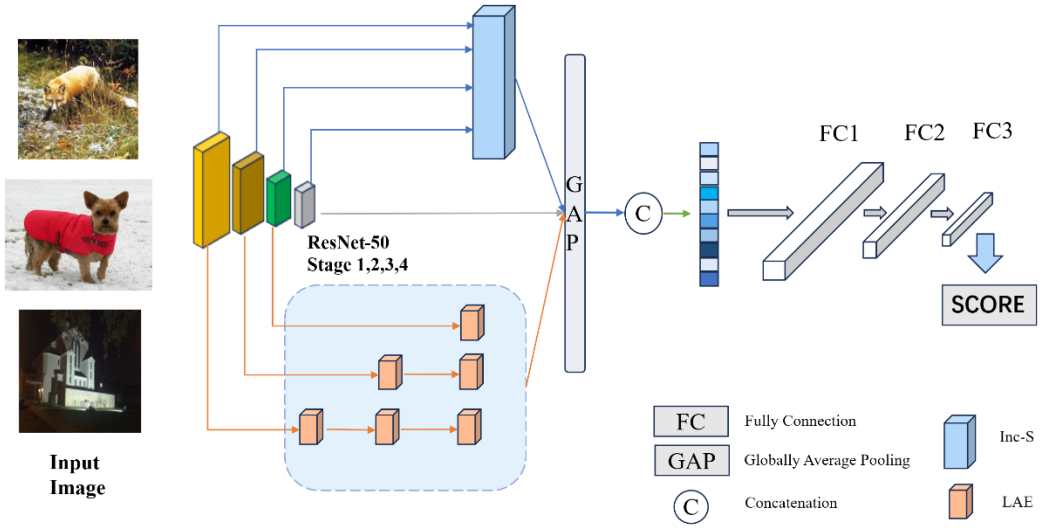

Fig. 1. Overall framework.

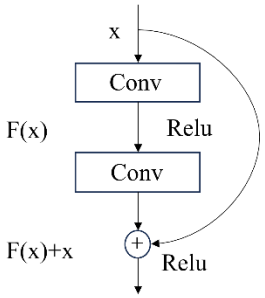

Fig. 2. Residual block structure.

In order to capture rich contextual information and deep semantic features, we use ResNet-50 as the backbone network to extract multi-level features of the image. These features come from the output of Conv2_x, Conv3_x, Conv4_x, and Conv5_x stages. The features of each stage contain information of different spatial resolutions and semantic levels. The multi-scale features are shown in Formula (3):

$$\Phi(I) = \left\{ f_l(I) \middle| l \in \{2,3,4,5\} \right\} \quad (3)$$

where, *I* denotes the input image, and *d* represents the feature map at the *l*-th layer.

B. Feature Enhancement Module

Human vision has a multi-level information processing mechanism [21]. In order to improve the network's ability to understand the content of distorted images, inspired by the inception module [22], this study designed a feature enhancement method that integrates a multi-scale attention mechanism, as shown in Fig. 3.

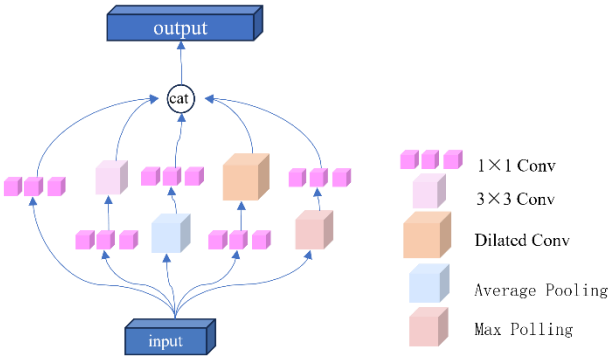

Fig. 3. Inc-s module.

Specifically, the method first extracts hierarchical features through the backbone network to obtain feature representations at four scales, and then inputs them into a new feature enhancement module to optimize feature expression capabilities. This module draws on the idea of a spatial pyramid structure and deploys multi-scale convolution operations (including 1×1, 3×3, and 5×5 convolutions) and pooling operations in parallel to achieve collaborative modeling of local details and global semantics, thereby enhancing the perception of image quality.

