

Article

No-Reference Image Quality Assessment based on a Multitask Image Restoration Network

Fan Chen ¹, Hong Fu ², Hengyong Yu ³ and Ying Chu ^{1,*}

¹ Department of Artificial Intelligence, Shenzhen University, Shenzhen, P.R.China, 518060

² Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, P. R. China

³ Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, U.S.A, 01854

* Correspondence: chuying@szu.edu.cn

Abstract: When the image quality is evaluated, the human visual system (HVS) infers the details in the image through its internal generative mechanism. In this process, the HVS integrates both local and global information of the image, utilizes contextual information to restore the original image information, and compares it with the distorted image information for image quality evaluation. Inspired by this mechanism, a no-reference image quality assessment method is proposed based on a multitask image restoration network. The multitask image restoration network generates a pseudo-reference image as the main task and produces structural similarity index measure map as an auxiliary task. By mutually promoting the two tasks, a higher quality pseudo-reference image is generated. In addition, when predicting the image quality score, both the quality restoration features and the difference features between the distorted and reference images are used, thereby fully utilizing the information from the pseudo-reference image. To enable the model to focus on both global and local features, a multi-scale feature fusion module is proposed. Experimental results demonstrate that the proposed method achieves excellent performance on both synthetically and authentically distorted databases.

Keywords: no-reference image quality assessment; multitask learning; image restoration; multi-level features

1. Introduction

As an important information carrier, image have been widely used in various fields, such as advertising, entertainment, medicine, education, *etc.* However, factors (e.g., lighting conditions during image acquisition and stability of the transmission path) may result in the loss of original image information, which can seriously affect users' visual experience and the effectiveness of image usage [1]. Image Quality Assessment (IQA) aims to evaluate the quality of images using certain methods and provide accurate evaluation results. Accurate and efficient IQA methods are of great significance in improving user experience and optimizing the performance and accuracy of computer vision-related applications. IQA methods can be categorized into three types based on the reference image usage: full-reference IQA (FR-IQA) [2-4], reduced-reference IQA (RR-IQA) [5, 6], and no-reference IQA (NR-IQA) [7-10]. NR-IQA has higher practicality, especially in situations where it is difficult to obtain reference images in practical applications. Therefore, NR-IQA has become one of the research hotspots in the field of IQA and has important practical research value [11].

The main challenge of the NR-IQA is the lack of reference images, which makes it impossible to evaluate the quality based on the feature differences between reference and distorted images. However, to perceive distorted images, based on prior knowledge in the brain, the human visual system (HVS) utilizes an internal generative mechanism [12-14] to reconstruct the reference image as much as possible. Then, the quality of the distorted image is evaluated based on the difference between the reconstructed reference and the distorted images [15], with greater differences indicating more severe image distortion. To acquire prior knowledge of reference images, some researchers have proposed methods based on generative adversarial network (GAN) to generate pseudo-reference images [16, 17]. Subsequently, image quality assessment is performed by evaluating the difference between the distorted image and the pseudo-reference image. However, GAN-based

methods typically face two challenges. First, the training process of GANs is often unstable, making it difficult to achieve satisfactory image restoration performance. Second, when dealing with severely distorted images, GANs may struggle to effectively restore image quality.

To address the aforementioned problems, a NR-IQA method based on a multitask image restoration network (MT-IRN) is proposed. The proposed method includes a multitask image restoration sub-network and a score prediction sub-network. The multitask image restoration sub-network is used to generate pseudo-reference images and structural similarity index measure (SSIM) maps between distorted images and reference images, as well as to extract quality restoration features. The score prediction sub-network is used to extract high-level features and multi-scale content features of the distorted image, and difference features between the distorted and pseudo-reference images, and it maps these features to quality scores after concatenation. Specifically, the contributions of this paper are as follows:

- First, a multitask image restoration network is proposed to generate high quality pseudo-reference images and enhance the overall performance of the model. The multitask image restoration network has a main task to generate pseudo-reference images and an auxiliary task to generate structurally similar images, with the two tasks mutually reinforcing each other to generate higher quality pseudo-reference images.
- Second, in addition to utilize the feature differences between distorted and pseudo-reference images, quality restoration features are also employed in the model to leverage rich semantic information in image restoration features, enabling the model to exploit not only differences between pseudo-reference and distorted images, but also the semantic information within quality restoration features.
- Third, a multi-scale feature fusion module is proposed to fully fuse quality restoration features and multi-scale content features of distorted images, enabling the model to extract both global and local features simultaneously.

The rest of this paper is organized as follows. Section 2 introduces related works for NR-IQA. Section 3 provides a detailed description of the proposed method. Section 4 reports the experiment results. Section 5 concludes this paper.

2. Related work

Due to lack of reference images, many traditional NR-IQA methods focus on specific types of distortion in distorted images and propose the corresponding evaluation algorithms. For example, filtering-based methods are used to estimate noise in images [18], and sharpness and blurriness estimation algorithms are used to evaluate the quality of blurry images [19]. These methods can achieve higher accuracy if the image distortion process or type is known. In addition, some NR-IQA methods [7-9] do not target specific distortion(s), but instead extract general quality features that can describe multiple types of distortion to evaluate the quality of distorted images. The focus and difficulty of this method lies in selecting which features to measure the level of distortion. This is generally manually extracted through natural scene statistics (NSS) in traditional methods, and it can be automatically learned in deep learning-based methods [20-23].

Moorthy and Bovik [7] first proposed a blind image quality index (BIQI) for general distortion types. This method fits the wavelet decomposition coefficients of images with a Generalized Gaussian Distribution (GGD) and uses the parameters of the GGD model as features. Mittal *et al.* [8] proposed a blind/referenceless image spatial quality evaluator (BRISQUE) to utilize NSS in the spatial domain. This approach first computes the multi-scale mean-subtracted contrast-normalized (MSCN) coefficients of distorted images. Then, it fits the MSCN coefficients and their related coefficients to predict quality scores. Ghadiyaram *et al.* [9] proposed a feature maps based referenceless image quality evaluation engine (FRIQUEE) for authentic distortion assessment. The aim of this method is to capture the statistical consistency or deviations from consistency in authentically distorted images, without assuming the presence of any specific type of distortion in the image.

With the development and improvement of deep learning, its application in IQA has received increasing attention. Deep learning has powerful fitting and generalization capabilities, enabling it

to learn feature representations from large amounts of training data and associate these features with image quality. Therefore, more and more researchers have explored the use of deep learning algorithms to improve the accuracy and reliability of IQA. Kang *et al.* [20] first proposed an IQA-CNN method that employs convolutional neural network (CNN). The method includes one convolutional layer and two pooling layers, and it uses two fully connected layers to map the features to quality scores. Bosse *et al.* [21] employed deeper CNN to extract high-level features of images. The method involves constructing ten convolutional layers and two fully connected layers for feature extraction and score prediction. Results indicate that it can significantly improve the performance of the model using high-level image features. Su *et al.* [22] proposed a HyperIQA for authentically distorted images. This method predicts image quality scores based on the perception of image content. Zhang *et al.* [23] proposed a deep bilinear CNN (DB-CNN), which utilizes two streams to extract synthetic distortion features and authentic distortion features of images. Pan *et al.* [24] proposed a NR-IQA method, called blind predicting similar quality map (BPSQM), which consists of a fully convolutional neural network and a pooling network. The global convolutional network is trained using the similarity maps from traditional FR-IQA methods, enabling the network to predict the corresponding similarity quality map for distorted images. The pooling network is then employed to regress the quality map into a quality score.

