symmetry

MDPI

Article

VGGNet and Attention Mechanism-Based Image Quality Assessment Algorithm in Symmetry Edge Intelligence Systems

Fanfan Shen [1], Haipeng Liu [1], Chao Xu [1,*], Lei Ouyang [2], Jun Zhang [3], Yong Chen [1] and Yanxiang He [4]

[1] School of Computer Science, Nanjing Audit University, Nanjing 211815, China; ffshen@nau.edu.cn (F.S.)
[2] North Information Control Research Academy Group Company Limited, Nanjing 221000, China
[3] College of Software, East China University of Science and Technology, Nanchang 330013, China
[4] School of Computer Science, Wuhan University, Wuhan 430072, China
* Correspondence: xuchao@nau.edu.cn

Abstract: With the rapid development of Internet of Things (IoT) technology, the number of devices connected to the network is exploding. How to improve the performance of edge devices has become an important challenge. Research on quality evaluation algorithms for brain tumor images remains scarce within symmetry edge intelligence systems. Additionally, the data volume in brain tumor datasets is frequently inadequate to support the training of neural network models. Most existing non-reference image quality assessment methods are based on natural statistical laws or construct a single-network model without considering visual perception characteristics, resulting in significant differences between the final evaluation results and subjective perception. To address these issues, we propose the AM-VGG-IQA (Attention Module Visual Geometry Group Image Quality Assessment) algorithm and extend the brain tumor MRI dataset. Visual saliency features with attention mechanism modules are integrated into AM-VGG-IQA. The integration of visual saliency features brings the evaluation outcomes of the model more in line with human perception. Meanwhile, the attention mechanism module cuts down on network parameters and expedites the training speed. For the brain tumor MRI dataset, our model achieves 85% accuracy, enabling it to effectively accomplish the task of evaluating brain tumor images in edge intelligence systems. Additionally, we carry out cross-dataset experiments. It is worth noting that, under varying training and testing ratios, the performance of AM-VGG-IQA remains relatively stable, which effectively demonstrates its remarkable robustness for edge applications.

Keywords: image quality assessment; attention mechanism; visual perception characteristics; medical images; symmetry edge systems

Academic Editor: Jie Yang

Received: 7 January 2025
Revised: 11 February 2025
Accepted: 19 February 2025
Published: 22 February 2025

Citation: Shen, F.; Liu, H.; Xu, C.; Ouyang, L.; Zhang, J.; Chen, Y.; He, Y. VGGNet and Attention Mechanism-Based Image Quality Assessment Algorithm in Symmetry Edge Intelligence Systems. *Symmetry* **2025**, *17*, 331. https://doi.org/10.3390/sym17030331

1. Introduction

Symmetry edge intelligence systems have emerged as a pivotal research area, bridging the gap between artificial intelligence algorithms and resource-constrained edge devices. Multiple symmetry edge nodes will collaboratively process data and enhance the system's performance. The era of edge intelligence has witnessed a massive influx of visual data from cameras embedded in smartphones, surveillance systems, autonomous vehicles, and industrial sensors. Image quality evaluation algorithms are essential for discerning the visual fidelity of these images, which directly impacts subsequent decision-making processes at the edge. Traditional image quality assessment (IQA) methods, initially developed for offline scenarios, are now being adapted and optimized to meet the resource-constrained and real-time requirements of symmetry edge devices.

Brain tumors are abnormal growths of brain cells in the midbrain, including primary brain tumors and brain metastases, and are considered a life-threatening disease. The best treatment method is complete surgical resection, but due to the non-resectable nature of normal brain tissue and the widespread infiltration of malignant tumors into the intracranial area, it is extremely difficult to undertake extensive and complete surgical resection. Therefore, how to reduce brain tissue damage, protect the central function of the brain, and maximize the elimination of tumors is still the current research direction and goal of oncologists. According to the relevant data, brain tumors account for over 85% of all primary central nervous system tumors worldwide, accounting for approximately 2% to 3% of cancer-related deaths, posing a huge threat to human health [1]. Therefore, early diagnosis and treatment of the brain tumors are particularly important. Magnetic resonance imaging (MRI) is a non-invasive imaging technique that can clearly display soft tissue lesions and is widely used in the diagnosis and treatment of brain tumor diseases. MRI image quality evaluation is an important step in clinical diagnosis and treatment planning, which can provide a basis for quantitative image analysis, auxiliary diagnosis, and surgical planning [2]. At present, most of the quality evaluation of brain tumor images still relies on manual analysis by doctors, which is time-consuming, laborious, and influenced by personal experience. Therefore, developing an accurate and reliable automatic brain tumor image quality evaluation algorithm is of great clinical significance.

In order to alleviate the storage and transmission challenges of massive data in symmetry edge intelligence systems, in addition to improving the network transmission speed and storage device capacity, we can also start from the image itself, filter out low-quality images through image quality evaluation, reduce the number of images from the source, and avoid unnecessary resource waste. In addition to reducing network and storage pressure, image quality assessment is also widely used in the decision-making field. An excellent quality evaluation model can not only reduce the workload of decision-makers, but also reduce the likelihood of decision-making errors. It can be seen that a reliable and versatile image quality assessment method has broad application prospects [3]. For example, in the military field, by evaluating the quality of satellite and infrared imaging, commanders can more confidently decide on operational plans and material allocation strategies. In the field of video live-streaming, guides can analyze the live-streaming effect in a timely manner through image quality assessment, and then adjust the live-streaming strategy and resource allocation. Especially in the medical field, a reliable image quality assessment method can not only provide specific quality scores for reference, greatly reducing doctors' workload, but also batch-label the distortion categories and levels of images, reducing the possibility of doctors making mistakes in diagnosing diseases [4]. By determining the distortion category and level in advance, experienced doctors can eliminate the interference of these distortions and distinguish the artifacts caused by distortion from the actual lesion area, achieving the effect of improving the accuracy of disease diagnosis [5].

It can be seen that image quality assessment has been increasingly widely applied in recent years, and this technology is becoming increasingly important in the medical field. However, there are some issues that need to be addressed. During the imaging process, images are inevitably affected by artifacts and noise, resulting in multimodality, non-uniformity, and blurriness [6–8]. Therefore, it is necessary to evaluate the quality of medical images to determine whether these images are correct and how trustworthy they are. By evaluating the quality of medical images, doctors can observe images from multiple directions and angles, thereby conducting more targeted quantitative analysis of lesion areas and improving the accuracy of disease diagnosis. In the process of image enhancement such as artifact removal and denoising in medical images, the main consideration is to improve the display effect of the image as much as possible without affecting

its credibility [9]. In addition, there are medical image reconstruction and other image processing processes that cannot be separated from the assessment of medical image quality, which is an important way to test the accuracy and credibility of images.

In the usage scenarios of medical images, the subjective judgment of doctors is often the most direct and accurate, but the drawbacks of long time consumption, high cost, and poor real-time applicability make subjective assessment difficult to achieve [10]. Therefore, objective medical image quality assessment has received attention. Many researchers have studied imaging instruments and imaging processes, but there is little research on how to objectively evaluate the quality of medical images after image processing [11]. Thus far, there is still no unified and recognized method. Traditional image quality evaluation involves obtaining a quality score through humans or machines, and then comparing the quality score with the original score label. However, medical image quality evaluation often involves obtaining the distortion type and level of the image through humans or machines, and then comparing them with the real label. The International Commission on Radiation Units and Measurements (ICRU) discussed the importance and basic methods of medical image quality assessment in Report No. 54 of 1996 [12]. At present, the peak signal-to-noise ratio (PSNR) is mainly used in medical image processing systems for the objective assessment of medical image quality [13]. This method is efficient in terms of speed; however, it overlooks the correlation between pixels, which can lead to significant deviations between the assessment results and human visual perception. Consequently, identifying more appropriate quality assessment criteria for medical images holds substantial importance for evaluating and optimizing medical image processing workflows.

Considering the above research needs, our research content is as follows: Using MRI images of brain tumors as the research object, this article first performs distortion processing on the original image, including various distortion categories and levels. Then, the sharpened quality score image is extracted from the distorted image, and the distorted image is combined with the quality score image. The distortion category and level are labeled as inputs for model training, Finally, the Visual Geometry Group Network (VGGNet) model in the convolutional neural network is adjusted and the attention mechanism module CBAM (Convolutional Block Attention Module) is added. After training, the final quality assessment model is obtained. Compared with traditional methods and similar methods, the distortion categories and levels obtained by our model have high consistency with real data and can make more accurate quality judgments on brain tumor images. In summary, our contributions of this paper are summarized as follows.

