Signal Processing: Image Communication 94 (2021) 116181

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

An optimized CNN-based quality assessment model for screen content image[☆]

Xuhao Jiang [b], Liquan Shen [a,b,*], Guorui Feng [b], Liangwei Yu [b], Ping An [b]

[a] *Key laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, China*

[b] *Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, China*

ARTICLE INFO

Keywords:
Image quality assessment
Screen content image
No-reference
Full-reference
Convolutional neural network
Quality pooling
Data selection

ABSTRACT

Most existing convolutional neural network (CNN) based models designed for natural image quality assessment (IQA) employ image patches as training samples for data augmentation, and obtain final quality score by averaging all predicted scores of image patches. This brings two problems when applying these methods for screen content image (SCI) quality assessment. Firstly, SCI contains more complex content compared to natural image. As a result, qualities of SCI patches are different, and the subjective differential mean opinion score (DMOS) is not appropriate as qualities of all image patches. Secondly, the average score of image patches does not represent the quality of entire SCI since the human visual system (HVS) is sensitive to image patches containing texture and edge information. In this paper, we propose a novel quadratic optimized model based on the deep convolutional neural network (QODCNN) for full-reference (FR) and no-reference (NR) SCI quality assessment to overcome these two problems. The contribution of our algorithm can be concluded as follows: 1) Considering the characteristics of SCIs, a valid network architecture is designed for both NR and FR visual quality evaluation of SCIs, which makes the networks learn the feature differences for FR-IQA; 2) with the consideration of correlation between local quality and DMOS, a training data selection method is proposed to fine-tune the pre-trained model with valid SCI patches; 3) an adaptive pooling approach is employed to fuse patch quality to obtain image quality, owns strong noise robust and effects on both FR and NR IQA. Experimental results verify that our model outperforms both current no-reference and full-reference image quality assessment methods on the benchmark screen content image quality assessment database (SIQAD). Cross-database evaluation shows high generalization ability and high effectiveness of our model.

1. Introduction

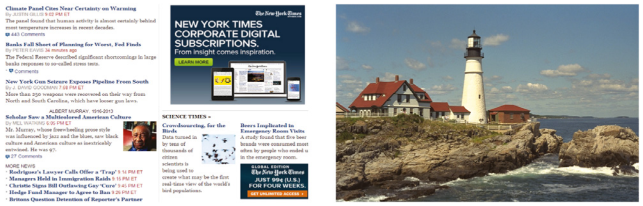
Nowadays screen content pictures have become quite common in our daily life with the rapid development of multimedia and social network. Numerous consumer applications, such as Facebook, Twitter, remote control and more, involve computer-generalize screen content images (SCIs). Fig. 1(a)–(b) shows two typical images, one is a natural image and the other is a SCI. There are significant differences between these two images. Natural images have rich color and slow color change, while SCIs contain more thin lines, sharp edges and little color variance for massive existence of texts and computer-generated graphics. During acquisition, processing, compression, storage, transmission and reproduction, digital images may introduce various types of distortions, and the visual quality of the images is degraded as a result. Image quality assessment (IQA) aims to objectively evaluate image quality, in order to solve the problem that the human spend much time on judging the image subjective quality. IQA methods can also be used to optimize image processing algorithms. Therefore, IQA plays a very important role in image processing community.

There are numbers of IQA methods for natural images designed in recent years including full reference (FR), reduced reference (RR) and no reference (NR). Considering the characteristics of the HVS, many FR approaches develop and become highly consistent with subjective quality scores, including structural similarity (SSIM) [1], multi-scale SSIM (MS-SSIM) [2], information weighted SSIM (IW-SSIM) [3], feature similarity (FSIM) [4], visual saliency-based index (VSI) [5], gradient magnitude similarity deviation (GMSD) [6], and gradient SSIM (GSIM) [7]. For RR methods, such as [8–10], only partial information of reference images is used for IQA. For NR methods, only the distorted

[☆] This work was supported in part by the National Natural Science Foundation of China under Grant 61931022 and Grant 61671282; in part by the Open Fund of Key Laboratory of Advanced Display and System Applications of Ministry of Education (Shanghai University); in part by the Shanghai Science and Technology Innovation Plan under Grant 18010500200; and in part by the Shanghai Shuguang Program under Grant 17SG37.