Some GAN-based methods have been proposed to address the problem of lack of reference image by generating pseudo-reference images. Ren *et al.* [16] proposed restorative adversarial nets for no-reference image quality assessment (RAN4IQA), which includes a restorer, a discriminator, and a predictor. The restorer and discriminator work together to restore the quality of distorted images, and then the predictor extracts features from the distorted and pseudo-reference images to map them to quality scores. Similarly, Lin *et al.* [17] proposed a Hall-IQA, which also generates pseudo-reference images using GAN and uses the difference image between the distorted and pseudo-reference images as input to a regression network for image quality score prediction. Both RAN4IQA and Hall-IQA have achieved good results, but GAN-based methods have a shaky training process and are difficult to achieve good image restoration performance, especially when faced with severely distorted images. Pan *et al.* [25] proposed a method based on visual compensation restoration (VCR), named as VCRNet, which uses the features in the image reconstruction process to avoid the performance degradation caused by the suboptimal quality of the pseudo-reference image. However, it does not use the differential features between the pseudo-reference and the distorted images.

Although the aforementioned methods have achieved remarkable results, there are still rooms for further improvement. Inspired by BPSQM, this paper proposes an NR-IQA method based on a multitask image restoration network, in which the main task is to generate pseudo-reference image and the auxiliary task is to generate the SSIM map. By leveraging the mutual promotion between the two tasks, high quality pseudo-reference image can be generated. Furthermore, when predicting quality scores, not only the quality restoration features but also the difference features between the distorted and pseudo-reference images are utilized. This fully exploits the information from the pseudo-reference image. Finally, a multi-scale feature fusion module is designed to enable the model to simultaneously focus on both global and local features.

3. Proposed Method

In this paper, an NR-IQA method is proposed based on a multitask image restoration network. The proposed method consists of a multitask image restoration sub-network and a score prediction sub-network, as shown in Figure 1. The multitask image restoration sub-network is used to generate pseudo-reference images and SSIM maps between the reference images and distorted images, and to extract quality restoration features. This enables the model to not only use the differences between the pseudo-reference and distorted images but also leverage rich semantic information from the quality restoration features. The score prediction sub-network is used to extract high-level features of the distorted images, multi-scale content features of the distorted images, and difference features between the distorted and pseudo-reference images. These features are concatenated and mapped to quality scores.

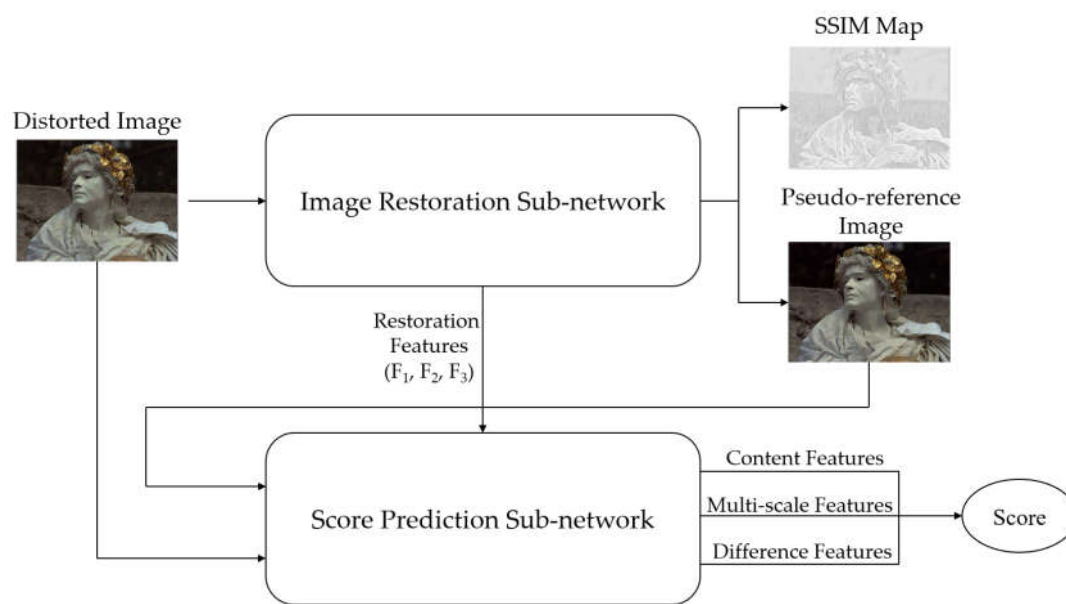


Figure 1. The architecture of the proposed method.

3.1 Multitask Image Restoration Sub-network

The quality of pseudo-reference images is crucial for the performance of the model. The proposed method utilizes a multitask image restoration network to generate higher-quality pseudo-reference images, thereby improving the overall performance of the model. Inspired by the BPSQM, the multitask image restoration sub-network takes the generation of pseudo-reference images as the main task and the generation of SSIM map as an auxiliary task, generating higher-quality-pseudo reference images through the mutual promotion of the two tasks. The SSIM map provides rich structural information for the reference image, and therefore, generating the auxiliary task that allows the image restoration sub-network to learn the structural information of the reference image, and thus improving the structural similarity between the pseudo-reference and reference images.

The proposed method employs a U-Net architecture [26-28] to construct the image restoration sub-network. U-Net is an encoder-decoder structure widely used in the fields of image segmentation and restoration. Its skip connection structure concatenates low-level features with high-level features, allowing the decoder to retain more detailed information during the upsampling process. The structure of the multitask image restoration sub-network is illustrated in Figure 2. The multitask learning mode adopts a hard sharing mode [29, 30], in which the two tasks share the parameters of

the feature extraction layer, and then different convolutional layers are set for each task to achieve their respective goals.

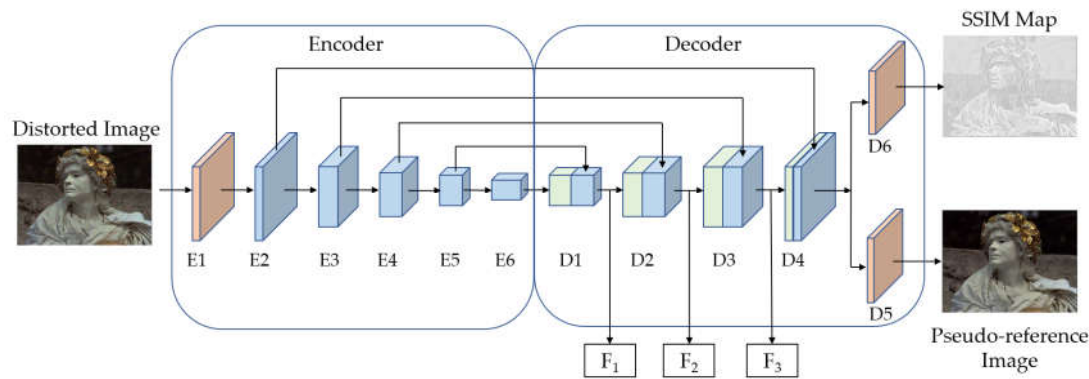


Figure 2. The architecture of the multitask image restoration sub-network.