- We propose a non-reference image quality assessment model based on VGGNet that integrates the characteristics of the human visual system: AM-VGG-IQA. We add the attention mechanism module CBAM to the VGGNet model and extract the visual saliency images of distorted images. The distorted images are fused with the corresponding visual saliency images as the training dataset. From the final training results, the accuracy of the model in predicting distortion types and levels has been improved.
- We simulate potential issues that may arise in real-world application scenarios and further augment the brain tumor MRI dataset. In practical medical image applications, various factors, including lighting conditions, machine hardware performance, and transmission fluctuations, can introduce image distortion. To address this, we perform a series of processing operations on the original brain tumor dataset. By doing so, we generate distorted images with diverse types and varying degrees of distortion, effectively resolving the issue of insufficient dataset size for deep learning training.
- The experimental results convincingly illustrate that the proposed AM-VGG-IQA significantly enhances the classification and prediction accuracy with respect to distortion types and levels across TID2008 and TID2013, as well as the expanded brain tumor

MRI datasets. Simultaneously, it demonstrates a remarkable ability to evaluate the quality of brain tumor images within symmetry edge intelligence systems.

The remainder of this paper is organized as follows. Section 2 introduces the relevant work. Section 3 provides a detailed introduction of the proposed AM-VGG-IQA algorithm. Section 4 describes the experimental setup and evaluation results. Finally, the conclusion is presented in Section 5.

2. Related Work

Image quality assessment can be divided into subjective image quality assessment and objective image quality assessment. Among them, subjective image quality assessment can be divided into the mean opinion score (MOS) and differential mean opinion score (DMOS) types [14]. Since most users of image quality assessment are humans, the results obtained by subjectively scoring are often the most in line with human perception. It is precisely for this reason that manually scored data are often used to verify the accuracy of objective quality assessment models [15]. However, such methods require a significant amount of manpower and resources and require evaluators to learn a standardized scoring standard before participating in the scoring work [16]. As evaluators are human, there is a possibility of making mistakes during the assessment process. In summary, it is difficult to adapt to the fast-paced requirements of today's society.

Therefore, objective image quality assessment has gradually developed in recent years. According to the different reference content and scope, it can be divided into full reference, reduced reference, and no reference [17]. This type of method simulates the scoring process of humans through algorithms, completely entrusting the task of quality assessment to machines. With the rapid development of computing power in recent years, image quality assessment algorithms based on deep learning have seen rapid development [18]. The trained models can quickly and accurately provide quality scores, which meets the development needs of various industries.

2.1. Subjective Image Quality Assessment

According to the recommendations and standards issued by the Video Quality Expert Group established by the International Telecommunication Union, subjective image quality assessment can be roughly divided into three categories: single single-stimulus grading, dual dual-stimulus grading, and paired comparative judgment. Among them, single single-stimulus rating refers to presenting the test image on the screen for a certain period of time, and the evaluator gives a rating of "excellent", "good", "medium", "poor", or "inferior" to the image quality [19]. If the evaluator feels that this assessment result is too vague, they can provide specific scores for the corresponding score range. The difference between dual dual-stimulus rating and single single-stimulus rating is that the former provides both reference and test images, and the evaluator gives corresponding quality scores based on the assessment criteria of the latter [20]. Paired comparison judgment involves presenting two images of the same scene to the evaluator, who then assesses and categorizes them as either low quality or high quality, while also quantifying the perceived difference between the two.

The mean opinion score (MOS) and differential mean opinion score (DMOS) are the two most commonly used measurement methods for subjective image quality assessment [21]. The larger the MOS value or the smaller the DMOS value, the higher the perceived image quality. Although the evaluation results of such methods align closely with human subjective perception, their limitations are significant and cannot be overlooked. Firstly, these methods rely on manual assessment, which is time-consuming, labor-intensive, and unsuitable for real-time applications. Secondly, the evaluation process is susceptible to numerous interfering factors, such as the personal preferences and mental state of the

evaluators, the accuracy and brightness of the display equipment, and variations in the testing environment. Lastly, cultural and regional differences can influence the assessment outcomes, making the results highly context-specific. Given these advantages and disadvantages, such methods are currently primarily employed for constructing various image databases.

2.2. Full Reference Image Quality Assessment

As the name suggests, this type of method can refer to all the information of the original image during use. The core idea of the method is to calculate the similarity or difference between the reference image and the test image. It can be mainly divided into methods based on the HVS model, methods based on image structure, and methods based on various statistical information and information theory [22]. Among these methods, the most classic are the mean square error (MSE) and peak signal-to-noise ratio (PSNR). Their primary advantage lies in their fast computational speed. However, they suffer from a significant drawback: they ignore the correlation between pixels. As a result, in certain scenarios, the assessment results often deviate substantially from the human perceptual experience.

2.3. Reduced Reference Image Quality Assessment

In some scenarios where the full information of reference images cannot be obtained but high real-time requirements are required, such as video live-streaming and video calls, semi-reference image quality assessment is often adopted in the industry. This type of method can only obtain the image features of a limited reference image, and the algorithm uses these features to calculate the similarity between the distorted image and the reference image [23]. Due to the high dependence of these methods on image features, there is a great deal of uncertainty. If the obtained image features cannot be helpful in the application scenario, or even have misleading effects, the consequences can be unimaginable. Therefore, this type of method is relatively niche, with limited application scenarios and high requirements for the experience of evaluators and the adaptability of assessment algorithms [24].

2.4. No Reference Image Quality Assessment

The assessment of image quality without reference does not require the use of any information from the reference image, so this type of method is currently the most promising for development and application, because in some scenarios, obtaining the original image is difficult or even impossible. However, precisely because only distorted images are processed, the design difficulty of such methods is also the greatest. According to the different principles of establishing models, non-reference methods can be divided into two categories: knowledge-driven and data-driven.

Knowledge-driven methods are mostly traditional, with the main idea of relying on prior knowledge from human visual systems or natural scene statistics to manually extract the required image features, form feature vectors, and then train on the dataset to obtain the mapping relationship between feature vectors and quality scores. This type of method relies more on the background knowledge and accumulated experience of algorithm designers. Typical methods include that of Moorthy et al. [25], extracting NSS (natural scene statistics) features in wavelet transform and proposing a two-stage image quality prediction method based on Support Vector Machine (SVM) and Support Vector Regression (SVR). Firstly, the distortion type of the image is predicted through SVM, and secondly, the corresponding quality score is obtained through SVR mapping. Saad et al. [26] proposed an image integrity index based on Discrete Cosine Transform (DCT) statistical information. Firstly, the image was subjected to DCT transformation, and the kurtosis

and anisotropic entropy of the coefficient histogram were calculated as features. Finally, a multivariate Gaussian distribution was used to establish a mapping relationship between features and quality scores. Wang et al. [27] proposed a natural statistical feature method based on the relative order of brightness, extracting features such as kurtosis, variance, entropy, and differential entropy to train an image quality regression model. Wu et al. [28] proposed using directional similarity patterns to describe the correlation between local pixels, and then trained the regression model through SVR.

Data-driven methods are currently a hot topic in the field of image quality assessment. With the rapid development of computing power, the practicality and superiority of these methods have been demonstrated. Among them, the most popular method is end-to-end learning and training of image quality based on deep learning or convolutional neural networks. Kang et al. [29] proposed a convolutional neural network (CNN)-based method that achieved results comparable to mainstream full reference methods at the time. Bosse et al. [30] were inspired by the structure of the visual cortex in the brain and designed a network model consisting of 10 convolutional layers and 2 fully connected layers. The model's feature expression ability was improved, and the prediction accuracy of quality scores was also improved. Ma et al. [31] proposed an end-to-end image quality assessment model based on multitasking. His contribution was to provide a new solution to the problem of insufficient training data. He first performed distortion processing on the reference image to obtain distorted images with five degrees of distortion. Coupled with the reference image itself, there were five image types, transforming the problem into a five classification problem. He also used GDN (generalized divisive normalization) as the activation function in the model, effectively reducing the number of network parameters. Su et al. [32] proposed an adaptive hypernetwork structure that divides image quality assessment into three stages: understanding image content, perceptual rule learning, and quality prediction.