[*] Corresponding author at: Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, China.
E-mail address: jsslq@163.com (L. Shen).

https://doi.org/10.1016/j.image.2021.116181
Received 24 March 2019; Received in revised form 8 August 2020; Accepted 24 January 2021
Available online 6 February 2021
0923-5965/© 2021 Elsevier B.V. All rights reserved.



Climate Panel Cites Near Certainty on Warming
By JUSTIN GILLIS 9:02 PM ET
The panel found that human activity is almost certainly behind most temperature increases in recent decades.
443 Comments

Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in large banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from North and South Carolina, which have looser gun laws.

ALBERT MURRAY, 1916-2013
Scholar Saw a Multicolored American Culture
By MEL WATKINS 6:05 PM ET
Mr. Murray, whose freewheeling prose style was influenced by jazz and the blues, saw black culture and American culture as inextricably entwined. He was 97.
27 Comments

MORE NEWS
- Rodriguez's Lawyer Calls Offer a 'Trap' 9:14 PM ET
- Managers Held in Immigration Raids 9:15 PM ET
- Christie Signs Bill Outlawing Gay 'Cure' 9:45 PM ET
- Hedge Fund Manager to Agree to Ban 9:26 PM ET
- Britons Question Detention of Reporter's Partner

The New York Times
NEW YORK TIMES CORPORATE DIGITAL SUBSCRIPTIONS.
From insight comes inspiration.
LEARN MORE

SCIENCE TIMES »

Crowdsourcing, for the Birds
Data turned in by tens of thousands of citizen scientists is being used to create what may be the first real-time view of the world's bird populations.

Beers Implicated in Emergency Room Visits
A study found that five beer brands were consumed most often by people who ended u in the emergency room.

GLOBAL EDITION
The New York Times
JUST 99¢ (U.S.) FOR FOUR WEEKS.
GET UNLIMITED ACCESS ›

(a)

(b)

Fig. 1. (a) is an example of screen content image and (b) is an example of natural image.

images are employed for IQA. Generally, NR IQA algorithms extract specific features from distorted images and train a regression model with these features and subjective rating by machine learning, such as [11–16].

Most existing IQA approaches devised for natural images are not effective for SCI quality evaluation without taking into account the difference in image content and characteristics between SCIs and natural images. Recognizing this difficulty, others have developed a variety of IQA approaches tailored to SCIs. Some representative works [17–22] of FR IQA have been published and achieve good results. Yang et al. propose a FR-IQA model based on SCIs segmentation [17]. It provides an effective text segmentation method to distinguish the pictorial and textual regions, and an activity weighting strategy is employed to fuse the visual quality scores of textual and pictorial regions to the overall quality scores. Based on text segmentation method [17], structural features based on gradient information and luminance features are extracted for similarity computation to obtain the visual quality of SCI [19]. In [20], Gu et al. propose a FR metric which mainly relies on simple convolution operators to detect salient areas. Ni et al. [21] design a FR metric based on local similarities extracted with Gabor filters in the LMN color space. These FR-IQA methods above achieve a superior performance of SCIs quality evaluation. Numerous NR IQA methods for SCIs are proposed [23,24]. Shao et al. [23] propose a blind quality predictor for SCI to explore the issue from the perspective of sparse representation. Local sparse representation and global sparse representation are conducted for textual and pictorial regions, respectively. Then, the local and global quality scores are estimated and combined to a total one. Another effective approach of no-reference SCI quality assessment is presented in [24], which obtains an overall quality score through extracting features from the histograms of texture and luminance and training these features based on SVR.

With the development of the CNN, many models [25–32] have started to build neural networks to process the problem for natural image quality assessment, and have achieved superior performance. These methods utilize image patches as a data augmentation, and design special patch-level neural networks for natural IQA. Kang et al. [25] propose a method based on CNN to accurately predict natural image quality without reference images. Bosse's method [32] promotes the CNN to learn the local scores and local weights of image patches, and then fuse the local scores to obtain a global quality score with the local weights. This work mainly considers the relative importance of local score to the global quality estimation. However, these patch-level CNN-based methods still do not consider the special characteristics of the screen content images where textual regions attract more attention than pictorial regions. Zhang et al. [33] propose a FR-IQA model for SCI taking fusion of textual and pictorial regions into consideration, where the IQAs of pictorial and textual regions are evaluated separately and fused with a region-size-adaptive quality-fusion strategy. Zuo et al. propose a NR method using classification models for SCI quality assessment in [34]. A novel classification network is designed to train the distorted images for getting a practical model, and weights of texture regions and pictorial regions are determined according to the gradient entropy adapt to the characteristics of the screen content image. Chen et al. [35] propose a naturalization module to transform IQA of SCIs into IQA of natural images. These patch-level CNN-based IQA methods for SCIs have their limitations. They divide SCIs into image patches aiming to obtain enough training data, and utilize DMOSs as ground truth. This brings two problems: lacking reliable ground truth of image patches and effective strategy of fusing local score. For SCIs, qualities of SCI patches are different, and the DMOS is not appropriate as qualities of all image patches. In addition, the HVS is sensitive to image patches containing texture and edge information. Thus, training CNN utilizing some valid image patches is reasonable, and image patches containing a large amount of texture information should be assigned bigger weights compared to patches containing simple graphics. The proposed model aims to solve these two problems employing two optimization including training with valid image patches and weighting based on the variance of local standard deviation (VLSD) in Sections 3.2 and 3.3.

