Journal of Big Data
DOI:10.32604/jbd.2020.015357
Article

Tech Science Press

Multi-Scale Blind Image Quality Predictor Based on Pyramidal Convolution

Feng Yuan and Xiao Shao*

School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China
*Corresponding Author: Xiao Shao. Email: shaoxiao@nuist.edu.cn
Received: 17 September 2020; Accepted: 04 December 2020

Abstract: Traditional image quality assessment methods use the hand-crafted features to predict the image quality score, which cannot perform well in many scenes. Since deep learning promotes the development of many computer vision tasks, many IQA methods start to utilize the deep convolutional neural networks (CNN) for IQA task. In this paper, a CNN-based multi-scale blind image quality predictor is proposed to extract more effectivity multi-scale distortion features through the pyramidal convolution, which consists of two tasks: A distortion recognition task and a quality regression task. For the first task, image distortion type is obtained by the fully connected layer. For the second task, the image quality score is predicted during the distortion recognition progress. Experimental results on three famous IQA datasets show that the proposed method has better performance than the previous traditional algorithms for quality prediction and distortion recognition.

Keywords: No-reference image quality assessment (NR-IQA); convolutional neural network; deep learning; feature extraction; image distortion recognition.

1 Introduction

Since the rapid development of digital technology, digital information has become ubiquitous in people's lives, such as electronic photo album, video stream and video websites. As the performance of the display device improves, people are more concerned about the quality of the images. However, digital images may occur a quality degrade during the process of image acquisition, storage and transmission, which lead to the deterioration of image quality. In order to predict the quality score of the received images, image quality assessment (IQA) has become more and more critical in the field of low-level computer vision task.

According to the final receiver of the image, IQA method can be categorized into subjective IQA method and objective IQA method. For subjective IQA method, the image score is obtained by the human observer. Though it can get reliable and accurate scores, however, collecting the mean opinion score (MOS) or differential mean opinion score (DMOS) for each image is laborious and consuming. Hence, it is crucial to design a computer algorithm to automatic predict the image score.

Generally, Objective IQA method can be divided into full-reference (FR) IQA, reduce-reference (RR) and no-reference (NR) IQA based on the availability of the reference image information. For FR-IQA methods, it uses the reference information and distorted image to obtain the final image score, such as PSNR, SSIM [1], MS-SSIM [2], FSIM [3] and VIF [4]. For RR-IQA methods [5–6], only partial reference information can be used in image quality prediction. Although the quality prediction of full-reference image quality assessment has been greatly improved in recent years, however, reference information is not available in many realistic scenes, such as in the wild scene. Hence, in order to make the IQA algorithm available in many real-world scenes, it is crucial to predict the image quality without the reference information. The goal of the NR-IQA method is to predict the image quality without the information of the

pristine image. Compare to the FR- and RR-IQA methods, NR-IQA (which is also called BIQA) is more applicable in many scenes since it is unnecessary to provide the reference information, which also makes it a challenging work to predict the quality score precisely.

Commonly, traditional objective NR-IQA methods use hand-crafted features [7–9] extracted from the distorted images to conduct the quality prediction. Many methods first extract the Natural Scene Statistics (NSS) based features, and then map these extracted features to the quality score through the support vector machines (SVR). Designing these hand-craft features require many people's efforts, but the effect of enhancement is not ideal. Since deep learning promotes the development of many computer vision task, e.g., object detection [10], image segmentation [11] and image enhancement [12]. Deep neural network applied to many IQA methods [13–14] boosts the performance of the quality prediction. Generally, convolutional neural networks (CNN) are used to extract the distortion features hide in the distorted image, then these features are feed into the fully connected (FC) layers to regression for the final quality score. Compare with the traditional IQA methods, deep learning-based IQA methods can extract more effective distortion features, and it can update the network parameters automatically through backpropagation without manual design features.