2.5. The Problems with the Above Methods

Traditional algorithms have been proven effective in practical usage scenarios throughout history, but with the richness of application scenarios and the sharp increase in sample size, the accuracy of algorithm evaluation begins to gradually decline, making it difficult to handle these complex scenarios. Meanwhile, traditional algorithms often only extract low-level features of images and cannot classify and generate high-level semantic information; thus, they cannot adapt to new tasks.

Deep learning-based algorithms can perfectly cope with massive data scenarios, but in the early stages, a large number of samples are required for training, and the samples must meet the application scenarios of the algorithm and contain corresponding type labels. Meanwhile, convolutional neural networks can simulate the workflow of the brain, but when analyzing image quality, they lack consideration for human eye perception. Therefore, the evaluation results of previous deep learning-based algorithms often have a certain gap with human eye perception.

In summary, there has been significant progress in the assessment of image quality without reference. However, these methods still face some problems and there is a lot of room for improvement:

(1) The number of samples required for training is too small, and the types covered by the dataset are also too few.

(2) Methods are needed to simulate the observation process of the human eye in the model to make the assessment results more in line with human eye perception.

(3) Methods are needed to improve the universality of the model to reduce its usage threshold and improve its practicality.

3. The Proposed AM-VGG-IQA

At present, there is limited research on image quality assessment in the medical field. Our algorithm takes brain tumors as the research object and proposes a network model based on VGGNet and the attention mechanism module CBAM, later referred to as AM-VGG-IQA.

As shown in Figure 1, our method can be divided into three main stages: image distortion processing, sharpness image fractional image fusion, and model training. Firstly, there is image distortion processing, with the main purpose of expanding the dataset and simulating the problems faced when using images in real-world scenarios. Secondly, there is fractional image fusion for sharpening images, with the main purpose of making tumor lesion areas and edges more prominent and improving recognition accuracy. Finally, there is training, where corresponding changes are made to the input image and output results, and attention mechanisms are added to the model. While accelerating training speed, it also improves the accuracy of prediction. Figure 2 shows the specific process of the AM-VGG-IQA algorithm. Next, we will elaborate on this process in detail.

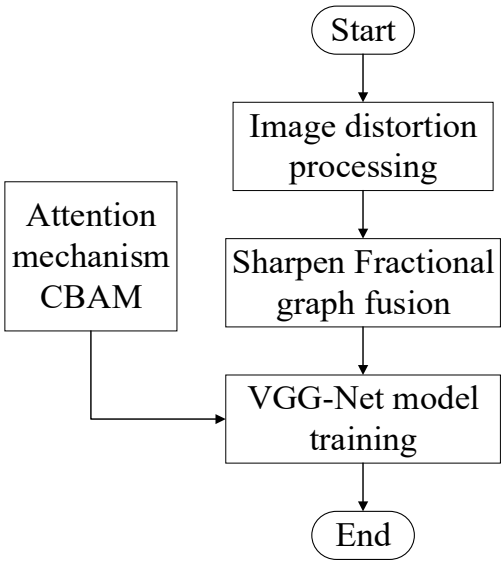

Figure 1. Algorithm flowchart of AM-VGG-IQA.

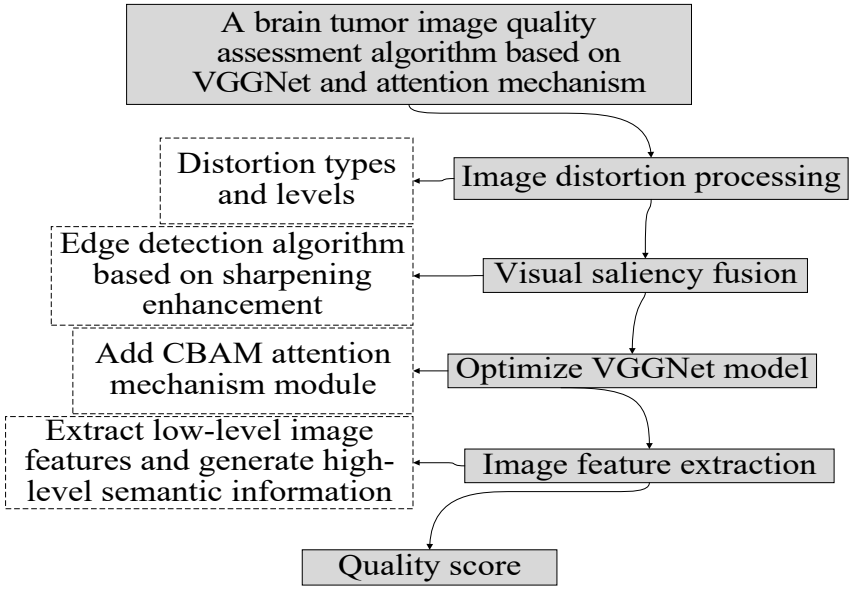

Figure 2. Specific process of AM-VGG-IQA.

3.1. Image Distortion Processing

Training the deep learning model requires a large number of training samples. The magnetic resonance imaging dataset of brain tumors we use contains a total of 5712 original images of four tumor types. Considering the particularity of medical images and the complexity of the model's tasks, we also need to expand the dataset to ensure the robustness of the model. In the actual application scenarios of medical images, we often encounter problems such as insufficient light or overexposure, machine hardware problems, network transmission fluctuations, etc. After the full investigation and study of true distorted medical images, we summarize the common distortion types into the following seven types: Gaussian blur, Gaussian noise, haze effect, JPEG compression, salt and pepper noise, overexposure, and underexposure, which are close to the true distortion effects and cover the types of distortion in most cases.

Due to the inability to collect a large number of realistic distorted images that meet the training requirements of the model, we have decided to use algorithms to synthesize distorted images. To cover all possible scenarios as much as possible, we will also adjust the specific parameters in the algorithm to achieve different levels of distortion effects. Among them, the first five distortion types each have five distortion levels, and the last two distortion types each have two distortion levels. We use "distortion type, distortion level" as the labels for the image, and a total of 29 labels are obtained. After completing the algorithm processing, we obtain 165648 distorted images, as shown in Figure 3. It is definitely unrealistic to use subjective image quality assessment methods to score images, and using other objective image quality assessment methods may result in controversial results. But we have a clear understanding of the type and level of image distortion, so we use a unique heat vector (corresponding to the distortion type and level of 1; all others are 0) as the image label, so that we can refer to the distortion type and level of the image in a unified form. At this point, we have solved the two major problems of insufficient database quantity and missing image training labels.

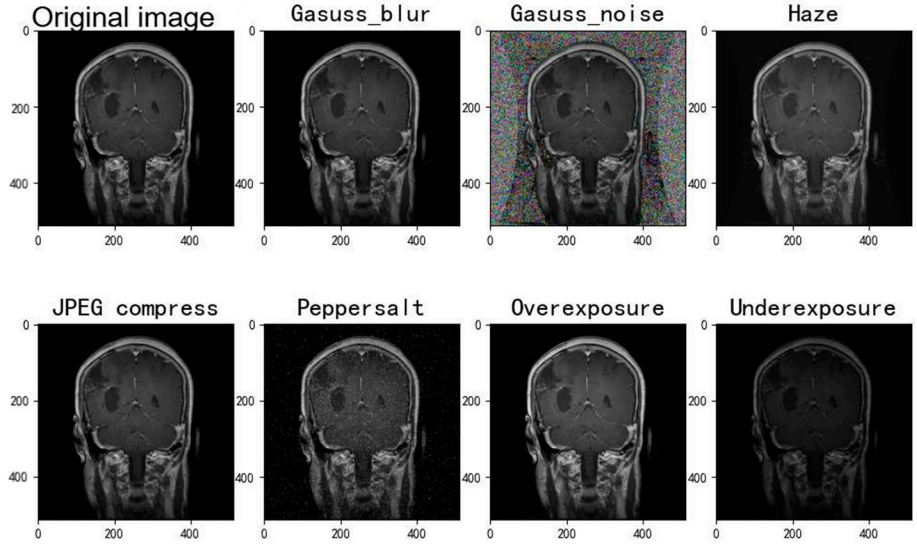

Figure 3. Image quality reduction processing.