In this work, we propose a novel algorithm for both NR and FR IQA of screen content images to solve limitations of the previous methods for screen content images based on patch-level CNN-based models. Unlike traditional patch-level CNN-based methods, our model selects part of all image patches as effective input data whose quality is relatively close to DMOS. For this purpose, the network is pre-trained with all of the training image patches to predict the quality scores of these patches, and employs the Euclidean distance between predicted scores and DMOS to pick valid image patches. The pre-trained model is fine-tuned with selected image patches to obtain a more accurate model. On this basis, an efficient and adaptive weighting method is designed to fuse the image patch quality with considering the effect of the different image patch content. The main contributions of our method are described as follows:

(1) Considering the characteristics of SCIs, a valid network architecture is designed for both NR and FR quality evaluation of SCIs. Moreover, reference information is extracted by independent layers, which is concatenated with distorted information in the shallow layer for FR-IQA. This confirms that networks learn the feature differences.

(2) Considering the connection between the histogram distribution of local scores predicted by pre-trained model and DMOSs, the Euclidean distance between local scores and DMOSs is utilized to evaluate the effectiveness of training image patches. A training data selection based on effectiveness is proposed to fine-tune pre-trained model for obtaining a higher-performance model.

(3) The noise robust index VLSD is utilized to evaluate weights of SCI patches. Our proposed adaptive weighting method based on VLSD is appropriate to fuse local quality score under different types and degrees of distortions.

The rest of this paper is organized as follows. Section 2 provides a brief review of the related work. In Section 3, an effective CNN-based IQA algorithm for SCI is proposed. Section 4 shows experimental results and compares performance of the proposed algorithm with the state-of-the-art methods. Finally, conclusions are given in Section 5.

2. Related work

2.1. Training patch-level deep IQA model with more reliable ground-truth

The traditional CNN-based approaches [25,32] work with image patches by assigning the DMOS of an image to all patches within it. These approaches suffer from limitation that quality scores of image patches within a large image vary even when the distortion is homogeneously applied [1]. Therefore, some works [27,36] make use of FR-IQA methods for quality annotation. Kim and Lee [36] pre-train the model with the predicted local score of an FR-IQA approach as the ground-truth and fine-tune it with DMOSs. Inspired by this work, Bare et al. [27] devise an accurate deep model which utilizes the FSIM [4] to generate training labels of image patches and adopt a deep residual network [37] showing strong ability to extract features in classification and regression tasks. Compared with approaches using