In addition, to further optimize computational efficiency and model performance, the feature enhancement module introduces two key technical improvements during implementation: 1) Atrous convolution replaces traditional large kernel convolution: 3×3 atrous convolution kernels are used to replace traditional 5×5 convolutions (the dilation factor is 1 in this experiment), which reduces the number of parameters while ensuring equivalent receptive fields; 2) Adding an average pooling branch. An additional average pooling branch is introduced to capture the global statistical information of the image, complementing other convolution branches. Average pooling can smooth feature distribution and reduce the impact of outliers and noise, thereby reducing the interference of local distortion on model prediction and improving the generalization ability of unknown distortions. In addition, since average pooling does not rely on trainable parameters, this branch enriches feature expression while increasing model complexity minimally, which helps to reduce the risk of overfitting. Formula (4) describes the output form of the Inception module:

$$F_{out} = \left[F_{1\times1},\ F_{3\times3},\ F_{dilated},\ F_{max},\ F_{avg} \right] \quad (4)$$

Among them, $F_{1\times1}$, $F_{3\times3}$,and $F_{dilated}$ represent convolution branches of different scales, and $F_{\max}$ and F_{avg} represent maximum pooling and flat pooling branches, respectively.

C. Lightweight Local Adaptive Extraction Module

In IQA tasks, accurately extracting local and global features of an image is the key to evaluating the degree of image distortion. In order to effectively preserve the edge and local detail information of the image, this method introduces the LAE module [23], as shown in Fig. 4. Compared with traditional convolution methods, the LAE module has obvious advantages in terms of the number of parameters and computational complexity, while extracting more semantically rich features. IQA is highly sensitive to edges, textures, and local details, while traditional convolutions tend to lose this key information during downsampling. The LAE module effectively retains and enhances this information through parallel feature extraction and adaptive information aggregation.

The LAE module is designed with a dual-branch architecture that shares parameters, combining Group Convolution and adaptive feature aggregation, which not only reduces the number of parameters but also retains key information in the image. By dividing the feature map into N groups for convolution operations, the number of parameters and the amount of calculation are reduced to 1/N compared to traditional convolution.

LAE introduces adaptive feature aggregation to alleviate the loss of edge and detail information caused by downsampling. Specifically, it changes the dimension of the feature map from four dimensions (batch, channel, height, width) to five dimensions (batch, channel, height, width, n), where “n” represents the sampling factor. The adaptive path aggregates four adjacent pixels and calculates their weights through average pooling and convolution operations, and uses softmax normalization to ensure that pixels with higher information entropy are retained in the channel dimension. Formula (5) describes the output mode of the LAE module. Given the input $X \in R^{h\times w\times c}$, the output of LAE is $Y \in R^{\frac{h}{2}\times\frac{w}{2}\times c}$.