3.2. VGGNet Network Model

VGGNet is a convolutional neural network developed by the Computer Vision Group of Oxford University and Google DeepMind Laboratory. It is characterized by simple structure and excellent performance. The simple and clear structure does not mean that there are few layers in the network. On the contrary, the original VGGNet has 11-19 layers, which belongs to the ultra deep convolutional network. This means that all convolution kernels

are 3*3 in size, and the model building process is relatively regular. By repeatedly stacking the 3*3 convolution core and the 2*2 pooled core, VGGNet's network depth continues to deepen, but this does not bring an explosive increase in the number of parameters, because most parameters are concentrated in the final full connection layer. The increase in the depth of the convolution layer, on the contrary, gives the model a stronger feature learning ability. The uniform use of 3*3 convolution kernels also greatly reduces the number of parameters in the network. The concatenation of two 3*3 convolution kernels is equivalent to a 5*5 convolution kernel, and the concatenation of three 3*3 convolution kernels is equivalent to a 7*7 convolution kernel. However, the number of parameters of three 3*3 convolution kernels is only half that of a 7*7 convolution kernel. Therefore, this strategy reduces the number of parameters while expanding the receptive field of the convolution layer, giving the model a better learning ability.

When training the model, it is required to input 224*224 images, so cutting is required. In order to strengthen the model, we also flipped the image randomly. When necessary, the data should also be regularized to reduce the complexity of the model in case of overfitting. By studying the operating mechanism of the human brain, it is found that the activation sequence of neurons in the human brain is relatively scattered and sparse when processing information coding, and the linear rectification unit ReLU can exactly imitate the working principle of this mechanism, so we uniformly use ReLU as the activation function of the convolution layer in the model. At the same time, we also set the step size to 1, so as to reduce the number of network parameters and make the process more conducive to extracting local image features. In the training process, the argmax function is selected to calculate the most likely distortion type and level. The general network structure is shown in Table 1.

Table 1. Improved convolutional network model structure based on VGGNet.

Layer	**Conv1_1**	**Conv1_2**	**Maxpool1**	**Conv2_1**
Kernel	64	64	/	128
Output	224,224,64	224,224,64	112,112,64	112,112,128
Layer	Conv2_2	Maxpool2	Conv3_1	Conv3_2
Kernel	128	/	256	256
Output	112,112,128	56,56,128	56,56,256	56,56,256
Layer	Conv3_3	Maxpool3	Conv4_1	Conv4_2
Kernel	256	/	512	512
Output	56,56,256	28,28,256	28,28,512	28,28,512
Layer	Conv4_3	Maxpool4	Conv5_1	Conv5_2
Kernel	512	/	512	512
Output	28,28,512	14,14,512	14,14,512	14,14,512
Layer	Conc5_3	Maxpool5	FC1	FC2
Kernel	512	/	/	/
Output	14,14,512	7,7,512	512*7*7,4096	4096,4096
Layer	FC3	Argmax	/	/
Kernel	/	/	/	/
Output	29	1	/	/

3.3. Attention Mechanism Module CBAM

CBAM (Convolutional Block Attention Module) was proposed in 2018, and its main working principle is similar to the human visual system. After scanning the global image, it focuses its attention on the area of interest, invests more attention in this area, reduces attention to other areas, and even directly ignores irrelevant information. This mechanism can help humans to quickly screen out useful information from a large amount of information. Similarly, this mechanism can also help network models to prioritize resource allocation to important tasks when computing power is limited, solving the problem of information overload. A large amount of practice has proven that within a certain range, the more parameters a model has, the stronger its expressive power. However, it also means that the computational load is greater. Therefore, the addition of the attention mechanism module not only preserves the model's strong learning ability, but also greatly reduces the computational load required by the model, improving the efficiency of model training. In addition, attention mechanism modules are mostly integrated as separate modules that can be added to the model conveniently.

The simplified model diagram of CBAM is shown in Figure 4, which consists of two sub modules: CAM (channel attention module) and SAM (spatial attention module), which perform channel and spatial processing, respectively.

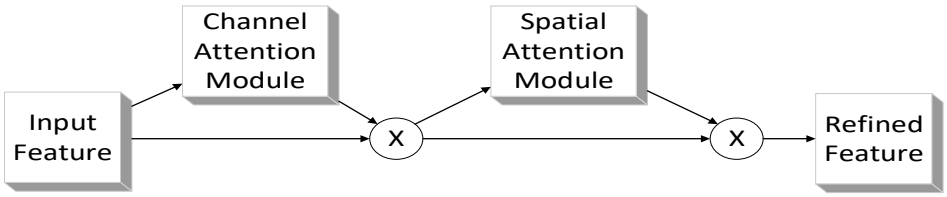

Figure 4. CBAM model diagram.

The working principle of the channel attention module is shown in Figure 5. The input feature map passes through two parallel MaxPool layers and AvgPool layers, changes the feature map from C*H*W to C*1*1, and then passes through the Shared MLP module. In this module, it first compresses the number of channels to 1/r (reduction) times the original number of channels, then expands to the original number of channels, and obtains two activated results through the ReLU activation function. Adding the two output results element by element, it then obtains the output result of the channel.

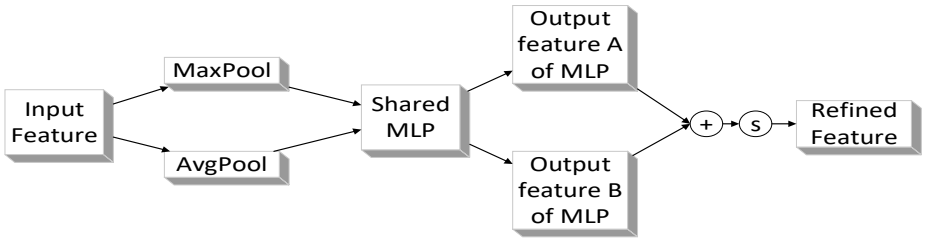

Figure 5. Channel attention module.

The attention is obtained through a sigmoid activation function. Then, the output result is multiplied by the original image to return to the size of C*H*W.

The working principle of the spatial attention module is shown in Figure 6. The input image feature map processed by the channel attention module passes through the MaxPool layer and AvgPool layer, and then uses a standard convolution layer for connection and convolution. Then, the sigmoid activation function is used to obtain the weight value of each feature point of the feature map, and finally the weight is added to the feature layer by multiplication.

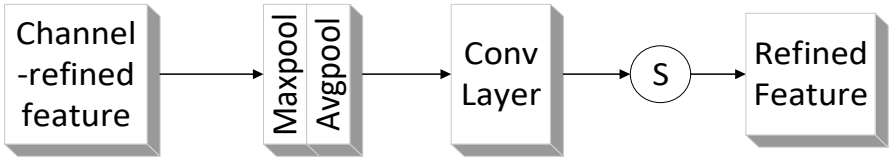

Figure 6. Spatial attention module.

3.4. Visual Saliency Fusion

In previous image quality assessment algorithms or models, the mainstream idea was to extract the required image features and process and evaluate these features to obtain the final quality score. These methods do not integrate the characteristics or working principles of the human visual system; thus, in some scenarios, the results obtained may have significant deviations from the results perceived by the human eye. In response to this deficiency, we incorporated visual saliency maps into the image, so that the model possesses the characteristics of the human visual system.

Among numerous visual saliency detection methods, we have comprehensively considered algorithm complexity and implementation difficulty. The final visual saliency calculation method we use is an edge detection algorithm based on sharpening enhancement. This algorithm first enhances the sharpening effect of the image to highlight the edge lines of the image. This feature is beneficial for our experimental subjects to quickly locate the lesion area of tumors and accelerate the model's learning of key information such as tumor shape, texture, color, etc. The algorithm uses the Laplace transform, which enhances the high-frequency information of the image, making the edge contours and texture details of the image body more prominent. At the same time, the low-frequency information of the image is suppressed, removing low-frequency information from the image. The reason why low-frequency information can be removed without affecting the accuracy of the algorithm is that the image subject in brain tumor MRI images is presented as high-frequency information, while low-frequency information is mainly distributed in the background part of the image, with little correlation with the image subject. Therefore, it can be safely removed to improve the efficiency of the algorithm. After obtaining the visual saliency image corresponding to the distorted image, we need to fuse the two. The method used here is to multiply the two, as shown in Formula (1), to obtain a quality feature map that integrates visual saliency information.

$$FM = DM \times VSM \tag{1}$$

As shown in Formula (1), FM represents the quality feature map that integrates distorted images and visual saliency images, DM represents distorted images, VSM represents corresponding visual saliency images, and FM serves as input for subsequent model training and testing. For $DM \times VSM$, the generated results can be understood as saliency-weighted distortion information. This means that in areas with high saliency, the distortion information will be amplified, while in areas with low saliency, the distortion information will be suppressed.