The multitask image restoration sub-network takes the distorted image as an input, and it outputs the pseudo-reference image, the SSIM map, and the quality restoration features F_1 , F_2 , and F_3 . The encoder consists of six convolutional modules, E1-E6. E1 is a single-layer convolutional layer composed of 3×3 convolution kernels with a stride of 1. To avoid gradient vanishing [31] when deepening the model and to reuse low-level features, E2-E6 use residual convolutional blocks for downsampling. The specific structures of two types of the residual blocks are shown in Figure 3. E2 uses the residual block 1 structure, which contains two 3×3 convolutional layers with a stride of 1, and this structure does not change the size of the input feature map. E3-E6 use the residual block 2 structure, which contains one 3×3 convolutional layer with a stride of 2 and one 3×3 convolutional layer with a stride of 1. The output feature map size is half of the input feature map size. Since the sizes of the input and output feature maps are inconsistent, a 1×1 convolution with a stride of 2 is performed on the input feature map during the residual connection to match their sizes.

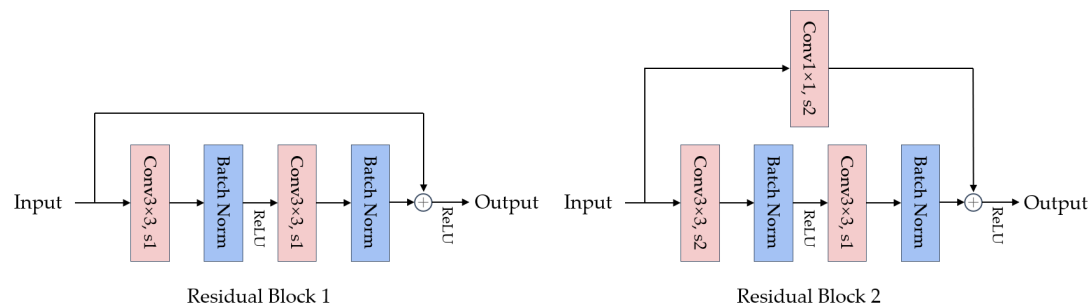


Figure 3. The structures of the residual blocks.

The decoder consists of six deconvolutional layers, D1-D6, which perform upsampling on the high-level features of distorted images to generate the pseudo-reference image and SSIM map. Through multi-level skip connections, the decoder effectively preserves the detailed information in the input image while avoiding the loss of feature details caused by pooling layers, thus improving the image restoration performance. D1-D6 are all composed of 3×3 deconvolutional layers, where the stride of D1-D4 is 2, and the stride of D5 and D6 is 1. The specific structural parameters of the image restoration sub-network are summarized in Table 1.

Table 1. The structural parameters of the image restoration sub-network, where $W \times H \times C$ denotes the width, height and channels of the feature map, respectively.

	Module	Layer	Input Size	Output Size
Encoder	E1	Conv3×3, s1	$W \times H \times 3$	$W \times H \times 16$
	E2	Conv3×3, s1 Conv3×3, s1	$W \times H \times 16$	$W \times H \times 16$
	E3	Conv3×3, s2 Conv3×3, s1	$W \times H \times 16$	$W/2 \times H/2 \times 32$
	E4	Conv3×3, s2 Conv3×3, s1	$W/2 \times H/2 \times 32$	$W/4 \times H/4 \times 64$
	E5	Conv3×3, s2 Conv3×3, s1	$W/4 \times H/4 \times 64$	$W/8 \times H/8 \times 128$
	E6	Conv3×3, s2 Conv3×3, s1	$W/8 \times H/8 \times 128$	$W/16 \times H/16 \times 256$
Decoder	D1	Deconv3×3, s2	$W/16 \times H/16 \times 256$	$W/8 \times H/8 \times 128$
	D2	Deconv3×3, s2	$W/8 \times H/8 \times 256$	$W/4 \times H/4 \times 64$
	D3	Deconv3×3, s2	$W/4 \times H/4 \times 128$	$W/2 \times H/2 \times 32$
	D4	Deconv3×3, s2	$W/2 \times H/2 \times 64$	$W \times H \times 16$
	D5	Deconv3×3, s1	$W \times H \times 16$	$W \times H \times 3$
	D6	Deconv3×3, s1	$W \times H \times 16$	$W \times H \times 1$

3.2 Score Prediction Sub-network

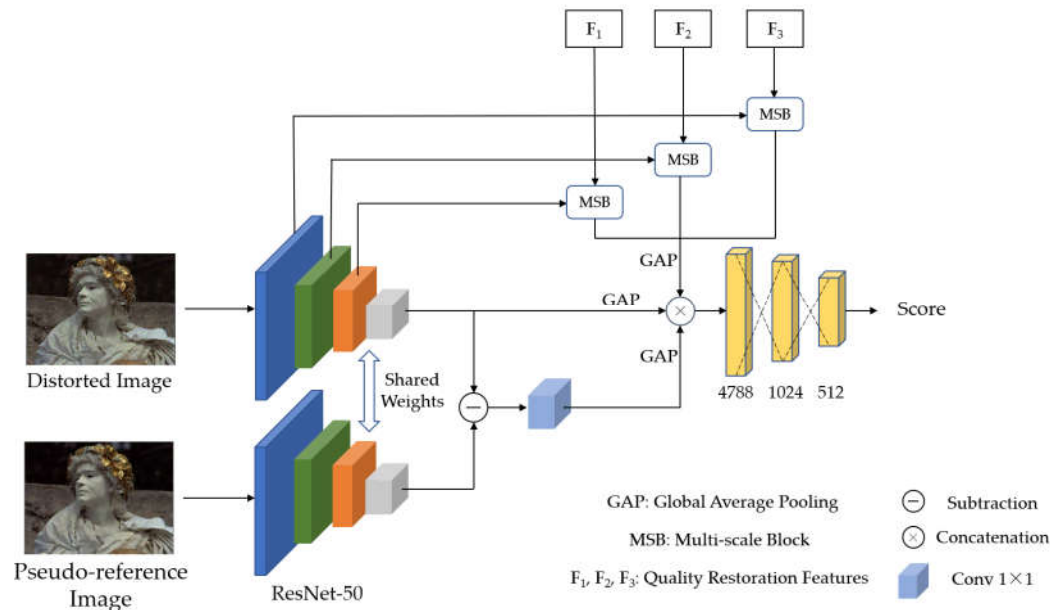


Figure 4. The architecture of the score prediction sub-network.

A score prediction sub-network is used to extract features from distorted and pseudo-reference images and to predict image quality scores, as shown in Figure 4. The score prediction sub-network takes distorted image, pseudo-reference image, and quality restoration features F_1 , F_2 , F_3 as inputs, and a pre-trained ResNet-50 [32] on ImageNet [33] is used as a feature extractor to extract the content features of the image. First, the outputs of Conv2_10, Conv3_12, and Conv4_18 in ResNet-50 are used as multi-scale features of the distorted image, and are respectively fully fused with the quality restoration features F_1 , F_2 , F_3 extracted by the image restoration sub-network through the multi-scale features fusion block. Then, the high-level features of the distorted image, namely the output of Conv5_9 in ResNet-50, are subtracted from the high-level features of the pseudo-reference image. After dimension reduction by a 1x1 convolution, the difference feature between the distorted and pseudo-reference images is obtained. Finally, the high-level feature, multi-scale feature, and

difference feature are globally average-pooled, concatenated, and mapped to quality scores by fully connected layers.