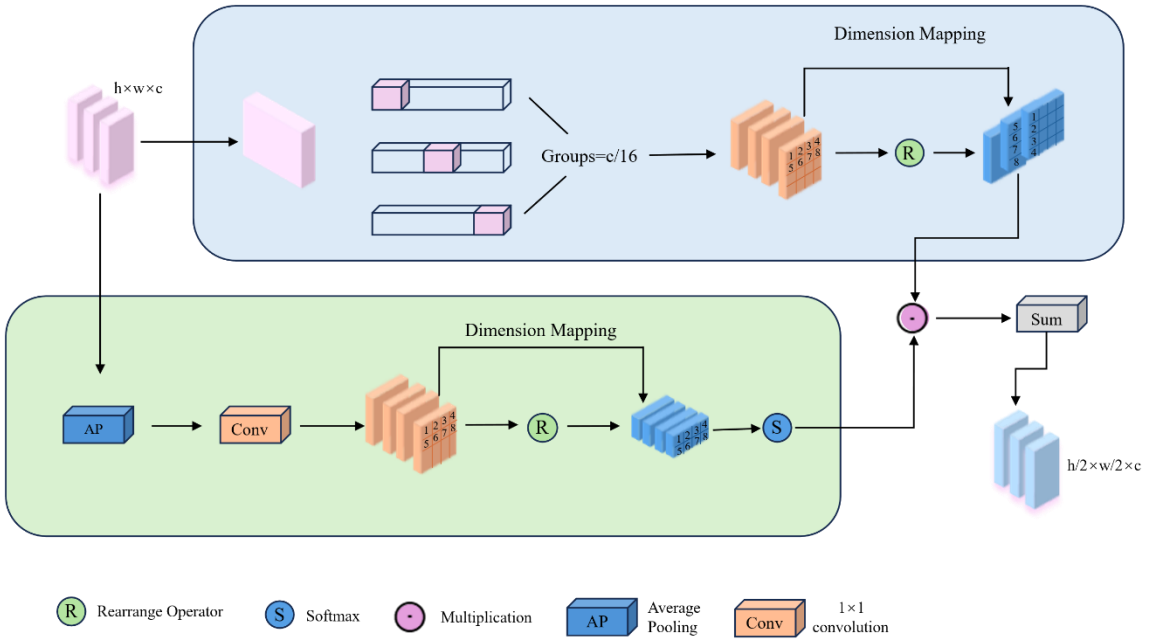

Fig. 4. Lightweight Local Adaptive Extraction module (LAE).

$$Y = \sum \begin{pmatrix} Rearrange(GroupConv(x)) \\ Softmax(Rearrange(Conv(AP(X)))) \end{pmatrix} \quad (5)$$

where, $X \in R^{h\times w\times c}$ is the input feature map, $Y \in R^{\frac{h}{2}\times\frac{w}{2}\times c}$ is the output feature map, $GroupConv(X)$ performs group convolution on the input feature map, $Rearrange(\bullet)$ compresses the spatial dimension to the channel dimension, $AP(X)$ performs average pooling on the input, and $Softmax(\bullet)$ generates adaptive weights.

IV. Experiments

A. Experimental Environment and Dataset

The model proposed in this study is implemented using Python 3.9 programming language and PyTorch 2.5.1 deep learning framework, which has strong compatibility and scalability. In the training and testing phases, the experiment uses NVIDIA GeForce RTX 4090 graphics processor as the main computing platform, combined with CUDA 12.1 toolkit to achieve efficient parallel computing support. This configuration not only effectively improves the running speed of the model, but also shortens the training cycle to a certain extent, enhancing the controllability of the experimental process and the repeatability of the results.

In this experiment, we tested and evaluated on four public IQA datasets (two synthetic distortions and two real distortions). For synthetic distortion, we used LIVE [24] and CSIQ [25]. The LIVE dataset contains 779 distorted images, covering five distortion types and five distortion levels, and the CSIQ dataset contains 866 images, covering six distortion types and four to five distortion levels. For real distortion datasets, we used CLIVE [26] and KonIQ-10k [27]. LIVE-C is a no-reference image quality assessment dataset containing 1162 real distorted images taken by different multimedia devices, and KonIQ-10k is a large-scale no-reference dataset containing 10,073 high-quality real distorted images. Table I is a summary of the databases used in this experiment.

TABLE I. IQA Dataset Used for Comparative Experiments

Databases	#of Dist. Images	#of Dist. Types	Distortions Type
LIVE	799	5	synthetic
CSIQ	866	6	synthetic
LIVEC	1162	-	authentic
KonIQ-10k	10073	-	authentic

B. Evaluation Indicators

For the performance evaluation of the proposed method, two commonly used standards are used. They are Spearman's rank order correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC). The range of SROCC and PLCC is from -1 to 1, and the larger the absolute value, the better the performance. SROCC is a non-parametric correlation coefficient used to measure the monotonic relationship between two variables (i.e., whether one variable increases or decreases as the other variable increases). It is based on the rank (sorting) of the variables rather than the original values, so it is not sensitive to outliers. The calculation formula of SROCC is shown in Formula (6):

$$SROCC = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \quad (5)$$

where, d_i represents the level difference between the true value and the predicted value of the distorted image, and n

represents the number of distorted images. The calculation formula of PLCC is shown in Formula (7):

$$\text{PLCC}=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (6)$$

where, X_i and Y_i represent the predicted score and true score of the distorted image, respectively. $\bar{X}$ and $\bar{Y}$ represent the mean of the predicted score and the true label score.