4. Experiments

To fully verify the performance of the algorithm, this experiment introduces the brain tumor MRI dataset, the TID2008 dataset, and the TID2013 dataset, which can not only verify the practicability of the algorithm in the field of brain tumor images, but also test the universality of the algorithm. First, the necessity of several elements or steps in the algorithm is verified through the control variates. Secondly, the Structural Similarity Index Measure (SSIM), PSNR, Natural Image Quality Evaluator (NIQE), Blind/Referenceless

Image Spatial Quality Evaluator (BRISQUE), Blind Image Quality Assessment Based on Convolutional Neural Networks (BIECON), and Cascade Hierarchical Deep Convolutional Network (CaHDC) algorithms are introduced for comparative tests. The performance of the algorithms in all aspects is reflected through the correlation indicators SROCC, KROCC, and PLCC. Finally, the generalization performance of the algorithms is tested through cross-dataset experiments.

4.1. Experimental Environment and Parameter Settings

(1) Experimental environment

The environmental configuration of this experiment is shown in Table 2.

Table 2. Experimental environment.

Environment	Statement
Hardware platform	Intel Core i7 CPU and GeForceRTX3090
Software platform	Electerm and Windows 10
IDE	PyCharm2021.2.1
Dataset	TID2008 and Tid2013 and Brain Tumor MRI
Source code of AM-VGG-IQA	https://github.com/rookiets/AM-VGG-IQA.git (accessed on 26 June 2023)

(2) Experimental parameter settings

In the process of image distortion processing, there are many parameters involved, with the aim of obtaining images with different degrees of distortion under different types of distortion. Taking the model in the comparative experiment as an example, we have listed some important model parameters during the training process, as shown in Table 3.

Table 3. Experimental parameter settings.

Parameter	Settings
Train and test ratio	4:1
Loss function	MSE
Optimizer	Adam
Regularization	L2 regularization method
Learning rate	0.0002

The reason for choosing MSE as the loss function is that its function curve is smooth, continuous, and differentiable everywhere, making it easy to use a gradient descent algorithm, which is a commonly used loss function. Moreover, as the error decreases, the gradient also decreases, which is beneficial for convergence. Even with a fixed learning rate, it can converge quickly to the minimum value. The reason for choosing the Adam optimizer is that it can adaptively adjust the learning rate of each parameter, thereby improving the convergence speed and generalization ability of the model. In addition, the computational complexity of the Adam optimizer is relatively small, which can improve the training speed of the model. The reason for choosing the L2 regularization parameter is that it can prevent overfitting—L2 regularization can reduce the complexity of the model, thereby improving generalization ability and reducing the occurrence of overfitting problems; it can aid in feature selection—L2 regularization can penalize weights to make the weights of some unimportant features approach zero, which can automatically perform feature selection, thereby simplifying the model; and it can solve the problem of collinearity—when there is collinearity (i.e., linear correlation) between features, L2 regularization can reduce the impact of this collinearity by reducing the weight of related features. As for

setting the learning rate to 0.0002, it takes into account both training speed and model overfitting issues.

4.2. Dataset Preparation and Processing

In Section 3, we introduced the brain tumor MRI dataset, introduced the distortion processing of the images in the dataset, and explained how to generate the tags for training. Due to the fact that the initial input data and final output results of most classic image quality assessment algorithms are specific quality scores, this dataset is not suitable for horizontal comparison between AM-VGG-IQA and other algorithms. It is only used for early training of the VGG model.

To verify the comprehensive performance of the algorithm, we introduced other image quality assessment algorithms using the classic datasets TID2008 and TID2013, which contain 17 distortion types, 1700 distorted images, and 256,428 MOS values. The latter is an enhanced version of the former, which includes 24 distortion types, 3000 distorted images, and 524,340 MOS values. These two databases contain a variety of distortion types and cover a wide range of image types, making them suitable for measuring the overall performance of algorithms. The databases and key information used in the experiment are shown in Table 4. The data preprocessing methods and model training environments are consistent with the existing models.

Table 4. Dataset basic information.

Dataset	Num	Type_Num	Label
Brain tumor	165,648	7	Distortion type and level
TID2008	1700	17	MOS
TID2013	3000	24	MOS

4.3. Evaluation Criterion

The performance assessment of image quality assessment algorithms is to examine the correlation between the objective scores calculated by the algorithm on images with different types and degrees of distortion. In the magnetic resonance image dataset of brain tumors, the image label is the distortion type and distortion level corresponding to the image. Therefore, when training the VGG network in the early stage, the inspection content is the accuracy of judging the distortion type and distortion level. Due to the fact that the calculation methods measure the correlation between real data and predicted data, we can use the same methods and indicators to test and compare on three databases.

In the experiment, the classic indicators used in statistics to describe data correlation were used:

(1) Spearman's rank correlation coefficient (SROCC);

(2) Kendall rank order correlation coefficient (KROCC);

(3) Pearson linear correlation coefficient (PLCC).

The Spearman coefficient and Kendall coefficient are mainly used to evaluate the hierarchical correlation between two groups of data, and can also be used to reflect the monotonicity of the algorithm; The Pearson linear correlation coefficient is used to measure the accuracy of model predictions.

The proposed model is utilized to compute the distortion type, distortion level, or quality score for the distorted images in the dataset. Subsequently, the correlation coefficients between the distortion labels (or scores) and the real labels (or MOS values) are calculated. Higher absolute values of SROCC, KROCC, and PLCC indicate superior algorithm performance.

4.4. Prediction Accuracy of AM-VGG-IQA

In our model, in addition to distorting the images in the dataset, we also calculated the visual saliency images corresponding to the distorted images and fused them. At the same time, we also added an attention mechanism module to the VGG model. To sum up, it can be summarized into two aspects: attention mechanism module and visual saliency fusion. Next, we will verify the necessity of these two steps through a series of experiments.

(1) Attention mechanism module

An attention mechanism is a way of thinking that imitates the human visual or auditory system. In the traditional neural network model, all input data are regarded as equally important. The addition of attention mechanism allows neural networks to apply different weights to input information at different positions and parts when processing sequence data, thereby improving the performance, robustness, and generalization ability of the model.

When we are looking at an image or listening to an audio clip, we do not allocate our attention equally to all information, but instead selectively focus on key information based on our own interests. Similarly, the attention mechanism will allocate different weights based on different parts of the input, focusing computational power on important areas to improve the performance of the model. In the working process of attention mechanism, input data are divided into two parts: query vector and key-value pair. The query vector represents the objects that need attention, while the key-value pair maps to various parts of input data. By calculating the similarity between the key-value pair and the query vector, different weights can be assigned to each part to determine the key information that needs attention. The common methods of similarity calculation include dot product, additive, and multilayer perceptron. After obtaining the weights of each part, a more accurate weighting sum can be performed.

Due to the uncertainty of the compatibility between the modules and our model, in order to select the most suitable attention mechanism for this model, we selected three attention mechanism modules: CBAM (Convolutional Block Attention Module), BAM (Bottleneck Attention Module), and DANet (Dual Attention Net), and added them to the model for experimental comparison. Among them, CBAM applies a combination of channel attention module and spatial attention module to process the input feature layers in the channel attention module and spatial attention module, respectively. Its working principle has been roughly introduced in the previous model introduction chapter. BAM can be integrated with any feedforward convolutional neural network, which infers an attention map along both channel and spatial paths to emphasize important elements. Placing the module at the bottleneck of the model significantly reduces the computational and parameter overhead, and it can be jointly trained with any feedforward model in an end-to-end manner. Unlike the previous two modules, the DANet module consists of a positional attention mechanism and a channel attention mechanism. The positional attention mechanism is responsible for capturing the spatial dependencies of feature maps at any two positions, and similar features are related to each other regardless of distance. The channel attention mechanism is responsible for integrating relevant features between all channel mappings to selectively emphasize the existence of interdependent channel mappings.