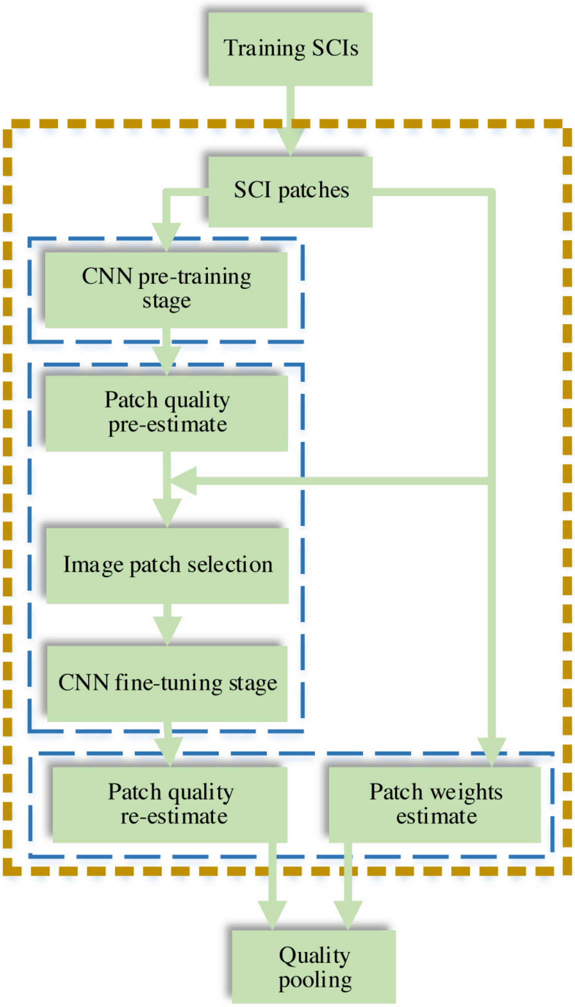


Fig. 2. Framework of the proposed algorithm.

DMOSs, models employing scores predicted by FR-IQA methods as ground-truth achieves better performance, because network is trained better while each image patch is labeled with a more accurate score. However, this brings a new problem that the accuracy of the FR-IQA models affects their performance.

Inspired by these works, we observe that the local scores within context of SCIs (e.g., the quality of a 32 × 32 patch within a large image) have great difference since SCIs contain complex content such as textual and pictorial regions. However, three high-performance FR-IQA methods for natural images and SCIs are utilized to predict the image patches quality and the performance with an average pooling strategy is poor in Section 3.2. The main reason is that statistical characteristics of SCI patches vary greatly for different characteristics of pictorial and textual regions existing in SCIs. Therefore, the method to train a deep model with the predicted local scores by FR models is not appropriate for SCIs.

2.2. Local quality score fusion for IQA

Most existing CNN-based models divide large image into image patches, and use image patches as input of CNN [25–36]. This leads to an problem that how to fuse local scores of image patches to obtain entire image quality. Some approaches use average pooling strategy to gain image quality, such as [27,31,35]. However, they do not consider the effect of image patch content on image quality. Bosse et al. [32] first propose a learning model to combine image patch content with NR-IQA model. In this work, it contains two sub-networks to separately learn local score and local weight of image patches. Then the image visual quality Q is evaluated by weighting the local score y_i of region i with the corresponding local weight w_i with

$$Q = \frac{\sum_i w_i y_i}{\sum_i w_i} \tag{1}$$

However, we observe that the method to learn local weight surely improves performance of IQA for natural images but has little effect on IQA for SCIs. For natural images, CNN-based model can precisely predict local score and show a high performance on IQA. For SCIs, CNN-based model does not achieve an expected performance of image patch score prediction. The biggest problem is mentioned in Section 2.1. Local weight of learning model has high correlation with local score, and thus the weight prediction will be poor without accurate quality prediction.

3. Proposed method

In this section, the proposed QODCNN for SCIs is described in details. As is shown in Fig. 2, QODCNN consists of three sub-steps accomplished by pre-training, fine-tuning and post-processing. Firstly, the designed CNN is trained with all the image patches to obtain an initial model of SCI visual quality assessment in the first stage, which learns the features of image distortion information and can effectively predict the quality of SCIs. Secondly, with the pre-trained CNN model, quality scores of all the training image patches are predicted, and then a data selection is applied according to the connection between predicted scores and DMOS. Fine-tuning the network aims to gain a more precise and valid model with selected data in the second stage. Third, considering the different importances of image patches containing different content for IQA of SCIs, the VLSD is designed to measure the local weights of image patches and fuse local scores. Finally, a learning model is obtained to effectively evaluate the visual quality of SCIs.

3.1. Network architecture

The proposed CNN architecture is applied to the pre-training and fine-tuning stages, and is shown in Fig. 3. The proposed NR-IQA architecture consists of six convolutional layers, three max pooling layers, and two fully connection layers. Compared to NR-IQA architecture, the FR-IQA architecture has two more convolutional layers, one more max pooling layer and one more concatenate layer. Each convolutional layer has a 3 × 3 filter with a stride of 1 pixel, and each pooling layer has a 2 × 2 pixel-sized kernel with a stride of 2 pixels. For each convolutional layer, zeros are padded around the border and a BN [38] layer is added to improve network training performance. The output feature map of the BN layer is calculated by Eq. (2),

$$y = alpha \times \frac{(Z_j - u)}{std} + beta \tag{2}$$

where Z_j is the input feature map of the jth batch normalized layer and y is the output. u and std respectively denote the mean and variance of the input map. $alpha$ and $beta$ are two parameters updated in training. The rectified linear unit (ReLU) [39] as activation function is added after the normalized layers. Feature maps extracted by convolutional layers and pooling layers are named, and the precise configurations are listed in Fig. 3.