C. Experimental Details

This experiment follows the training strategy of the existing IQA algorithm. During training, 30 224×224 pixel patches are randomly selected from each distorted image, and random data augmentation is performed, and their quality scores are consistent with the source image. The Adam optimizer is used for training, with a weight decay coefficient of 5×10-4, a maximum number of training rounds (epochs) of 5, and a mini-batch size of 32. The initial learning rate is set to 2×10-5 and is decreased by 10% after each round of training. In the test phase, 30 patches are randomly selected from each test image, their quality scores are predicted, and the average is taken as the final score. The experiment uses the same settings and randomly divides the dataset into 80% training set and 20% test set using 10 different seeds. The test data does not participate in the training, and the median of the SROCC and PLCC of the 10 groups of experiments is taken as the result. In order to optimize the model, capture subtle changes in image quality, reduce the impact of outliers, accelerate convergence and improve stability, the SmoothL1Loss loss function is used in training, as shown in Formula (8):

$$\text{SmoothLoss}=\begin{cases}0.5(x-y)^2, & if\ |x-y|<\beta\\ \beta(|x-y|-0.5\beta), & otherwise\end{cases} \quad (7)$$

where, x is the prediction score of the model and y is the true label value of the distorted image.

D. Comparative Experiment

In order to fully verify the effectiveness and robustness of the proposed model, we conducted comparative experiments with six representative image quality evaluation methods on four mainstream NR-IQA datasets, covering traditional methods (such as BRISQUE and ILNIQE) and recent deep learning methods (such as WaDIQaM, DBCNN, HyperIQA and TReS). As shown in Table II and Table III, on the real distortion datasets KONIQ-10k and LIVEC, our method achieved good performance, reaching SROCC 0.898 and PLCC 0.909 on KONIQ-10k, and SROCC 0.852 and PLCC 0.872 on LIVEC, respectively, which is better than traditional methods such as BRISQUE and ILNIQE, and is comparable to advanced models such as HyperIQA and TReS, with only a small gap or close performance, reflecting the high generalization ability of the model in complex real scenes. On synthetic distortion datasets such as LIVE and CSIQ, our method also shows strong robustness. On the LIVE dataset, we achieved a high score of 0.967 for both SROCC and PLCC, which is close to or even slightly better than other mainstream methods, indicating that the model has accurate perception capabilities when dealing with standard synthetic distortion types (such as blur, compression, noise, etc.); on the CSIQ dataset, our method's SROCC and PLCC are 0.935 and 0.944, respectively, which also surpasses most baseline models and is second only to some highly optimized models such as DBCNN, but the overall performance is still stable and competitive. In order to more intuitively show the performance differences between the methods, we introduced the bar charts shown in Fig. 5 to Fig. 8, which visualize the evaluation results of the LIVEC, KONIQ-10k, LIVE and CSIQ datasets, respectively. From the figures, we can clearly observe the leading trend of the proposed method in various indicators, which further enhances the intuitiveness and persuasiveness of the experimental results.

TABLE II. PERFORMANCE COMPARISON OF EACH EVALUATION METHOD ON THE REAL DISTORTION DATASET

Methods	LIVEC		KONIQ-10k	
	SROCC	***PLCC***	***SROCC***	***PLCC***
BRISQUE	0.608	0.629	0.701	0.705
ILNIQE	0.432	0.508	0.507	0.527
WaDIQaM	0.671	0.681	0.806	0.807
DBCNN[28]	0.851	0.869	0.877	0.884
HyperIQA[29]	0.855	0.875	0.901	0.910
TReS[30]	0.846	0.877	0.905	0.908
Ours	0.852	0.872	0.898	0.909

TABLE III. PERFORMANCE COMPARISON OF EACH EVALUATION METHOD ON THE SYNTHETIC DISTORTION DATASET

Methods	LIVE		CSIQ	
	SROCC	***PLCC***	***SROCC***	***PLCC***
BRISQUE	0.929	0.944	0.748	0.812
ILNIQE	0.898	0.903	0.815	0.854
WaDIQaM	0.954	0.963	0.844	0.852
DBCNN	0.968	0.971	0.946	0.959
HyperIQA	0.962	0.966	0.923	0.942
TReS	0.968	0.969	0.922	0.942
Ours	0.967	0.967	0.935	0.944