Although the convolutional neural network facilitates the extraction of effective features, however, the standard square convolutional layer has the weakness in handling the multi-scale features. To solve this problem, a multi-scale blind image quality predictor based on pyramidal convolution is proposed to focus on extracting the multi-scale distortion features for quality regression. Different from the standard square convolutional layer, pyramidal convolution [15] is capable of handling the image through multiple convolutional kernels. Hence, one standard square convolutional layer and three pyramidal convolutions are adopted to our network to learn complicated relationships between multi-scale distortion features and predicted quality score. For the human visual system (HVS), humans can easily judge the distortion types when they receive a distorted image. To mimic the HVS, a distortion recognition task is added to enhance the learning ability of the quality prediction. The distortion type recognition task is realized by the fully connected layer by mapping the feature maps to the n-node (n denotes the number of the distortion types) distortion types. The contributions of the proposed blind image quality predictor are summarized as follows:

(1) A multi-scale blind image quality predictor is proposed to mapping the relationships between the distortion features and the quality score, and it is realized by end-to-end training without the need to designed the hand-crafted features.

(2) To mimic the behavior of the HVS, the distortion recognition task is proposed to assist the quality prediction task. To enhance the ability of the feature extraction, pyramidal convolution is adopted to our network to achieve the multi-scale feature extraction ability.

(3) Experiments conducted on three famous IQA datasets have proved the effectiveness of our proposed method.

2 Related Works

For the FR-IQA method, it needs to obtain the full reference images, and the quality score is obtained by comprehensively comparing the distorted image and the corresponding distortion-free image. Compare with the RR- and NR-IQA methods, FR-IQA methods is relatively mature. The simplest way of the FR-IQA method is MSE (mean squared errors), it is realized by calculating the average variance of the pixel points of the distorted image and the reference image. PSNR (peak signal-to-noise ratio) is another corresponding way to calculate the difference between the distorted image and the corresponding distortion-free image. Although these two methods are simple to implement and widely used in the early stage, the prediction results are not consistent with the subjective IQA method. With the exploring of the human visual system (HVS), many novel methods are proposed. Wang et al. [1] proposed SSIM (structural similarity image metric) to mimic the HVS, which has been the most representative FR-IQA method. SSIM considers the brightness, contrast and structural information of the distorted image to extract the representative features, and achieves a great result on quality prediction. Then, many scholars made a series improvement

on the original SSIM. Wang et al. [16] proposed the MS-SSIM (multi-scale structural similarity image metric) to supply more multi-scale features than the original SSIM with introducing more view conditions. Chen et al. [17] proposed the GSSIM (gradient-based structural similarity), which considers the gradient information when extracting the features.

Before the machine learning and deep learning applied to the IQA domain, the dominant approach was to rely on the NSS features extract from the image, which is used to distinguish the distorted image and the pristine image. Generally, no-reference image quality assessment can be classified into hand-crafted-based methods and the learning-based methods. For hand-crafted-based methods, Wang et al. [18] proposed a quality method for handling the JPEG compression images. Saad et al. [19] proposed the BLIINDS-II to handling the distortion in DCT domain by extracting the contrast and structure features. Mittal et al. [20] proposed the NIQE by using the multivariate model to conduct the prediction task.

For learning-based methods, distortion features are extracted by the deep neural network instead of the elaborately designed features. Kang et al. [13] design the network with only one convolutional layer and two pooling layers to do the quality regression. To augment the training samples, images are cropped to 32×32 pixel patches to feed the network. Then they update the network by adding another task for distortion recognition [14]. Bosse et al. [21] use the deeper network with ten convolutional layers and max-pooling layers to extract the features, and the weighed strategy is proposed to calculate the influence of each patch for the final score. Kim et al. [22] enhance the training data by generating the error map in the first stage of training, then use the pre-trained model to do the quality regression in the second stage. Though these methods achieve a great result in handling the quality prediction, there still challenge remains. The distortion information in the image is multi-scale instead of single-scale. It is impossible to extract the multi-scale distortion features effectively through the single-scale convolutional layers.

3 Method Description

The overall architecture of the proposed blind image quality predictor is shown in Fig. 1. To effectively extract the distortion features of the distorted images, a multi-task blind image quality predictor is proposed to solve the NR-IQA problem. The proposed method contains two tasks: (1) Distortion recognition task and (2) Quality prediction task. Given a distorted image I_d, we crop patches from I_d to group $\{P_i^d, i = 1,2,\cdots,N\}$. Before entering the training progress, local normalization is used to preprocess the image patch P_i^d, then the local normalized patch $P_i^{d\prime}$ is feed into the network to train the distortion type and final quality score. The details of the proposed method are described as follows.