For FR-IQA, this model extracts the feature maps of reference image patches and distorted image patches, and fuses these maps with a concatenate layer in shallow layer of network. These fused feature maps are regressed by the remaining network layers. The function for FR models is defined as follows,

$$\hat{q} = f_3(f_1(I_d) + f_2(I_r)) \tag{3}$$

where $\hat{q}$ is the predicted quality of the distorted image patch, I_d is the distorted image patch, I_r is the reference image patch, f_1 is the

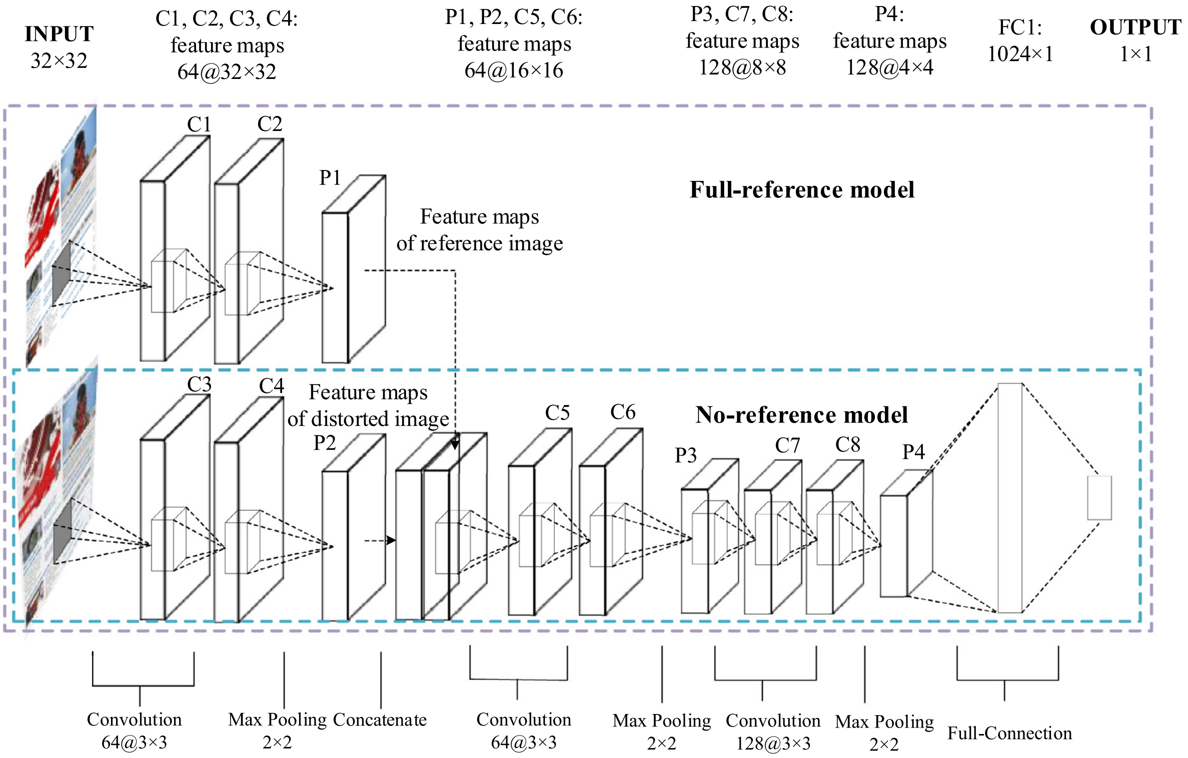


Fig. 3. An illustration of the architecture of our CNN model.

function of networks to extract features from distorted image, f_2 is the function of networks to extract features from reference image and f_3 is the function of fusion features and regression quality. For NR-IQA, the branch of extracting feature maps of reference images is abolished, and thus the adjusted model extracts features only from the distorted images. The function for NR models is defined as follows,

$$\hat{q} = f_4(f_1(I_d)) \tag{4}$$

where f_4 is the regression function to predicted quality of distorted image. The $l_1 - norm$ loss function for both NR and FR models is defined as follows,

$$l_1 = \frac{1}{N}\sum_{i=1}^{N} |\hat{q} - q| \tag{5}$$

where N is the image patch number of an input mini-batch , $\hat{q}$ is the network output of an input image patch, and q is the DMOS value of the corresponding image.