A multi-scale features fusion module is used to effectively fuse the multi-scale content features and the quality restoration features, as shown in Figure 5. Considering that these two features come from different network structures, there may have different feature scales. Therefore, multi-scale convolutions are applied to each feature separately to extract features at various scales, as shown in Equation 1. The multi-scale convolution uses two concatenated 3×3 convolutions to achieve a receptive field size of 5×5 while reducing the number of parameters.

$$MC(F_i) = Conv_{3 \times 3}(F_i) \otimes Conv_{3 \times 3}(Conv_{3 \times 3}(F_i)) \otimes Conv_{1 \times 1}(F_i), \quad (1)$$

where $MC(\cdot)$ denotes the multi-scale convolution operation, F_i represents the input feature map, $Conv_{3 \times 3}(\cdot)$ represents the 3×3 convolution layer with stride 1 and padding 1, $Conv_{1 \times 1}(\cdot)$ represents the 1×1 convolution layer with stride 1, and \otimes indicates the concatenation operation.

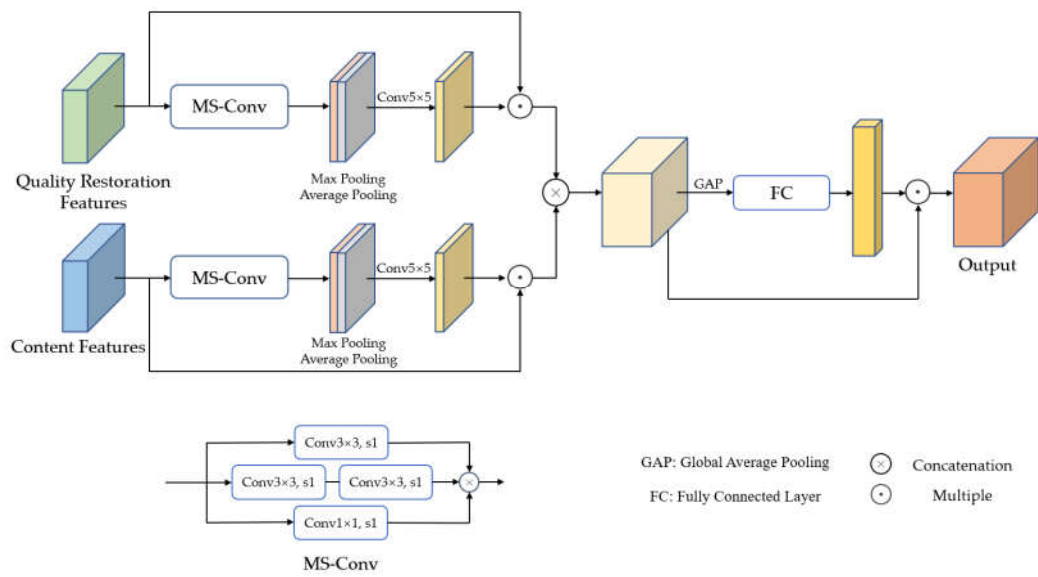


Figure 5. The structure of the multi-scale features fusion module.

Subsequently, the spatial attention module is utilized to extract prominent spatial features from two different features. Spatial attention module [34] improves model's performance by retaining critical information while ignoring unimportant regions. Specifically, we first perform max-pooling and average-pooling operations on the multi-scale convolutional feature maps to generate two 2D feature maps. We concatenate these two feature maps and learn spatial weights through a 5×5 convolution operation, followed by element-wise multiplication with the input feature map to obtain the final spatial attention feature map. Finally, the two spatial attention feature maps are concatenated, as shown in Equation 2:

$$\begin{cases} F_{MR} = MC(F_R) \\ F_{MC} = MC(F_C) \\ F_{WR} = Conv_{5 \times 5}(MaxPool(F_{MR}) \otimes AvgPool(F_{MR})), \\ F_{WC} = Conv_{5 \times 5}(MaxPool(F_{MC}) \otimes AvgPool(F_{MC})), \\ F_S = (F_R \odot F_{WR}) \otimes (F_C \odot F_{WC}) \end{cases} \quad (2)$$

where F_R denotes the quality restoration feature maps, and F_C denotes the content feature maps. $Conv_{5 \times 5}(\cdot)$ represents the 5×5 convolution layer, $MaxPool(\cdot)$ and $AvgPool(\cdot)$ represent the max pooling and average pooling operation, respectively. \otimes indicates the concatenation operation, and \odot indicates the multiple operation.

To further fuse the spatial attention feature maps, a channel attention module [35] is utilized to learn the importance of the concatenated feature maps across different channels to better capture the

relationships between different channels. The concatenated feature maps are processed using global average pooling to obtain a one-dimensional vector, and then a fully connected layer is applied to learn the weights of each channel, generating a weight vector. The weight vector is multiplied with the concatenated feature maps to obtain a fully fused output feature, as shown in Equation 3:

$$\begin{cases} F_{WS} = GAP(F_S) \\ W = FC(F_{WS}) \\ F_M = F_S \odot W \end{cases}, \quad (3)$$

where $GAP(\cdot)$ denotes the global average pooling, $FC(\cdot)$ denotes the fully connected layer, and \odot denotes the multiple operation.

3.3 Network Training

The training of the model consists of two parts: pre-training of the image restoration sub-network and overall model training. First, we pre-train the multi-task image restoration sub-network with an auxiliary task of generating SSIM map, using the Waterloo Exploration Database [36]. We randomly crop 224x224 sized image patches from the distorted images to expand the training data, set the learning rate to 0.001 and batch size to 64, and train for 100 epochs using the Adam [37] optimizer. The loss function is the L_1 loss between the generated SSIM map and the ground truth SSIM map, as shown in Equation 4:

$$l_s = \frac{1}{N} \sum_{i=1}^N \|I_{SSIM}^{(i)} - \hat{I}_{SSIM}^{(i)}\|_{l_1}, \quad (4)$$

where N denotes the number of training image patches, $I_{SSIM}^{(i)}$ denotes the SSIM map between the i^{th} image patch and its corresponding reference image patch, $\hat{I}_{SSIM}^{(i)}$ denotes the predicted SSIM map of the i^{th} image patch by the model, and l_1 denotes the l_1 -norm.

After the training of the auxiliary task, the image restoration sub-network has well learned the structural information in the reference images. Based on this, the main task of generating pseudo-reference images can be trained. Training is conducted by randomly cropping 224x224 sized image patches from distorted images in the Waterloo Exploration Database. The learning rate is set to 0.0001, the batch size is set to 64, and the Adam optimizer is used to optimize the network for 50 epochs. The loss function is the L_1 loss between the pseudo-reference and the reference images, as shown in Equation 5:

$$l_r = \frac{1}{N} \sum_{i=1}^N \|I_r^{(i)} - \hat{I}_r^{(i)}\|_{l_1}, \quad (5)$$

where N denotes the number of training image patches, $I_r^{(i)}$ denotes the i^{th} reference image patch, $\hat{I}_r^{(i)}$ denotes the i^{th} pseudo-reference image patch generated by the network, and l_1 denotes the l_1 -norm.

After training the image restoration sub-network, the entire model needs to be trained on the target database. During the first 10 epochs of training, the parameters of the image restoration sub-network are frozen, and only the score prediction sub-network is trained. Then, the entire network is finely trained for another 40 epochs. To perform data augmentation, following the strategy from [17] and [38] the images are randomly horizontally flipped during training, and 5 randomly sampled 224x224 image patches are extracted from each image to increase the number of training samples. The quality scores of the image patches are the same as those of the corresponding distorted images. During testing, 5 randomly sampled 224x224 image patches are also extracted from each testing image, and their quality scores are predicted. The mean of these scores is used as the quality score of the testing image.