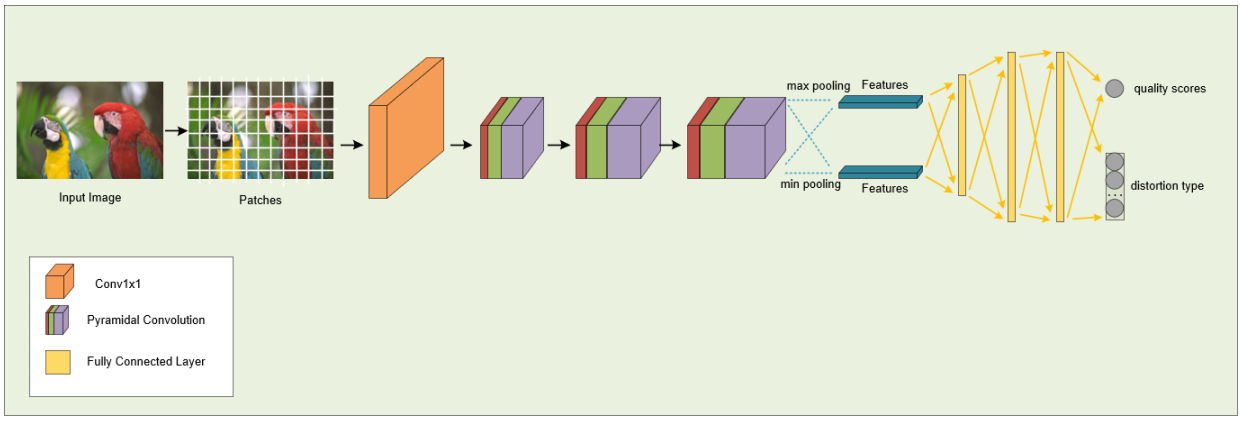

Figure 1: The overall architecture of the proposed network

3.1 Model Architecture

Motivated by [13], the proposed network uses convolutional neural network to extract distortion features. The network consists of four convolutional layers, which include one standard convolutional layer

and three pyramidal convolution layers, then three fully connected layers are used to map the feature maps to quality score and distortion types. The detailed architecture of the proposed network is shown in Tab. 1. The first convolutional layer with kernel size 1 is used to expend the channels of the feature maps to match the next pyramidal convolution layers. Then three pyramidal convolution layers with three kernel size 3, 5, 7 are used to extracted the multi-scale distortion features, and the output features are $128 \times 8 \times 8$. After that, max pooling and mini pooling layers are used to reduce the feature maps to $128 \times 1 \times 1$. Finally, three fully connected layers followed by PReLu [23] map the relationships between the extracted features and predicted distortion types and quality score. Motivated by [13], dropout is adopted after the first FC layers to avoid the overfitting problem, and the dropout probability is set to 0.5 in the network.

Table 1: Network details of the proposed method

	Input	Kernel	Output
Conv1	$3 \times 64 \times 64$	k = 1×1, s=1	$16 \times 64 \times 64$
Pyconv1	$16 \times 64 \times 64$	$\begin{bmatrix} k = 7 \times 7, c = 16, G = 8 \\ k = 5 \times 5, c = 8, G = 4 \\ k = 3 \times 3, c = 8, G = 1 \end{bmatrix}$	$32 \times 32 \times 32$
Pyconv2	$32 \times 32 \times 32$	$\begin{bmatrix} k = 7 \times 7, c = 32, G = 8 \\ k = 5 \times 5, c = 16, G = 4 \\ k = 3 \times 3, c = 16, G = 1 \end{bmatrix}$	$64 \times 16 \times 16$
Pyconv3	$64 \times 16 \times 16$	$\begin{bmatrix} k = 7 \times 7, c = 64, G = 8 \\ k = 5 \times 5, c = 32, G = 4 \\ k = 3 \times 3, c = 32, G = 1 \end{bmatrix}$	$128 \times 8 \times 8$
Max pooling	$128 \times 8 \times 8$	/	$128 \times 1 \times 1$
Mini pooling	$128 \times 8 \times 8$	/	$128 \times 1 \times 1$
FC1	256-d	/	800-d
FC2	800-d	/	800-d
FC3	800-d	/	1-d / n-d