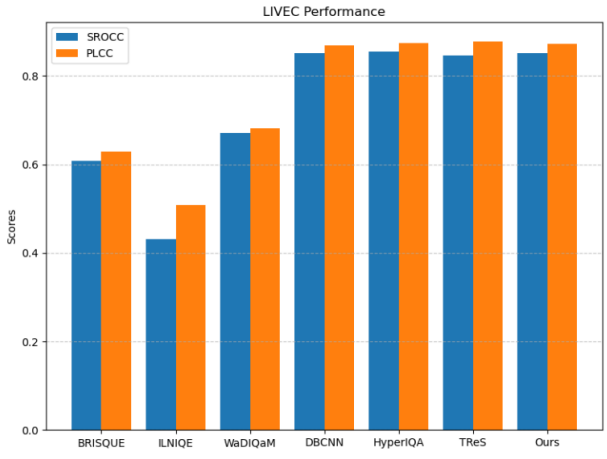

Fig. 5. Results of the LIVEC.

Fig. 6. Results of the KONIQ-10k.

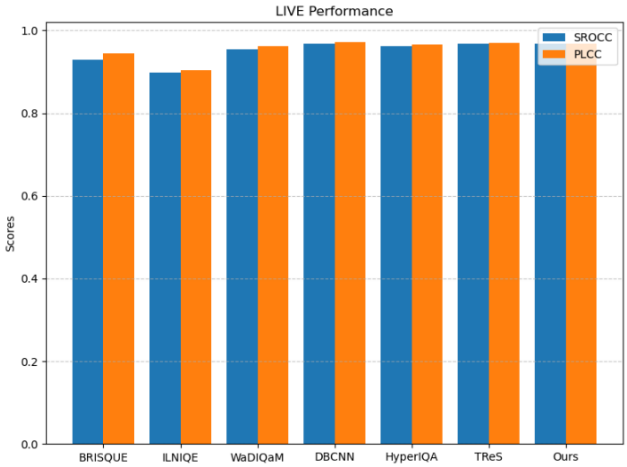

Fig. 7. Results of the LIVE.

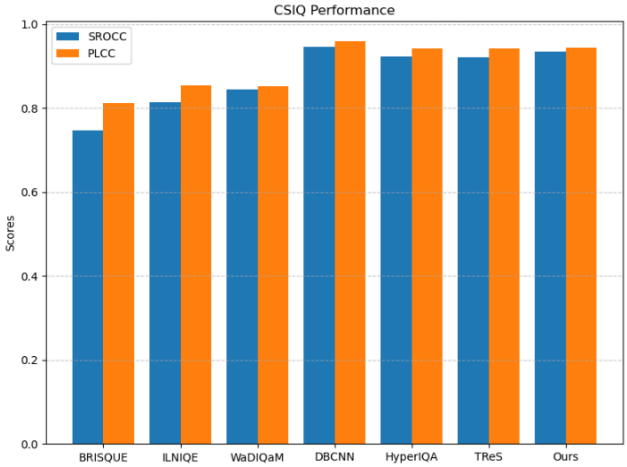

Fig. 8. Results of the CSIQ.

E. Ablation Experiment

In order to verify the contribution of each module to this model in this experiment, an ablation experiment was designed and conducted. The ablation experiment also adopted the above experimental rules, conducted ten experiments, and took the median of SROCC and PLCC as the experimental results for testing on the CSIQ and LIVEC databases. Table IV shows the statistical results of different modules to help understand the impact of each module on the performance of the model. The baseline model uses ResNet-50. From Table IV, it is not difficult to see that by integrating the feature enhancement module Inc-s into the basic model, SROCC and PLCC are improved on the CSIQ dataset and LIVE-C dataset. By integrating the lightweight local adaptive module LAE into the baseline model, all evaluation indicators on the two datasets are also improved. Finally, both modules are integrated into the baseline model to form our final model. It can be seen that our final model can show the best performance.

TABLE IV. PERFORMANCE COMPARISON OF EACH MODULE

Baseline	Inc-s	LAE	CSIQ		LIVE-C	
			SROCC	***PLCC***	***SROCC***	***PLCC***
√			0.903	0.928	0.817	0.850
√	√		0.921	0.940	0.838	0.856
√		√	0.915	0.933	0.845	0.866
√	√	√	0.925	0.944	0.852	0.872