The L_1 loss is used to train the entire model, as shown in Equation 7:

$$l = \frac{1}{N} \sum_{i=1}^N \|q_i - \hat{q}_i\|_{l_1}, \quad (6)$$

where N denotes the number of the image patches, q_i denotes the score of i^{th} image patch predicted by the model, and \hat{q}_i denotes the ground truth score of i^{th} image patch.

The Adam optimizer is used to optimize the parameters of the entire model with a weight decay rate of 5×10^{-4} . The model is trained for 50 epochs with a batch size of 48 and an initial learning rate of 5×10^{-5} . The learning rate is multiplied by 0.9 every 10 epochs during training. The proposed method is implemented with Pytorch and the experiments are conducted on NVIDIA 3080Ti GPU.

4. Experiments

4.1 Databases and Experimental Protocols

To evaluate the performance of the proposed method, experiments are conducted on both synthetically distorted databases and authentically distorted databases, and the state-of-the-art methods are compared. The synthetically distorted databases include LIVE [39], CSIQ [40], TID2013 [41] and KADID-10k [42], while the authentically distorted databases include LIVEC [43] and KonIQ-10k [44]. The details of the databases are summarized in Table 2.

Table 2. Details of the IQA databases.

Database	Ref. imgs	Dist. imgs	Dist. types	Score's type
LIVE [39]	29	779	5	DMOS
CSIQ [40]	30	866	6	DMOS
TID2013 [41]	25	3000	24	MOS
KADID-10k [42]	81	10125	25	DMOS
LIVEC [43]	/	1162	/	MOS
KonIQ-10k [44]	/	10073	/	MOS

For synthetically distorted databases, 80% images are used as a training set and the remaining 20% as a testing set, divided by the reference images, to avoid image content overlap between the training and testing sets. For authentically distorted databases, the training and testing sets are directly divided into 80% and 20% proportions. For each dataset, the random partitioning process is repeated 10 times according to the aforementioned rules, and the median of the results from the 10 experiments is taken as the final result. Spearman's rank correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC) are used to evaluate the performance of the proposed method. The SROCC measures the monotonicity between the predicted scores and the ground truth scores, while the PLCC measures the linear correlation between them. Both SROCC and PLCC have a range of $[-1, 1]$, with a larger absolute value indicating better performance of the model.

4.2 Performance on Individual Database

The experiments on individual database utilize four synthetically distorted databases, including LIVE, CSIQ, TID2013, and KADID, as well as two authentically distorted databases, LIVEC and KonIQ. The results are summarized in Tables 3 and 4. The methods compared with the proposed method include three traditional methods (PSNR, SSIM [2], and BRISQUE [13]), seven deep learning-based methods (IQA-CNN [20], BIECON [45], MEON [46], DIQaM-NR [21], HyperIQA [22], DB-CNN [23], and TS-CNN [47]), two GAN-based methods (RAN4IQA [16] and Hall-IQA [17]), and a visual compensation restoration-based method, namely VCRNet [25].

Table 3. The SROCC and PLCC results on synthetically distorted databases. The top two results are shown in bold font.

Method	LIVE		CSIQ		TID2013		KADID	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
PSNR	0.866	0.856	0.806	0.800	0.636	0.706	0.674	0.681
SSIM [2]	0.913	0.931	0.876	0.861	0.637	0.691	0.783	0.780
BRISQUE [13]	0.940	0.942	0.746	0.829	0.604	0.694	0.519	0.554
IQA-CNN [20]	0.956	0.953	0.876	0.905	0.701	0.752	0.651	0.607
BIECON [45]	0.961	0.962	0.825	0.838	0.717	0.762	0.685	0.691
MEON [46]	0.943	0.954	0.839	0.850	0.828	0.811	0.813	0.822
DIQaM-NR [21]	0.960	0.972	0.901	0.908	0.835	0.855	0.840	0.843
HyperIQA [22]	0.962	0.966	0.923	0.942	0.840	0.858	0.852	0.845
DB-CNN [23]	0.968	0.971	0.946	0.959	0.816	0.865	0.801	0.806
TS-CNN [47]	0.969	0.978	0.892	0.905	0.779	0.784	0.745	0.744
RAN4IQA [16]	0.961	0.962	0.914	0.938	0.820	0.859	/	/
Hall-IQA [17]	0.976	0.978	0.892	0.906	0.879	0.880	/	/
VCRNet [25]	0.973	0.974	0.943	0.955	0.846	0.846	0.850	0.857
MT-IRN	0.969	0.970	0.928	0.943	0.852	0.877	0.877	0.878

Table 4. The SROCC and PLCC results on authentically distorted databases. The top two results are shown in bold font.

Method	LIVEC		KonIQ	
	SROCC	PLCC	SROCC	PLCC
BRISQUE [13]	0.607	0.585	0.673	0.692
IQA-CNN [20]	0.516	0.536	0.655	0.671
BIECON [45]	0.595	0.613	0.618	0.651
MEON [46]	0.693	0.688	0.754	0.760
DIQaM-NR [21]	0.606	0.601	0.722	0.736
HyperIQA [22]	0.859	0.882	0.906	0.917
DB-CNN [23]	0.851	0.869	0.875	0.884
TS-CNN [47]	0.655	0.667	0.722	0.729
RAN4IQA [16]	0.586	0.612	0.752	0.763
VCRNet [25]	0.856	0.865	0.894	0.909
MT-IRN	0.865	0.872	0.899	0.912

From the experimental results in Tables 3 and 4, it can be observed that the proposed method outperforms the traditional methods on all six databases. This is mainly due to the powerful learning ability of deep learning, which enables the model to extract richer features. Compared with deep learning-based methods, our method performs better than most methods on the synthetically distorted databases, except DB-CNN on the CSIQ dataset. On the authentically distorted databases, our method performs slightly lower than Hyper-IQA, but it still achieves a better performance than the GAN-based RAN4IQA. Particularly, on the LIVEC, our method's SROCC is approximately 45.9% higher than that of RAN4IQA. This is mainly because RAN4IQA is pre-trained on the synthetically distorted databases, while our method's score prediction sub-network uses ResNet-50, which is pre-trained on ImageNet and has a stronger ability to extract features from authentic distortions. Compared with the visual compensation restoration-based VCRNet, our method lags behind the LIVE and CSIQ, but it still maintains a leading performance on the other four databases. This is

because our method not only uses quality restoration features but also utilizes the difference features between the distorted and the pseudo-reference images. Additionally, the quality of the pseudo-reference image is improved by a multi-task restoration network, which further enhances the model's performance.

4.3 Performance on Individual Distortion Types

To compare the performance of the proposed method with the state-of-the-art methods on specific types of distortions, experiments are conducted on LIVE, CSIQ and TID2013. The SROCC results of the experiments are summarized in Tables 5.

Table 5. The SROCC results of the individual distortion type on the LIVE, CSIQ and TID2013. The top two results are shown in bold font. "Count" refers to the number of times a method achieves top two results.