3.2 Image Preprocess

For the human visual system (HVS), the HVS is insensitive to the changes in the low-frequency band. And for image distortion progress, the distortion only affects the high-frequency information of the image but has little impact on the low-frequency information. Hence, to mimic the human visual system and make the training progress more stable, input image patches need to be preprocessed before entering the training progress. In this step, local normalization is used to preprocess the input image as following [13]. Given an image patch P_i^d, the intensity value of a (i, j) pixel is denoting as $P_i^d(i, j)$, where *i* and *j* denotes the width and height location of the image patch. The local normalization progress is summarized as follows:

$$\widehat{P_i^d}(i,j) = \frac{P_i^d(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{1}$$

$$\mu(i,j) = \sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} I(i+m, j+n) \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} \big(I(i+m, j+n) - \mu(i,j)\big)^2} \tag{3}$$

where *C* denotes a small positive constant to prevent dividing by zero, *M* and *N* indicates the size of the normalized window. As suggest in [13], we set $M = N = 3$ to achieve the best performance.

3.3 Pyramidal Convolution

As shown in Fig. 2, distortion information hides in the image vary a wide range, from shallow to deep. Hence, the standard convolutional layer used in [13–14] cannot handle the multi-scale distortion features well. To solve this problem, pyramidal convolution [15] is adopted in the proposed network to extract more effective distortion features. As described in [15], pyramidal convolution contains a pyramid of kernels, different size of the kernel size varying different depth has the ability to extract the different levels of distortion information hide in the image.

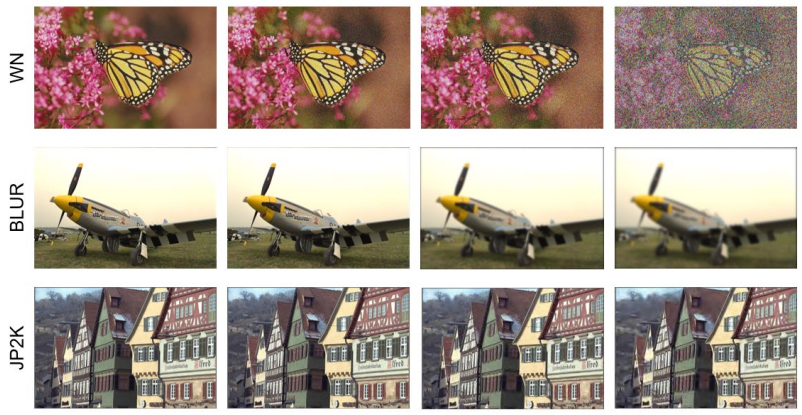

Figure 2: Image samples from the LIVE dataset. From left to right, the image quality deteriorates

3.4 Loss Functions

During the backpropagation progress, a well-designed loss function can not only accelerate the network converging but also improve the accuracy of the quality prediction. To achieve the best training effect, we design the mixed loss function with two different loss functions: L_i and L_d. Compare with the large dataset designed for object detection, the IQA datasets are too small for training the deep learning-based IQA method. To solve this problem, the input image is divided into 64×64 pixel patches for augmenting the training dataset. For each cropped image patch P_i^d, the corresponding score is obtained by the original image I_d. During the test progress, the quality score of the distorted images is calculated by averaging all patches cropped from the original image:

$$q = \frac{1}{N}\sum_{i=1}^{N} f\left(P_i^d; \theta\right) \tag{4}$$

where P_i^d denotes the i-th patch of distorted image, N denotes the number of the image patches cropped from the image, $f(\cdot)$ indicates the network of the proposed method, and θ denotes the network parameters of the network.