The CNN parameters are learned end-to-end by minimizing the sum of loss function and l_2 regularization for the predicted quality, on all training tuples:

$$min\{\frac{1}{N}\sum_{i=1}^{N} |\hat{q} - q| + \frac{\alpha}{2N}\sum w^2\} \tag{6}$$

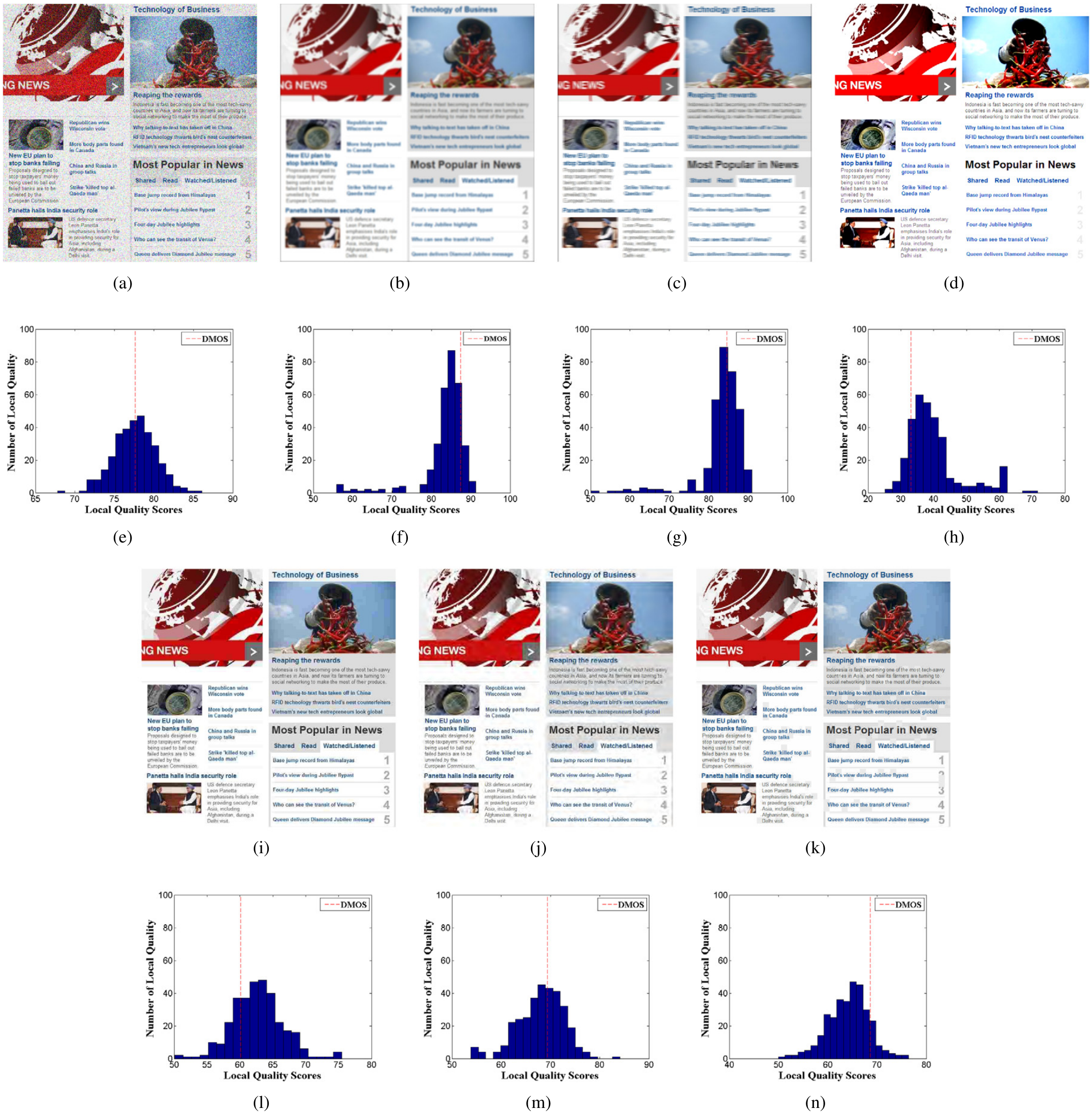
where α represents the penalty factor and w is the weight of CNN model.