	Dist. type	IQA-CNN [20]	DIQA [48]	HyperIQA [22]	RAN4IQA [16]	Hall-IQA [17]	VCRNet [25]	MT-IRN
LIVE	JP2K	0.936	0.961	0.949	0.958	0.969	0.975	0.977
	JPEG	0.965	0.976	0.961	0.923	0.975	0.979	0.980
	WN	0.974	0.986	0.982	0.973	0.992	0.988	0.985
	GB	0.952	0.962	0.926	0.964	0.973	0.978	0.973
	FF	0.906	0.912	0.934	0.893	0.953	0.962	0.965
CSIQ	JP2K	0.930	0.927	0.960	0.927	0.924	0.962	0.963
	JPEG	0.915	0.931	0.934	0.904	0.933	0.956	0.958
	WN	0.919	0.835	0.927	0.923	0.942	0.939	0.934
	GB	0.918	0.870	0.915	0.889	0.901	0.950	0.942
	PN	0.900	0.893	0.931	0.844	0.842	0.899	0.946
	CC	0.786	0.718	0.874	0.860	0.861	0.919	0.906
TID2013	AGN	0.784	0.916	0.942	0.866	0.923	0.844	0.892
	ANC	0.758	0.755	0.916	0.753	0.880	0.785	0.768
	SCN	0.762	0.878	0.947	0.842	0.945	0.787	0.961
	MN	0.776	0.734	0.801	0.462	0.673	0.795	0.781
	HFN	0.816	0.939	0.955	0.908	0.955	0.942	0.894
	IN	0.807	0.844	0.855	0.855	0.810	0.876	0.892
	QN	0.616	0.858	0.726	0.849	0.831	0.847	0.875
	GB	0.921	0.920	0.969	0.833	0.832	0.906	0.899
	DEN	0.872	0.788	0.941	0.839	0.957	0.937	0.880
	JPEG	0.874	0.892	0.898	0.939	0.914	0.934	0.897
	JP2K	0.910	0.812	0.947	0.912	0.624	0.906	0.918
	JGTE	0.686	0.862	0.934	0.566	0.460	0.762	0.852
	J2TE	0.678	0.813	0.892	0.778	0.782	0.865	0.892
	NPN	0.286	0.160	0.808	0.234	0.664	0.457	0.596
	BW	0.219	0.408	0.361	0.339	0.122	0.601	0.728
	MS	0.565	0.300	0.374	0.135	0.182	0.509	0.542
	CC	0.182	0.447	0.753	0.578	0.376	0.595	0.786
	CCS	0.081	0.151	0.857	0.484	0.156	0.855	0.719
	MGN	0.644	0.904	0.899	0.787	0.850	0.845	0.900
	CN	0.534	0.656	0.960	0.819	0.614	0.804	0.840
	LCNI	0.810	0.830	0.897	0.895	0.852	0.816	0.913
	ICQD	0.272	0.937	0.901	0.822	0.911	0.945	0.867
	CHA	0.892	0.757	0.870	0.762	0.381	0.932	0.828
	SSR	0.910	0.909	0.910	0.917	0.616	0.948	0.922
Count		2	3	17	1	8	18	22

From Table 5, it can be seen that the proposed method achieves the best performance on the JP2K, JPEG, GB, and FF distortions in the LIVE database. In terms of the WN distortion, the performance of the proposed method is not as good as that of Hall-IQA and VCRNet, but it still outperforms RAN4IQA and other deep learning-based algorithms. Overall, the proposed method exhibits outstanding performance on all five types of distortions in the LIVE dataset, and its performance is stable across various types of distortions without any obvious weaknesses. For the CSIQ database, the proposed method achieves the best performance on the JP2K, JPEG, GB, and PN distortion types. For the CC distortion, most methods have difficulties in achieving an SROCC of 0.900, whereas our proposed method achieves an SROCC of 0.906, second only to VCRNet. Although our method does not achieve the top two performances on the WN distortions, it still remains highly competitive, with a difference of only 0.014 compared to the best method. As for TID2013, most methods have difficulty achieving satisfactory results and cannot reach an SROCC of 0.500 for the complex distortion types such as NPN, BW, MS, CC, and CCS, while our proposed method still achieves relatively good results, with SROCCs of 0.596, 0.728, 0.542, 0.786, and 0.719, respectively.

Overall, the proposed method achieves the top two performances on 22 out of 35 distortion types, outperforming other methods. This demonstrates good performance for specific distortion types, even when facing relatively complex distortions. This is mainly due to the multitask restoration network used in our method, which improves the quality of the generated pseudo-reference images through the mutual promotion of the two tasks. In addition, our method not only uses differential features but also uses quality restoration features, enabling the score prediction sub-network to make more comprehensive and accurate predictions of image quality scores.

4.4 Performance Across Different Databases

Table 6 presents the SROCC results of cross database test on the LIVE, CSIQ, TID2013, and LIVEC, to test the generalization performance of the proposed method and compare it with the state-of-the-art methods.

Table 6. The SROCC results of cross database test. The top two results are shown in bold font.

Training		LIVE			CSIQ	
Testing	CSIQ	TID2013	LIVEC	LIVE	TID2013	LIVEC
DIIVINE [49]	0.582	0.373	0.300	0.815	0.419	0.366
CORNIA [50]	0.620	0.382	0.431	0.843	0.331	0.393
HOSA [51]	0.598	0.470	0.455	0.770	0.341	0.309
DB-CNN [23]	0.758	0.524	0.567	0.877	0.540	0.452
RAN4IQA [16]	0.632	0.462	0.157	0.806	0.471	0.116
Hall-IQA [17]	0.668	0.486	0.126	0.833	0.491	0.107
VCRNet [25]	0.768	0.502	0.615	0.886	0.542	0.463
MT-IRN	0.783	0.565	0.600	0.892	0.573	0.467
Training		TID2013			LIVEC	
Testing	LIVE	CSIQ	LIVEC	LIVE	CSIQ	TID2013
DIIVINE [49]	0.714	0.585	0.230	0.362	0.417	0.337
CORNIA [50]	0.829	0.662	0.267	0.578	0.456	0.403
HOSA [51]	0.844	0.609	0.253	0.537	0.336	0.399
DB-CNN [23]	0.891	0.807	0.457	0.746	0.697	0.424
RAN4IQA [16]	0.795	0.673	0.101	0.297	0.286	0.153
Hall-IQA [17]	0.786	0.683	0.116	-	-	-
VCRNet [25]	0.822	0.721	0.307	0.746	0.566	0.416
MT-IRN	0.897	0.739	0.375	0.758	0.546	0.419

Overall, in 12 tests, the proposed method achieves top two performance in 11 of them, outperforming the other methods. When cross database test is conducted on the synthetically distorted databases of LIVE, CSIQ, and TID2013, most methods achieve good performance as the

distortion types are relatively similar among the databases. However, the TID2013 contains more complex distortion types, and the performance of the model will decline when it is tested on this dataset. Nevertheless, the proposed method still achieves the highest SROCC, demonstrating its good generalization performance. When cross database test between synthetically and authentically distorted databases is conducted, many methods struggle to achieve good performance. However, the proposed method achieves top two performance in all the tests, surpassing other deep learning-based and GAN-based methods.

4.5 Ablation Experiments

To evaluate the impact of each module on the performance of the proposed method, ablation experiments are conducted on the LIVE, CSIQ, and LIVEC databases, and the results are summarized in Table 7.

First, the score prediction sub-network with only distorted image as input is used as the baseline, and its performance is the worst, with SROCC of 0.950, 0.894, and 0.820 on the three databases, respectively. Then, the single-task image quality restoration sub-network is added, and the image restoration features are directly concatenated with the multi-scale content features of the distorted images. At this point, the model is able to utilize some information from pseudo-reference images, resulting in an improvement in performance, with SROCC increased by 0.008, 0.013, and 0.013, respectively. Next, the multi-task image quality restoration sub-network is used, but only the image restoration features are used. The quality of the pseudo-reference images is improved, resulting in an improvement in model performance, with SROCC increased by 0.003, 0.013, and 0.009, respectively. Then, the image difference feature is introduced, allowing the model to more fully utilize the information from the pseudo-reference images, resulting in further improvements in SROCCs of 0.004, 0.006, and 0.005, respectively. Finally, the multi-scale feature fusion module is introduced, allowing for the full fusion of image multi-scale content features and restoration features, and the model's performance reaches its best, with SROCCs improved by 0.004, 0.008, and 0.005, respectively.