During the network training, the goal of the network is to narrow the gap between the predict scores and the ground truth score. The loss function is used to evaluate the predict image quality score $f\left(P_i^d; \theta\right)$ and ground truth score q_i. For the image quality prediction task, we adopt the commonly used objective function as:

$$L_i = \frac{1}{M}\left\| f\left(P_i^d; \theta\right) - q_i \right\|_1 \tag{5}$$

where M denotes the number of images, $f(\cdot)$ indicates the network of our proposed method and q_i denotes the i-th ground truth score.

For the distortion recognition task, cross entropy loss is adopted as the loss function and it can be described as:

$$L_d = -\log \frac{e^{x_i}}{\sum_{j=1}^{D} e^{x_j}} \tag{6}$$

where i, j indicates the i-th and j-th distortion types, respectively. x denotes the input vector. D denotes the number of the distortion types, e.g., for LIVE dataset, $D = 5$, for CSIQ dataset, $D = 6$.

In the end, to achieve the best performance on quality prediction and distortion recognition, the mixed loss function is defined as:

$$L = \alpha * L_i + \beta * L_d \quad (7)$$

where α and β denote the weight factor of L_i and L_d, respectively. In order to balance the training progress and keep the loss function on the same order of magnitude, we set $\alpha = \beta = 1$.

3.5 Training of the Network

All the training patches are cropped from the distorted image with the size of 64×64 pixel, and the step is set to 64 pixels. Our method is implemented using the Pytorch [24] on NVIDIA GTX 1080. We use Adam [25] with a learning rate of 10^{-4} to train our network. For every 100 epochs, the learning rate is decreased by 0.1. In addition, the momentum factor, weight decay factor and batch size are set to 0.9, 10^{-4} and 128 respectively.

4 Results

4.1 Experimental Setup

4.1.1 Datasets

In order to test the performance of quality prediction and distortion recognition, three famous synthetically IQA databases: LIVE [26], TID2013 [27] and CSIQ [28] are chosen to conduct the experiments. The details of the three databases (e.g., the number of reference images and distorted images, the number of distortion types) are tabulated in Tab. 2.

LIVE [26] database contains 779 distorted images which are generated from 29 different pristine images under the laboratory environment. The distorted images are under five different distortion types (such as, JP2K, JPEG, WN, GBLUR and FF) at 7 to 8 degradation levels. In addition, it provides Differential Mean Opinion Scores (DMOS) for each distorted image, and the range of the DMOS is from 0 to 100. The higher DMOS denotes the image has the worse quality.

TID2013 [27] database contains 3000 distorted images which are generated from 25 pristine images. For the distortion types and levels, it contains 24 different distortion types, and the degradation level is five, which makes it the most abundant synthetically IQA database according to the distortion types. Different from LIVE database, the Mean opinion Scores (MOS) is provided for each distorted image, and the value is from 0 to 9. The lower MOS denotes the lower image quality.

Table 2: Detailed information of three IQA databases

Databases	Ref. Images	Dist. Images	Dist. Types	Label	Range
LIVE [26]	29	779	5	DMOS	[1, 100]
TID2013 [27]	25	3000	24	MOS	[0, 9]
CSIQ [28]	30	866	6	DMOS	[0, 1]

CSIQ [28] database contains 866 distorted images generated from 30 pristine images. Each reference image contains six distortion types at 4 to 5 degradation levels: JPEG, JPEG2000, Gaussian blurring, Gaussian pink noise, Gaussian white noise and contrast change. Same as the LIVE database, the DMOS is provided for each distorted image, and the value is from 0 to 1. The higher value means the image has the bad visual quality.

4.1.2 Performance Criteria

For conduct the experiment, we choose two widely used metrics to evaluate each IQA algorithm: Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC). PLCC measures the linear correlation between the labeled quality scores and the network predicted quality, and it is formulated as:

$$PLCC = \frac{\sum_{i=1}^{N}(q_i-\bar{q})(\hat{q}_i-\bar{\hat{q}})}{\sqrt{\sum_{i}^{N}(q_i-\bar{q})}\sqrt{\sum_{i}^{N}(\hat{q}_i-\bar{\hat{q}})^2}} \tag{8}$$

where q_i denotes the labeled quality score of i-th image, and $\hat{q}_i$ denotes the predicted quality score of i-th image. $\bar{q}$ denotes the mean of the ground truth image quality scores, and $\bar{\hat{q}}$ indicates the mean of the predicted quality scores.