V. SUMMARIZE

This study proposes an NR-IQA model Inc-LAENet, that combines convolutional neural networks, Inc-s, and a lightweight local adaptive extraction (LAE) module. ResNet-50 extracts global and local features of the image through a residual structure, greatly enhancing the model's perception of complex distortions. Inc-s strengthens the expression of features by combining multi-scale convolution and pooling. LAE efficiently extracts local feature information while reducing computational overhead, thereby improving local feature sensitivity. Experimental results show that the model Inc-LAENet proposed in this study performs well on multiple public datasets, with SROCCs of 0.967 and 0.935 on two synthetic distortion datasets, LIVE and CSIQ, respectively, and SROCCs of 0.852 and 0.898 on two real distortion datasets, LIVEC and Koniq-10k, respectively, proving its effectiveness in NR-IQA. Future research can further explore deeper feature fusion, loss function optimization, and other technologies to further improve the performance and application potential of the model.

VI. DISCUSSION

This section will conduct an in-depth analysis of the experimental results of the proposed method, focusing on its design concept, performance advantages, scope of application and existing deficiencies, and further highlighting the practical value and potential impact of this study.

A. Analysis of the Causes of Performance Advantages

The experimental results show that the proposed method has achieved excellent performance on multiple public image quality assessment datasets. The overall performance is comparable to the current mainstream image quality assessment model, and even slightly exceeds it in some indicators. Further analysis of the reasons for its performance improvement is mainly reflected in the following two aspects: on one hand, the

Inc-s feature enhancement module designed in this study effectively integrates multi-scale contextual information, enabling the model to focus on the global structure and local details of the image at the same time, thereby improving the comprehensiveness of feature expression; on the other hand, the introduction of a lightweight local adaptive enhancement module (LAE) enhances the model's sensitivity to local areas of the image, especially distorted areas such as noise and compression artifacts, and shows stronger discrimination capabilities in complex distorted scenes. The two modules work together in parallel, effectively improving the model's global perception and local modeling capabilities in the task of no-reference image quality assessment, thereby significantly enhancing the generalization and stability of the model.

B. Generalization Ability and Application Prospects

It is not difficult to find from the experimental results that the method proposed in this study shows relatively consistent performance in both real distortion and synthetic distortion scenarios, indicating that it has good robustness in dealing with various types of image distortion. This result not only verifies the applicability of the method on multiple mainstream no-reference image quality assessment (NR-IQA) datasets but also further highlights its potential in practical engineering applications. This method can be widely used in many typical scenarios such as image compression quality assessment, image enhancement effect verification, and image transmission stability detection, and has good promotion prospects and practical value.

C. Model Limitations and Future Development Directions

Although the method proposed in this study has achieved excellent performance on multiple typical datasets and has shown good generalization ability and practical application potential, there is still room for further exploration in terms of model optimization. At present, this method mainly relies on convolutional neural networks for feature extraction and quality regression, which can effectively capture the structural information and local distortion characteristics in the image. However, when faced with image content with complex semantic relationships or strong contextual dependencies, the prediction results of the model may still have certain deviations on individual samples.

To further enhance the model's ability to understand complex image quality degradation scenarios, subsequent research may consider introducing structures such as self-attention mechanisms or Transformers to improve the modeling effect of global semantic information. In addition, in more challenging cross-database evaluation or zero-shot image quality prediction tasks, it is necessary to explore more efficient transfer learning strategies or unsupervised optimization solutions to further improve the robustness and adaptability of the model. These improvement directions not only help to improve the theoretical system of the model, but also lay the foundation for its application in practical scenarios.

REFERENCES

[1] Hu Z, Yang G, Du Z, et al. No-reference image quality assessment based on global awareness[J]. Plos one, 2024, 19(10): e0310206.

[2] Chen D , Wu T , Ma K ,et al.Toward Generalized Image Quality Assessment: Relaxing the Perfect Reference Quality Assumption[J]. 2025.

[3] Lan X , Jia F , Zhuang X ,et al.Hierarchical degradation-aware network for full-reference image quality assessment[J].Information Sciences, 2025, 690(000).DOI:10.1016/j.ins.2024.121557.

[4] Mao Q , Liu S , Li Q ,et al.No - Reference Image Quality Assessment: Past, Present, and Future[J].Expert Systems, 2025, 42(3).DOI:10.1111/exsy.13842.

[5] Yang C, Liu Y, Li D, et al. Exploring vulnerabilities of no-reference image quality assessment models: A query-based black-box method[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024.