In order to test the effectiveness of the module, we conducted a horizontal comparative experiment. The dataset used is a distorted brain tumor MRI dataset. The training purpose is to predict the distortion type and level of the distorted image. There are a total of four models participating in the comparison, and the differences between them are shown in Table 5.

Table 5. Differences in models.

Model A	VGGNet
Model B	VGGNet+CBAM
Model C	VGGNet+BAM
Model D	VGGNet+DANet

Since the task of this model is to predict the type and level of distortion in distorted images, we use accuracy as a measure. The training process of the four models is shown in Figure 7.

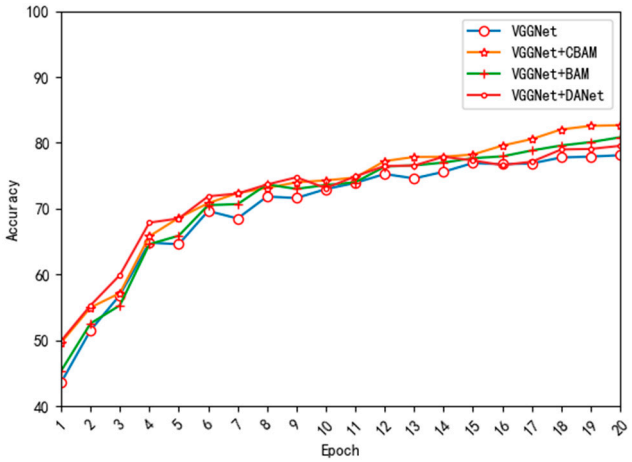

Figure 7. Comparison of prediction accuracy.

From Figure 7, we can see that the accuracy of models B, C, and D with the added attention mechanism module is higher than that of model A without the added attention mechanism module. Specifically, compared to Model A, Model B with the addition of the CBAM module increased its prediction accuracy from 78.10% to 82.67%, an increase of 4.57%. Model C with the addition of the BAM module increased by 3.5% compared to Model A, and Model D with the addition of the DANet module increased by 1.8% compared to Model A. This shows that the attention mechanism module can not only reduce the number of network parameters and accelerate training speed, but also improve the performance of the model. Among them, Model D with the DANet module performs best in the early training period, which is reflected in the fastest improvement in model accuracy. This is because the DANet module location attention mechanism will focus more quickly on the lesion area of brain tumor, and the structure and lines of this area are often the most complex and obvious. Therefore, through the calculation and comparison of this area, it is possible to determine the type and level of distortion in an image with faster speed and higher accuracy. In addition, the final performance of Model B is the best among the four models, so in subsequent experiments, all models that added attention mechanism modules were added with CBAM modules.

(2) Visual saliency fusion

In Section 3, we mentioned adding visual saliency fusion processing to the model. The specific operation is to obtain the visual saliency image of the image through an edge detection algorithm based on sharpening enhancement, and fuse the image with the corresponding distorted image to obtain new training data. Therefore, in order to verify the effectiveness of this step, we need to conduct comparative experiments with the original distorted image. In this experiment, the models used were the original model A without the CBAM module and the new model B with the CBAM module added. Based on whether the image had undergone visual saliency fusion, we divided the dataset used for training

into dataset A and dataset B. Dataset A contains the original distorted image, while dataset B contains the distorted image fused with visual saliency images. Finally, we generated four solutions, with specific details shown in Table 6.

Table 6. Differences in schemes.

Scheme	Model	Dataset
VGGNet+MRI	VGGNet	Brain tumor MRI dataset
VGGNet+MRI-VSF	VGGNet	Brain tumor MRI dataset with visual saliency fusion
VGGNet-CBAM+MRI	VGGNet+CBAM	Brain tumor MRI dataset
VGGNet-CBAM+MRI-VSF	VGGNet+CBAM	Brain tumor MRI dataset with visual saliency fusion

After the accuracy stabilized, we selected three rounds of experimental results to enhance the stability of the results. They were divided into three groups, and their comparison is shown in Figure 8. For each group, there are four schemes, the first and second schemes are based on the VGGNet model, and the third and fourth scheme are based on the VGGNet+CBAM model. For the comparison, the second and fourth schemes have visual saliency fusion. We can see that Model B (VGGNet+CBAM) with the addition of the CBAM module performs better than Model A (VGGNet), indicating the improvement effect of the attention mechanism module on the model. Taking Group 3 as an example, Scheme 2 (VGGNet model with VSF) shows a 3.5% improvement in accuracy compared to Scheme 1 (VGGNet model without VSF), while Scheme 4 (VGGNet+CBAM model with VSF) shows a 2.7% improvement in accuracy compared to Scheme 3 (VGGNet+CBAM model with VSF), we can see that the dataset processed by visual saliency fusion has improved its performance in model training. Compared to Scheme 1, Scheme 4 has improved its accuracy by at least 8.7%, indicating that the combination of the CBAM module added model and the visual saliency fusion processed dataset achieved the best results.

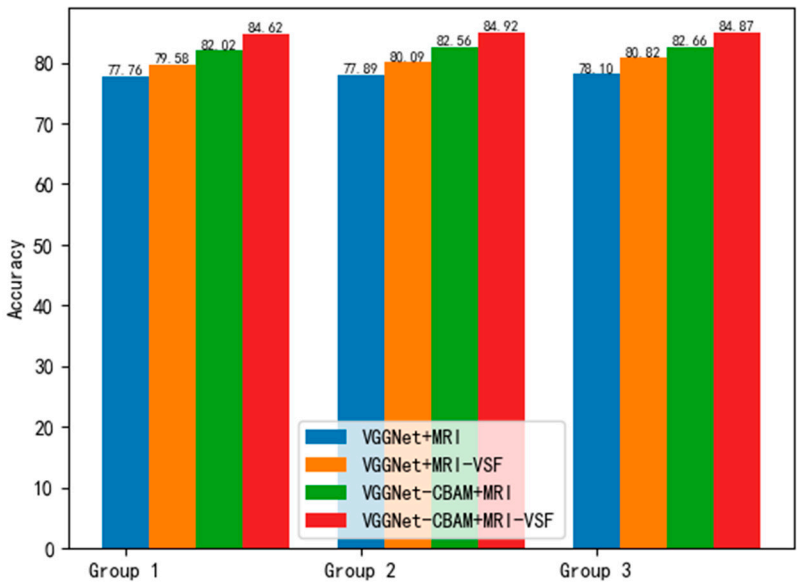

Figure 8. Comparison of accuracy between models.

Based on the above argument, we conclude that the additions of the attention mechanism module CBAM and the image processing of visual saliency fusion have improved our model. Therefore, in subsequent experiments, we will make this model scheme our

final choice. Next, we will conduct a series of comparative experiments to compare the performance of this model with other classic and excellent models.

4.5. Performance Comparison with Other Image Quality Assessment Methods

(1) Experimental Subjects and Objectives

To verify the performance of AM-VGG-IQA, we need to compare it horizontally with some classic algorithms. In this experiment, we introduced three types of image quality assessment algorithms: traditional full reference methods: SSIM and PSNR; non-reference methods based on natural image features: NIQE and BRISQUE; and deep learning-based non-reference methods: BIECON and CaHDC. The purpose of comparing with the first type of method is to examine how AM-VGG-IQA performs compared to the full reference method. The second and third types of methods are both non-reference methods. The former is based on natural image features, while the latter is based on deep learning. By comparing with these methods, we can understand whether deep learning has improved the reliability of image quality assessment. AM-VGG-IQA is of the same type as the third class of method, so we can visually demonstrate the performance of AM-VGG-IQA by comparing it with the third class of method.

(2) Parameter Settings for Datasets and VGGN Models

In algorithms that require training to obtain or adjust model parameters, we randomly divide the dataset into training and testing sets in a 4:1 ratio, while performing unified cropping and random horizontal flipping on the images. In the early stage, we used the brain tumor MRI dataset to train our model and learn the relevant features of a brain MRI. Considering that the model learns too few image types, this is not conducive to comparing it with other algorithms in the later stage. Therefore, before conducting a horizontal comparison, we separately trained our model using the TID2013 dataset to finetune the parameters in the network. At the same time, we changed the output of the model from the distortion type and distortion level to the specific quality score, and we selected MSE (mean square error) as the loss function in the global quality regression stage. Since the network model has many parameters, in order to prevent overfitting in the training process, we adopted the L2 regularization method; that is, the L2 norm of the weight vector was added to the loss function. This can reduce the impact of fluctuations in quality independent parameters. This model uses the Adam optimization algorithm to update the weights of the neural network, with an initial learning rate set to 0.0002.