In summary, the multitask image restoration sub-network, image restoration features, image difference feature, and multi-scale feature fusion module proposed in this paper can effectively improve the model's performance, as evidenced by the results of the above experiments.

Table 7. The SROCC results of the ablation experiments.

Baseline	✓	✓	✓	✓	✓
Single-task Restoration Sub-network		✓			
Multitask Restoration Sub-network			✓	✓	✓
Image Restoration Feature		✓	✓	✓	✓
Difference Feature				✓	✓
Multi-scale Feature Fusion Module					✓
LIVE	0.950	0.958	0.961	0.965	0.969
CSIQ	0.894	0.907	0.914	0.920	0.928
LIVEC	0.820	0.833	0.842	0.847	0.852

4.6 Performance of Image Restoration

To evaluate the performance of the image restoration subnetwork, image restoration experiments are conducted on the LIVE, CSIQ, and TID2013 datasets. Average PSNR and average SSIM of the distorted and pseudo-reference images are used to assess the restoration performance. The performance of single-task and multitask image restoration networks are tested separately, and the experimental results are summarized in Table 8.

From Table 8, it can be observed that the average PSNR and average SSIM of the multitask generated reference images are higher than those of the single-task generated reference images and distorted images on all three datasets. This suggests that the multitask image restoration network

exhibits superior quality restoration performance, and the introduction of multi-tasking effectively enhances the quality of generated reference images.

Table 8. Average PSNR and SSIM of pseudo-reference images and distorted images.

	LIVE		CSIQ		TID2013	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Distorted Images	27.499	0.715	27.433	0.770	26.864	0.773
Single-task Pseudo-reference Images	28.707	0.732	29.238	0.788	27.992	0.816
Multitask Pseudo-reference Images	29.085	0.751	29.607	0.800	29.009	0.827

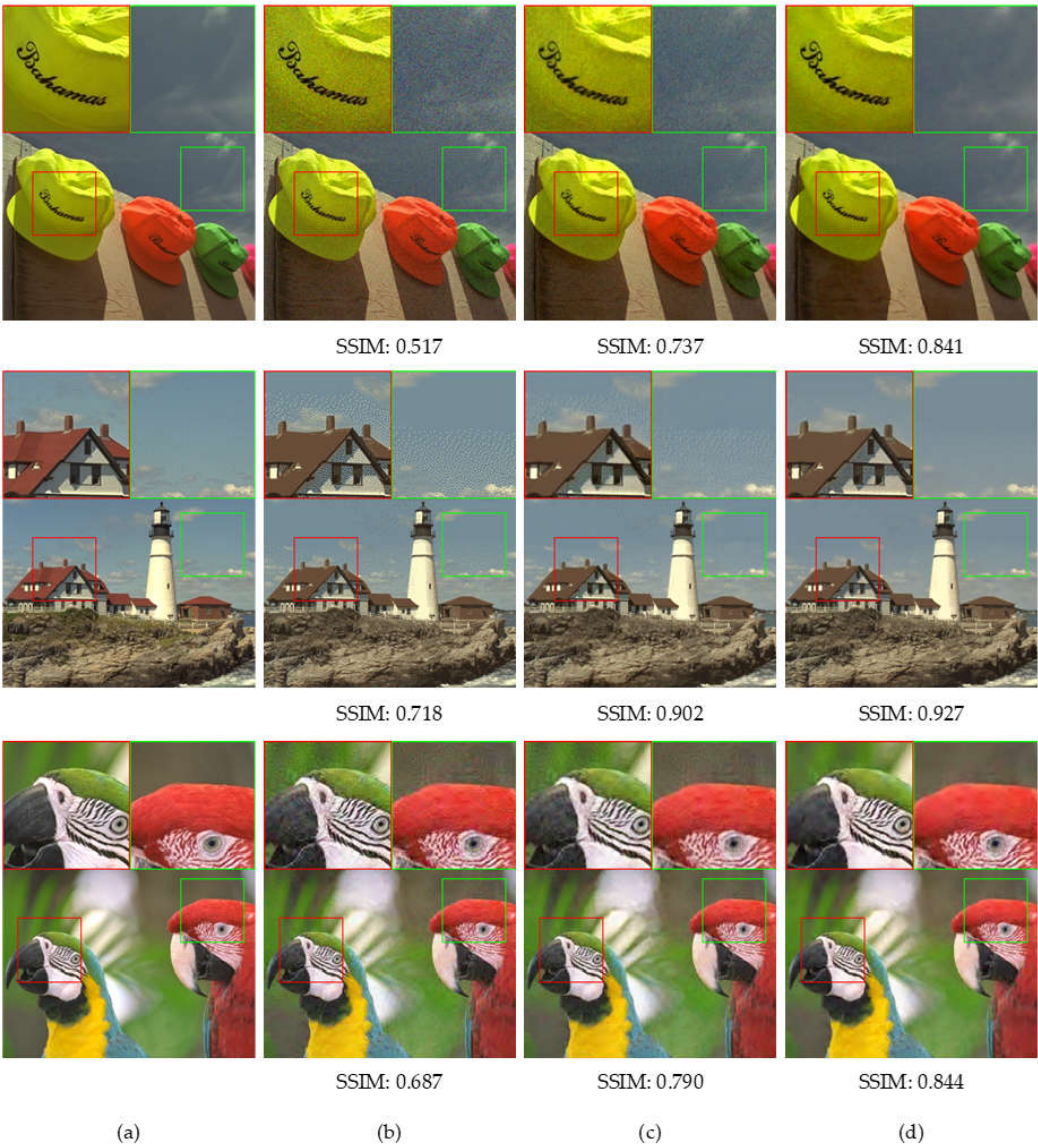


Figure 6. Comparison of pseudo-reference image quality. (a) are reference images, (b) are distorted images, (c) are single-task generated pseudo-reference images, and (d) are multitask generated pseudo-reference images.

Figure 6 shows a comparison between single-task generated pseudo-reference images and multitask generated pseudo-reference images. From Figure 6, it can be visually observed that the multitask generated pseudo-reference images exhibit significant advantages in terms of visual perceptual quality. Compared with the single-task generated pseudo-reference images and distorted images, the image quality of multitask generated pseudo-reference images is closer to the reference image. This suggests that in the framework of multitask learning, the generation of reference images

can better restore the visual quality and perceptual details of images, thereby improving the quality and usability of pseudo-reference images.

5. Conclusions

In this paper, inspired by the internal generative mechanism of HVS, an NR-IQA method is proposed based on a multitask image restoration network. The method consists of a multitask image restoration sub-network and a score prediction sub-network. First, a multitask image restoration sub-network is employed to restore the distorted image to generate a higher quality pseudo-reference image and extract quality restoration features. Second, a score prediction sub-network is used to extract the high-level features of the distorted image and difference features between the distorted image and the pseudo-reference image, and then fuse the multi-scale features of the distorted image with the quality restoration features in the image restoration sub-network. Finally, the high-level features of the distorted image, the fused multi-scale features, and the difference features between the distorted and the pseudo-reference images are utilized for quality score prediction. Experimental results on commonly used datasets demonstrate that the proposed method has achieved performance comparable to the state-of-the-art methods.