SROCC measures the prediction monotonicity and is defined as:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N}(r_i-p_i)^2}{M(M^2-1)} \tag{9}$$

where M denotes the number of the images, r_i indicates the rank of the ground truth score q_i in ground truth scores, and p_i denotes the rank of the predicted score $\hat{q}_i$ in predicted scores.

4.2 Experimental Results on Single Dataset

To verify the consistency between the model prediction results and human subjective evaluation, we conduct the single dataset evaluation on three synthetically IQA databases: LIVE [26], TID2013 [27] and CSIQ [28]. In the experiment, each database is randomly divided into two groups, 80 per cent of the reference images and the corresponding distorted images are selected to group for training the IQA algorithms, and the rest of them are used to group the testing set. The selection process is completely random. This procedure is repeated ten times to erase the bias caused by the database, the median results of SROCC and PLCC are chosen as the final results.

Table 3: SROCC and PLCC results on LIVE, TID2013 and CSIQ

Metrics	LIVE		TID2013		CSIQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
BLIINDS-II	0.912	0.916	0.536	0.628	0.780	0.832
DIIVINE	0.925	0.923	0.549	0.654	0.835	0.855
IL-NIQE	0.902	0.908	0.521	0.648	0.821	0.865
CORNIA	0.942	0.935	0.549	0.613	0.714	0.781
CNN++	0.953	0.953	/	/	/	/
Ours	**0.962**	**0.963**	**0.732**	**0.761**	**0.843**	**0.852**

For better evaluate the quality prediction and distortion recognition performance of the proposed method, several standard IQA methods are chosen to conduct the experiments, including BLIINDS-II [19], DIIVINE [29], IL-NIQE [30], CORNIA [31], and two deep learning-based method CNN [13] and CNN++ [14]. SROCC and PLCC results on three datasets are shown in Tab. 3, the best results are marked with bold face. From Tab. 3, our proposed method achieves the best results for all three IQA databases, and it reaches (0.962, 0.963), (0.732, 0.761), (0.843, 0.852) respectively. Compare with the deep learning-based method CNN and CNN++, our method even achieves better results on the LIVE dataset. Some crucial conclusions can be drawn from the experimental results that: (1) The pyramidal convolution layer introduced to our network can effectively extract the multi-scale distortion features than the standard convolutional layer,

which leads to the improvement in prediction results. (2) The multi-task training progress can better simulate the human visual system by predicting the image quality score while predicting the distortion types, which promotes the prediction accuracy.

Due to the distortion recognition ability is important for the accuracy of the quality prediction, we conduct the experiments to test the accuracy of the proposed method in distortion recognition. For the experiment, we compare our method with BLIINDS-II [19], BRISQUE [32], CORNIA [31], CNN++ [14]. The classification accuracy is tabulated in Tab. 4. It can be observed that the proposed method achieved the highest accuracy of 96.2%, which indicated that our method could have the ability to recognize the distortion type.

Table 4: Comparison of distortion recognition accuracy of multi-task IQA

Methods	Accuracy
BLIINDS-II	83.8%
BRISQUE	88.6%
CORNIA	87.5%
CNN++	95.1%
Ours	96.2%

4.3 Experimental Results on Different Distortion Types

A good IQA algorithm should be able to predict not only general distortion types but also for the individual distortion types. In this section, to verify the prediction ability for IQA methods on different distortion types, experiments are conduct on different types of LIVE database.