(3) Performance Comparison in Image Quality Assessment Experiments

In order to verify the comprehensive performance of the algorithm, we selected two datasets, namely TID2008 and TID2013. By comparing the performance of the algorithm on the TID2008 and TID2013 datasets, we can know the universality of AM-VGG-IQA in other image types or other application scenarios. This is also a part of algorithm performance.

As shown in Tables 7 and 8, the best results under each assessment criterion are highlighted in bold font. In the training tests of the TID2008 and TID2013 datasets, the performance of SROCC and PLCC of AM-VGG-IQA was the best among all algorithms, which means that the hierarchical correlation between the quality score judged by the algorithm and the actual score is the best, and the accuracy of model prediction is also the best. This is thanks to our model using the VGGNET network, which has been pre trained on large-scale ImageNet datasets for image classification tasks and extracts relevant features of real-world distorted images. In addition, AM-VGG-IQA also used distorted brain tumor MRI images to modify the pre-trained VGGNET network. At the same time, we also performed visual saliency fusion on distorted images and added attention mechanism modules to the model, all of which provided assistance in improving the performance of the model. Finally, we also used TID2013 to finetune the network parameters and change

the output of the network. Therefore, the AM-VGG-IQA network model can achieve good results on the TID2008 and TID2013 datasets, which verifies the universality of AM-VGG-IQA.

Table 7. Performance comparison on the TID2008 dataset.

Dataset	**TID2008**		
Index	SROCC	KROCC	PLCC
SSIM	0.893	0.914	0.906
PSNR	0.905	0.919	0.897
NIQE	0.656	0.732	0.726
BRISQUE	0.910	0.904	0.917
CaHDC	0.895	**0.920**	0.905
AM-VGG-IQA	**0.913**	0.894	**0.926**

Table 8. Performance comparison on the TID2013 dataset.

Dataset	**TID2013**		
Index	SROCC	KROCC	PLCC
SSIM	0.856	0.807	0.867
PSNR	0.889	0.848	0.847
NIQE	0.593	0.695	0.677
BRISQUE	0.872	**0.885**	0.903
CaHDC	0.862	0.839	0.878
AM-VGG-IQA	**0.893**	0.872	**0.912**

Next, in order to verify the robustness of AM-VGG-IQA, we will compare the performance changes in the algorithm under three different schemes by changing the ratio of the training and testing sets. The dataset used is still TID2013, and the assessment criteria are the SROCC values. The training and testing ratio in the previous experiment was 4:1. In the following experiment, we added a control group with training and testing ratios of 1:1 and 1:4. The test results are shown in Table 7.

From Table 9, it can be seen that AM-VGG-IQA achieved good results in training and testing ratios of 4:1, 1:1, or 1:4, with no significant fluctuations in SROCC values, which proves that AM-VGG-IQA has a better robustness.

Table 9. Comparison of SROCC on the TID2013 dataset.

Proportion of training and testing	4:1	1:1	1:4
AM-VGG-IQA's SROCC	0.893	0.859	0.822

4.6. Cross-Dataset Experiments

To verify the generalization ability of the AM-VGG-IQA algorithm, we also conducted cross-dataset experiments, namely training on one dataset and testing on another dataset. The databases used include TID2008 and TID2013, and the common distortion types of images in both datasets are selected. The first group of experiments used TID2008 for training and TID2013 for testing. The second group of experiments used TID2013 for training and TID2008 for testing, with SROCC and PLCC as assessment indicators. The experimental results are shown in Table 10. Comparative algorithms include BRISQUE, BIECON, and CaHDC.

As shown in Table 10, the best-performing data are highlighted in bold font. Under the standards of SROCC, the BRISQUE algorithm performs the best, possibly due to its ability to extract mean subtracted contrast normalized (MSCN) coefficients from images, fit MSCN

coefficients into an asymmetric generalized Gaussian distribution (AGGD), and extract features of the fitted Gaussian distribution. At the same time, MSCN does not have a strong dependence on the strength of the texture. The features extracted in this way are more applicable. The AM-VGG-IQA algorithm achieved the best or near the best results in both indicators in two sets of cross-dataset experiments. In some indicators, there is even a 20% improvement compared to some algorithms. Especially under the standards of PLCC, AM-VGG-IQA performs the best, indicating that the correlation between the results obtained by the algorithm and subjective scores is the strongest. This is thanks to the addition of visual saliency fusion processing to the algorithm, which fully simulates the characteristics of human eyes when observing images, thus improving the performance of the algorithm. It can be seen that the AM-VGG-IQA algorithm has a strong generalization ability.

Table 10. Cross-dataset experimental results.

Train	**TID2008**		**TID2013**	
Test	**TID2013**		**TID2008**	
Index	SROCC	PLCC	SROCC	PLCC
BRISQUE	**0.82**	0.84	**0.88**	0.87
BIECON	0.67	0.66	0.71	0.75
CaHDC	0.74	0.71	0.75	0.78
AM-VGG-IQA	0.80	**0.85**	0.87	**0.90**

5. Conclusions

This paper proposes an image quality assessment model based on VGGNET that integrates visual saliency and attention mechanisms and integrates the feature extraction and quality score prediction of distorted images into an optimized framework. The image is subjected to distortion processing and visual saliency fusion, which improves the accuracy of image feature extraction while expanding the dataset. The addition of an attention mechanism module improved the efficiency of image feature extraction and calculation, further improving the performance of the model. The necessity of the above operation has been verified through prediction accuracy experiments. Through comparative experiments, we can conclude that our method has a better prediction accuracy and higher consistency with the subjective quality perception of the human eye compared to traditional methods, based on natural scene statistics and some deep learning methods. The results of cross-dataset experiments indicate that the proposed method also has a good generalization performance for symmetry edge intelligence systems.

The proposed model is optimized for MRI images of specific resolutions, and further evaluation across different resolutions is required. The proposed IQA scheme plays a pivotal role in a wide range of practical applications, spanning from image processing and medical imaging to computer vision and multimedia communication. Its primary utility lies in evaluating and optimizing the visual fidelity of images, ensuring that they meet the desired standards for specific tasks. For instance, in medical imaging, IQA is crucial for enhancing diagnostic accuracy by assessing the clarity and detail of medical scans. In multimedia applications, it helps to maintain high-quality video streaming and efficient image compression, balancing quality and bandwidth usage. Despite its widespread adoption, IQA is not without limitations. One major challenge is the subjective nature of image quality, as the human perception of quality can vary significantly and is influenced by contextual factors. Furthermore, most IQA methods are designed for specific types of distortions and may not generalize well to unseen or complex distortions. Looking ahead, future research should focus on developing more robust and generalizable no-reference IQA methods that can handle diverse and complex distortions. Incorporating advanced

deep learning techniques and leveraging large-scale datasets could further enhance the accuracy and adaptability of IQA models. Additionally, exploring the integration of IQA with emerging technologies such as virtual reality (VR) and augmented reality (AR) could open new avenues for improving immersive experiences. Addressing these challenges and opportunities will be essential for advancing the field of image quality assessment and expanding its practical applications.

Author Contributions: Conceptualization, F.S. and H.L.; methodology, F.S., H.L. and C.X.; software, F.S., C.X. and L.O.; validation, F.S. and H.L.; formal analysis, F.S.; investigation, F.S. and J.Z.; resources, F.S. and J.Z.; data curation, F.S.; writing—original draft preparation, F.S. and H.L.; writing—review and editing, F.S., Y.C. and H.L.; visualization, F.S., H.L. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Basic Science (Natural Science) Research Project of Colleges and Universities in Jiangsu Province (24KJA520005, 22KJA520004), and the National Natural Science Foundation of China (61902189, 62472227, 62162002 and 61972293).

Data Availability Statement: The original data presented in this study are included in this article, and further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors sincerely thank all the reviewers for their valuable comments and suggestions, which have greatly helped to improve the quality and clarity of this paper. Their insightful feedback allowed us to refine our methodology, enhance the presentation of the results, and strengthen the overall contribution of this work. We deeply appreciate their time and effort in providing thoughtful reviews.