The proposed method has achieved excellent performance, but there are still rooms for further improvement. For instance, training the image restoration subnetwork with a more diverse set of distortion types could enhance the model's generalization ability. Additionally, currently, the pre-training of the image restoration subnetwork only employs synthetically distorted images, and how to pre-train using both synthetically and authentically distorted images is a research direction.

Author Contributions: Conceptualization, F.C.; methodology, F.C.; software, F.C.; validation, F.C.; formal analysis, F.C.; investigation, F.C.; resources, F.C.; data curation, F.C.; writing—original draft preparation, F.C.; writing—review and editing, H.F., H.Y. and Y.C.; visualization, F.C.; supervision, H.F., H.Y. and Y.C.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Stabilization Support Plan for Shenzhen Higher Education Institutions, grant number 20200812165210001.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rehman A.; Kai Z.; Zhou W. Display device-adapted video quality-of-experience assessment. Proceedings of Spie the International Society for Optical Engineering, 2015, 9394
2. Wang Z.; Bovik A C.; Sheikh H R.; E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 2004, 13(4): 600-612.
3. Wang Z.; E.P. Simoncelli.; Bovik A C. Multiscale structural similarity for image quality assessment. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2: 1398-1402.
4. Zhang L.; Zhang L.; Mou X.; Zhang D. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 2011, 20(8): 2378-2386.
5. Liu M.; Gu K.; Zhai G.; Le Callet P.; Zhang W. Perceptual reduced-reference visual quality assessment for contrast alteration. *IEEE Transactions on Broadcasting*, 2016, 63(1): 71-81.
6. Wu J.; Liu Y.; Shi G.; Lin W. Saliency change based reduced reference image quality assessment. 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017: 1-4.
7. Moorthy A K.; Bovik A C. A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 2010, 17(5): 513-516.
8. Mittal A.; Moorthy A K.; Bovik A C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 2012, 21(12): 4695-4708.
9. Ghadiyaram D.; Bovik A C. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of vision*, 2017, 17(1): 32-32.
10. Zhang L.; Zhang L.; Bovik A C. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 2015, 24(8): 2579-2591
11. Wang Z.; Bovik A C. Reduced-and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 2011, 28(6): 29-40.

12. Friston K.; Kilner J.; Harrison L. A free energy principle for the brain. *Journal of physiology-Paris*, 2006, 100(1-3): 70-87.
13. Friston K. The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, 2010, 11(2): 127-138.
14. Knill D C.; Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 2004, 27(12): 712-719.
15. Xu L.; Lin W.; Ma L.; Zhang Y.; Fang Y.; Ngan K N.; Yan Y. Free-energy principle inspired video quality metric and its use in video coding. *IEEE Transactions on Multimedia*, 2016, 18(4): 590-602.
16. Ren H.; Chen D.; Wang Y. RAN4IQA: Restorative adversarial nets for no-reference image quality assessment. Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).
17. Lin K Y.; Wang G. Hallucinated-IQA: No-reference image quality assessment via adversarial learning. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 732-741.
18. Joshi P.; Prakash S. Continuous wavelet transform based no-reference image quality assessment for blur and noise distortions. *IEEE Access*, 2018, 6: 33871-33882.
19. Li L.; Yan Y.; Lu Z.; Wu J.; Gu K.; Wang S. No-reference quality assessment of deblurred images based on natural scene statistics. *IEEE Access*, 2017, 5: 2163-2171.
20. Kang L.; Ye P.; Li Y.; Doermann D. Convolutional neural networks for no-reference image quality assessment. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1733-1740.
21. Bosse S.; Maniry D.; Müller K R.; Wiegand T.; Samek W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*, 2017, 27(1): 206-219.
22. 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. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3667-3676.
23. Zhang W.; Ma K.; Yan J.; Deng D.; Wang Z. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 30(1): 36-47.
24. Pan D.; Shi P.; Hou M.; Ying Z.; Fu S.; Zhang Y. Blind predicting similar quality map for image quality assessment. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6373-6382.
25. Pan Z.; Yuan F.; Lei J.; Fang Y.; Shao X.; Kwong S. VCRNet: Visual compensation restoration network for no-reference image quality assessment. *IEEE Transactions on Image Processing*, 2022, 31: 1613-1627.
26. Ronneberger O.; Fischer P.; Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
27. Ren W.; Liu S.; Ma L.; Xu Q.; Xu X.; Cao X.; Yang M H. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 2019, 28(9): 4364-4375.
28. Pan Z.; Yuan F.; Lei J.; Li W.; Ling N.; Kwong S. MIEGAN: Mobile image enhancement via a multi-module cascade neural network. *IEEE Transactions on Multimedia*, 2021, 24: 519-533.
29. Zhang Y.; Yang Q. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 34(12): 5586-5609.
30. Liu S.; Johns E.; Davison A J. End-to-end multi-task learning with attention. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 1871-1880.
31. Glorot X.; Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010: 249-256.
32. He K.; Zhang X.; Ren S.; Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
33. Deng J.; Dong W.; Socher R.; Li L J.; Li K.; Fei-Fei L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. 2009: 248-255.
34. Woo S.; Park J.; Lee J Y.; Kweon I S. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
35. Hu J.; Shen L.; Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
36. Ma K.; Duanmu Z.; Wu Q.; Wang Z.; Yong H.; Li H.; Zhang L. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 2016, 26(2): 1004-1016.
37. Kingma D P.; Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

38. Kim J.; Zeng H.; Ghadiyaram D.; Lee S.; Zhang L.; Bovik A C. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal processing magazine*, 2017, 34(6): 130-141.
39. Sheikh H R.; Sabir M F.; Bovik A C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 2006, 15(11): 3440-3451.
40. Larson E C.; Chandler D M. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 2010, 19(1): 011006-011006-21.
41. Ponomarenko N.; Jin L.; Ieremeiev O.; Lukin V.; Egiazarian K.; Astola J.; Kuo C C J. Image database TID2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 2015, 30: 57-77.
42. Lin H.; Hosu V.; Saupe D. KADID-10k: A large-scale artificially distorted IQA database. 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2019: 1-3.
43. Ghadiyaram D.; Bovik A C. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 2015, 25(1): 372-387.
44. Hosu V.; Lin H.; Sziranyi T.; Saupe D. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 2020, 29: 4041-4056.
45. Kim J.; Lee S. Fully deep blind image quality predictor. *IEEE Journal of selected topics in signal processing*, 2016, 11(1): 206-220.
46. Ma K.; Liu W.; Zhang K.; Duanmu Z.; Wang Z.; Zuo W. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 2017, 27(3): 1202-1213.
47. Yan Q.; Gong D.; Zhang Y. Two-stream convolutional networks for blind image quality assessment. *IEEE Transactions on Image Processing*, 2018, 28(5): 2200-2211.
48. Kim J.; Nguyen A D.; Lee S. Deep CNN-based blind image quality predictor. *IEEE transactions on neural networks and learning systems*, 2018, 30(1): 11-24.
49. Moorthy A K.; Bovik A C. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 2011, 20(12): 3350-3364.
50. Ye P.; Kumar J.; Kang L.; Doermann D. Unsupervised feature learning framework for no-reference image quality assessment. 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 1098-1105.
51. Xu J.; Ye P.; Li Q.; Du H.; Liu Y.; Doermann D. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 2016, 25(9): 4444-4457.