Table 5: SROCC results of different distortion types on LIVE database

Method	JP2K	JPEG	WN	BLUR	FF
BLIINDS-II	0.929	0.942	0.969	0.923	0.889
DIIVINE	0.913	0.910	**0.984**	0.921	0.863
CORNIA	0.943	0.955	0.976	0.969	0.906
CNN	0.952	**0.977**	0.978	0.962	0.908
Ours	**0.953**	0.962	0.973	**0.972**	**0.911**

Table 6: PLCC results of different distortion types on LIVE database

Method	JP2K	JPEG	WN	BLUR	FF
BLIINDS-II	0.935	0.968	0.980	0.938	0.896
DIIVINE	0.922	0.921	**0.988**	0.923	0.888
CORNIA	0.951	0.965	0.987	0.968	0.917
CNN	0.953	**0.981**	0.984	0.953	0.933
Ours	**0.962**	0.973	0.977	**0.979**	**0.919**

In this individual distortion experiment, all the model is train and test the model on each distortion types. We choose four NR-IQA methods: BLIINDS-II [19], DIIVINE [29], CORNIA [31], CNN [13] to compare with our method. The SROCC and PLCC values are shown in Tab. 5 and Tab. 6. From Tab. 5 and Tab. 6, the proposed method achieves the highest prediction accuracy for JP2K, BLUR and FF distortion types, and the results are (0.953, 0.962), (0.972, 0.979) and (0.911, 0.919), respectively. In summary, compare with CNN and CNN++, the method can handle different distortion type well.

5 Conclusion

In this paper, a multi-scale blind image quality predictor based on pyramidal convolution is proposed to solve the problem for NR-IQA, which includes two tasks: A quality prediction task and a distortion recognition task. With the introducing of the distortion recognition task, the accuracy of the quality prediction can be further improved. In addition, to enhance the network learning ability, pyramidal convolution is adopted to the backbone feature extractor of the proposed method to extract the multi-scale features. Extensive experiments on three famous IQA databases: LIVE, TID2013 and CSIQ demonstrate the effectiveness of the proposed method for quality prediction and distortion recognition.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

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

[2] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc IEEE Asilomar Conference on Signals*, vol. 2, pp. 1398–1402, 2003.

[3] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[4] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing,* vol. 15, no. 2, pp. 430–444, 2006.

[5] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. of SPIE-The International Society for Optical Engineering*, vol. 5666, pp. 149–159, 2005.

[6] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[7] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[8] M. A. Saad, A. C. Bovik and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[9] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 779–788, 2016.

[11] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Int. Con. on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2015.

[12] Lore, Kin Gwn, Adedotun Akintayo and Soumik Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition,* vol. 61, pp. 650–662, 2017.

[13] L. Kang, P. Ye, Y. Li and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1733–1740, 2014.

[14] L. Kang, P. Ye, Y. Li and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in 2015 *IEEE Int. Conf. on Image Processing (ICIP)*, Quebec City, QC, pp. 2791–2795, 2015.

[15] I. C. Duta, L. Liu, F. Zhu and L. Shao, "Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition," in *2020 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[16] Z. Wang, P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. of the 37th Asilomar Conf. on Signals, Systems and Computers*, pp. 1398–1402, 2003.

[17] G. H. Chen, C. L Yang and S. L. Xie, "Gradient-based structural similarity for image quality assessment," in *IEEE 2006 Int. Conf. on Image Processing*, 2006.

[18] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *IEEE Int. Conf. on Image Processing,* vol. 1, pp. I-477–I-480, 2002.

[19] M. A. Saad, A. C. Bovik and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[20] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[21] S. Bosse, D. Maniry, K. R. Mller, T. Wiegand and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[22] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2017.

[23] K. He, X. Zhang, S. Ren and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *2015 IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, pp. 1026–1034, 2015.

[24] Paszke A, Gross S and Massa F, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, pp. 8026–8037, 2019.

[25] D. Kingma, P. Diederik and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[26] H. R. Sheikh, M. F. Sabir and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[27] N. Ponomarenko, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing Image Communication*, vol. 30, pp. 57–77, 2015.

[28] E. C. Larson and D. M. Chandler, "Most apparent distortion: Fullreference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006, 2010.

[29] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[30] L. Zhang, L. Zhang and A. C. Bovik. "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 24, no. 8, pp. 2579-2591, 2015.

[31] P. Ye and D. Doermann. "No-reference image quality assessment using visual codebooks," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129–3138, 2012.

[32] A. Mittal, A. Moorthy and A. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.