Conflicts of Interest: Author Mr. Lei Ouyang was employed by the company "North Information Control Research Academy Group Company Limited". The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Dendi, S.V.R.; Channappayya, S.S. No-reference video quality assessment using natural spatiotemporal scene statistics. *IEEE Trans. Image Process.* **2020**, *29*, 5612–5624. [CrossRef] [PubMed]
2. Wu, J.; Ma, J.; Liang, F.; Dong, W.; Shi, G.; Lin, W. End-to-end blind image quality prediction with cascaded deep neural network. *IEEE Trans. Image Process.* **2020**, *29*, 7414–7426. [CrossRef]
3. Chen, D.; Wang, Y.; Gao, W. No-reference image quality assessment: An attention driven approach. *IEEE Trans. Image Process.* **2020**, *29*, 6496–6506. [CrossRef] [PubMed]
4. Hosu, V.; Lin, H.; Sziranyi, T.; Saupe, D. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Trans. Image Process.* **2020**, *29*, 4041–4056. [CrossRef]
5. Zhou, W.; Jiang, Q.; Wang, Y.; Chen, Z.; Li, W. Blind quality assessment for image super resolution using deep two-stream convolutional networks. *Inf. Sci.* **2020**, *528*, 205–218. [CrossRef]
6. Ma, J.; Wu, J.; Li, L.; Dong, W.; Xie, X.; Shi, G.; Lin, W. Blind image quality assessment with active inference. *IEEE Trans. Image Process.* **2021**, *30*, 3650–3663. [CrossRef]
7. 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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 14–19 June 2020; IEEE: New York, NY, USA, 2020.
8. Shi, C.; Lin, Y. Full reference image quality assessment based on visual salience with color appearance and gradient similarity. *IEEE Access* **2020**, *8*, 97310–97320. [CrossRef]
9. Peng, S.; Jiang, W.; Pi, H.; Li, X.; Bao, H.; Zhou, X. Deep Snake for Real-Time Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8530–8539.
10. Pan, Z.; Yuan, F.; Lei, J.; Fang, Y.; Shao, X.; Kwong, S. VCRNet: Visual Compensation Restoration Network for No-Reference Image Quality Assessment. *IEEE Trans. Image Process.* **2022**, *31*, 1613–1627. [CrossRef] [PubMed]
11. Li, S.; Wang, M. No-Reference Stereoscopic Image Quality Assessment Based on Convolutional Neural Network with A Long-Term Feature Fusion. In Proceedings of the IEEE International Conference on Visual Communications and Image Processing, Macau, China, 1–4 December 2020; pp. 318–321.

12. Sun, G.; Shi, B.; Chen, X.; Krylov, A.S.; Ding, Y. Learning Local Quality-Aware Structures of Salient Regions for Stereoscopic Images via Deep Neural Networks. *IEEE Trans. Multimed.* **2020**, *22*, 2938–2949. [CrossRef]
13. Liu, X.; Zhang, L.; Lu, K. A 3D image quality assessment method based on vector information and SVD of quaternion matrix under cloud computing environment. *IEEE Trans. Cloud Comput.* **2020**, *8*, 326–337. [CrossRef]
14. Jiang, Q.; Zhou, W.; Chai, X.; Yue, G.; Shao, F.; Chen, Z. A full-reference stereoscopic image quality measurement via hierarchical deep feature degradation fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9784–9796. [CrossRef]
15. Wan, Z.; Gu, K.; Zhao, D. Reduced reference stereoscopic image quality assessment using sparse representation and natural scene statistics. *IEEE Trans. Multimed.* **2020**, *22*, 2024–2037. [CrossRef]
16. Kroner, A.; Senden, M.; Driessens, K.; Goebel, R. Contextual encoder-decoder network for visual saliency prediction. *Neural Netw.* **2020**, *129*, 261–270. [CrossRef]
17. Jia, H.; Wang, T.-h.; Fu, P. Multi-feature fusion based image quality assessment method. *J. Pattern Recognit. Artif. Intell.* **2019**, *32*, 669–675.
18. Kim, J.; Ncuyen, A.; Lee, S. Deep CNN-based blind image quality predictor. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 11–24. [CrossRef] [PubMed]
19. Greffier, J.; Viry, A.; Durand, Q.; Hajdu, S.D.; Frandon, J.; Beregi, J.P.; Dabli, D.; Racine, D. Brain image quality according to beam collimation width and image reconstruction algorithm: A phantom study. *Phys. Medica* **2023**, *108*, 108102558. [CrossRef] [PubMed]
20. Verma, S.; Chug, A.; Singh, A.P.; Singh, D. Plant Disease Detection and Severity Assessment Using Image Processing and Deep Learning Techniques. *SN Comput. Sci.* **2023**, *5*, 83. [CrossRef]
21. Liu, Y.; Yin, X.; Yue, G.; Zheng, Z.; Jiang, J.; He, Q.; Li, X. Blind omnidirectional image quality assessment with representative features and viewport oriented statistical features. *J. Vis. Commun. Image Represent.* **2023**, *91*, 103770. [CrossRef]
22. Duan, H.; Min, X.; Zhu, Y.; Zhai, G.; Yang, X.; Le Callet, P. Confusing Image Quality Assessment: Towards Better Augmented Reality Experience. *IEEE Trans. Image Process.* **2022**, *31*, 7206–7221. [CrossRef]
23. Chen, B.; Zhu, L.; Kong, C.; Zhu, H.; Wang, S.; Li, Z. No-Reference Image Quality Assessment by Hallucinating Pristine Features. *IEEE Trans. Image Process.* **2022**, *31*, 6139–6151. [CrossRef] [PubMed]
24. Madhusudana, P.C.; Birkbeck, N.; Wang, Y.; Adsumilli, B.; Bovik, A.C. Image Quality Assessment using Contrastive Learning. *IEEE Trans. Image Process.* **2022**, *31*, 4149–4161. [CrossRef] [PubMed]
25. Zhao, J.; Wang, Y.; Mancenido, M.V.; Chiou, E.K.; Maciejewski, R. Evaluating the Impact of Uncertainty Visualization on Model Reliance. *IEEE Trans. Vis. Comput. Graph.* **2023**, *30*, 4093–4107. [CrossRef] [PubMed]
26. Liu, X.; Li, J.; Lu, G. Modeling Realistic Clothing from a Single Image under Normal Guide. *IEEE Trans. Vis. Comput. Graph.* **2023**, *30*, 3995–4007. [CrossRef] [PubMed]
27. Wei, Y.; Wang, Z.; Wang, Z.; Dai, Y.; Ou, G.; Gao, H.; Yang, H.; Wang, Y.; Cao, C.C.; Weng, L.; et al. Visual Diagnostics of Parallel Performance in Training Large-Scale DNN Models. *IEEE Trans. Vis. Comput. Graph.* **2023**, *30*, 3915–3929. [CrossRef]
28. Giacomo, E.D.; Didimo, W.; Liotta, G.; Montecchiani, F.; Tappini, A. Comparative Study and assessment of Hybrid Visualizations of Graphs. *IEEE Trans. Vis. Comput. Graph.* **2023**. *Early Access*.
29. Deng, Z.; Chen, S.; Xie, X.; Sun, G.; Xu, M.; Weng, D.; Wu, Y. Multilevel Visual Analysis of Aggregate Geo-Networks. *IEEE Trans. Vis. Comput. Graph.* **2022**, *30*, 3135–3150. [CrossRef]
30. Rydow, E.; Borgo, R.; Fang, H.; Torsney-Weir, T.; Swallow, B.; Porphyre, T.; Turkay, C.; Chen, M. Development and assessment of Two Approaches of Visual Sensitivity Analysis to Support Epidemiological Modeling. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 1255–1265. [PubMed]
31. Xie, T.; Ma, Y.; Kang, J.; Tong, H.; Maciejewski, R. FairRankVis: A Visual Analytics Framework for Exploring Algorithmic Fairness in Graph Mining Models. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 368–377. [CrossRef] [PubMed]
32. Park, H.; Das, N.; Duggal, R.; Wright, A.P.; Shaikh, O.; Hohman, F.; Chau, D.H.P. NeuroCartography: Scalable Automatic Visual Summarization of Concepts in Deep Neural Networks. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 813–823. [CrossRef] [PubMed]