[6] Mao Q, Liu S, Li Q, et al. No-Reference Image Quality Assessment: Past, Present, and Future[J]. Expert Systems, 2025, 42(3): e13842.

[7] Mittal A , Moorthy A K , Bovik A C .No-Reference Image Quality Assessment in the Spatial Domain[J].IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2012, 21(12):4695.DOI:10.1109/TIP.2012.2214050.

[8] Saad M A , Bovik A C , Charrier C .A DCT Statistics-Based Blind Image Quality Index[J].IEEE Signal Processing Letters, 2010, 17(6):583-586.DOI:10.1109/LSP.2010.2045550.

[9] Saad M A , Bovik A C , Charrier C .Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain[J].IEEE Transactions on Image Processing, 2012, 21(8):3339-3352.DOI:10.1109/TIP.2012.2191563.

[10] Zhang L , Zhang L , Bovik A C .A Feature-Enriched Completely Blind Image Quality Evaluator[J].IEEE Transactions on Image Processing, 2015, 24(8):2579-2591.DOI:10.1109/TIP.2015.2426416.

[11] Srinath S, Mitra S, Rao S, et al. Learning generalizable perceptual representations for data-efficient no-reference image quality assessment[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 22-31.

[12] Shi J, Gao P, Qin J. Transformer-based no-reference image quality assessment via supervised contrastive learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(5): 4829-4837.

[13] Kim J , Lee S .Fully Deep Blind Image Quality Predictor[J].Selected Topics in Signal Processing, IEEE Journal of, 2017, 11(1):206-220.DOI:10.1109/JSTSP.2016.2639328.

[14] Bosse S , Maniry D , Muller K R ,et al.Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment[J].IEEE Transactions on Image Processing, 2017:1-1.DOI:10.1109/TIP.2017.2760518.

[15] Zhao W, Li M, Xu L, et al. A Multi-Branch Network with Multi-Layer Feature Fusion for No-Reference Image Quality Assessment[J]. IEEE Transactions on Instrumentation and Measurement, 2024.

[16] Miao H, Sang Q. McmIQA: Multi-Module Collaborative Model for No-Reference Image Quality Assessment[J]. Mathematics, 2024, 12(8): 1185.

[17] Kang L , Ye P , Li Y ,et al.Convolutional Neural Networks for No-Reference Image Quality Assessment[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition.IEEE, 2014.DOI:10.1109/CVPR.2014.224.

[18] Kang L , Ye P , Li Y ,et al.Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks[J].IEEE, 2015.DOI:10.1109/ICIP.2015.7351311.

[19] Liu X , Weijer J V D , Bagdanov A D .RankIQA: Learning from Rankings for No-Reference Image Quality Assessment[C]//2017 IEEE International Conference on Computer Vision (ICCV).IEEE, 2017.DOI:10.1109/ICCV.2017.118.

[20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[21] Choudhary B K, Sinha N K, Shanker P. Pyramid method in image processing[J]. Journal of Information Systems and Communication, 2012, 3(1): 269.

[22] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[23] Yu Z, Guan Q, Yang J, et al. Lsm-yolo: A compact and effective roi detector for medical detection[J]. arXiv preprint arXiv:2408.14087, 2024.

[24] Sheikh H R, Sabir M F, Bovik A C. A statistical evaluation of recent full reference image quality assessment algorithms[J]. IEEE Transactions on image processing, 2006, 15(11): 3440-3451.

[25] Larson E C, Chandler D M. Most apparent distortion: full-reference image quality assessment and the role of strategy[J]. Journal of electronic imaging, 2010, 19(1): 011006-011006-21.

[26] Ghadiyaram D, Bovik A C. Massive online crowdsourced study of subjective and objective picture quality[J]. IEEE Transactions on Image Processing, 2015, 25(1): 372-387.

[27] Hosu V, Lin H, Sziranyi T, et al. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment[J]. IEEE Transactions on Image Processing, 2020, 29: 4041-4056.

[28] Network A D B C N. Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network[J]. Deep Bilinear Convolutional Neural.

[29] Su S, Yan Q, Zhu Y, et al. Blindly assess image quality in the wild guided by a self-adaptive hyper network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 3667-3676.

[30] Golestaneh S A, Dadsetan S, Kitani K M. No-reference image quality assessment via transformers, relative ranking, and self-consistency[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 1220-1230.