In our model, combination of two 3 × 3 convolutional layers and one pooling layers is employed for owning a larger view to extract features with less data. Considering that SCIs contain lots of edge and gradient information, the max pooling layer extracts texture features which is applied to capture texture changes degraded by noise. For FR-IQA, using a difference map calculated by reference and distorted images as input of model is a normal idea, which reduces the amount of model calculations. However, this operation will lose a lot of information of distorted and reference images which is important for IQA. The difference map only considers the pixel difference between the reference and distorted images, refer to MSE algorithm for IQA. In the proposed model, reference information is extracted by independent layers and concatenated with distorted information in the shallow layer. This confirms that networks learn the differences between features extracted from distorted and reference SCIs rather than the differences between distorted and reference SCIs. Compared with features in deep layer, features in shallow layer retain a large amount of original information with less information loss. In addition, BN layers and l_2 regularization significantly improve the speed of the model's regression and the performance of fitting. Most existing models adopt $l_2 - norm$ loss for natural IQA. However, considering the big statistics differences between image patches of SCIs, $l_1 - norm$ loss is applied to reduce the impact of some abnormal image patches.

3.2. Training with valid image patches

Most existing patch-level CNN-based models all face a problem that local score of a large image is labeled with an inaccurate quality score. This problem is more serious when employing a patch-level CNN model to predict visual quality of SCIs. Compared with natural images, SCIs contain more complex content such as pictorial and textual regions. Some approaches [27,36] utilize FR-IQA methods to predict local scores of natural scene image. However, it has limited effect for SCI visual quality evaluation. Here, SSIM [1], FSIM [4] and SQMS [20] are employed to test the performance where these three FR-IQA methods predict local scores of SCI patches and fusing local scores with average pooling is applied to obtain image quality on SIQAD database [17]. From Table 1, it can be observed that these methods illustrate poor performance due to existing big difference between image patches of SCIs. Therefore, we consider solving this problem through training CNN with valid image patches.

In our model, training with valid image patch is considered to solve this problem including two training steps. In the first step, an initial model is obtained by training neural network of Section 3.1 with all the image patches for SCI quality assessment. To test effectiveness



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k)

(l) (m) (n)

Fig. 4. The samples of predicted local quality. The first and third rows show the distorted SCIs; the second and fourth rows show the correspond local quality histograms. (a–d, i–k) are distorted SCIs with the most serious noise of GN, GB, MB, CC, JPEG, JPEG2000, and LSC in SIQAD.

Table 1
Patch-level performance of three FR-IQA methods.

Index	PLCC	SRCC	RMSE
SSIM [1]	0.7630	0.7602	10.8837
FSIM [4]	0.8295	0.8328	9.2777
SQMS [20]	0.8104	0.8156	10.2790

of the pretrained model, this model predicts all the training image patches and the corresponding distribution maps of these predicted patch scores are provided in Fig. 4. As shown in Fig. 4, the distorted images of seven different distortion types and the corresponding local quality histograms are listed. For gaussian noise (GN), gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG compression (JC), JPEG2000 compression (J2C) and layer segmentation-backed coding (LSC) distortions, the most serious noise is used to analysis as typical examples. These histograms of local quality show that predicted local quality of training image patches is distributed around the DMOSs. This also verifies the problem mentioned in the introduction that the local score of each image patch varies in a large image. It can be noted the local scores of some image patches are far away from the DMOSs, which damages the performance of the learning model.

Inspired by histograms of local scores, a training data selection is more reasonable and benefits the deep model learning, since CNN can be learned better with training data labeled with precise ground-truth. Data selection abandons those image patches whose local scores deviates from ground-truth greatly and selects those local scores closes to DMOSs. Therefore, the Euclidean distance is employed to evaluate the effectiveness of training image patches and calculated as follows,

$$E = |\hat{q} - q| \tag{7}$$

where $\hat{q}$ is the predicted score of an image patch, q is the DMOS value of the corresponding image and E denotes the effectiveness index. In order to maintain enough information of each image, image patches are selected from each image with a fixed ratio. The ratio is computed by

$$P = \frac{N_s}{N} \quad (E \leq T) \tag{8}$$

where P presents the data selection ratio, N_s and N are the number of selected image patches and all image patches of a SCI, and T denotes an

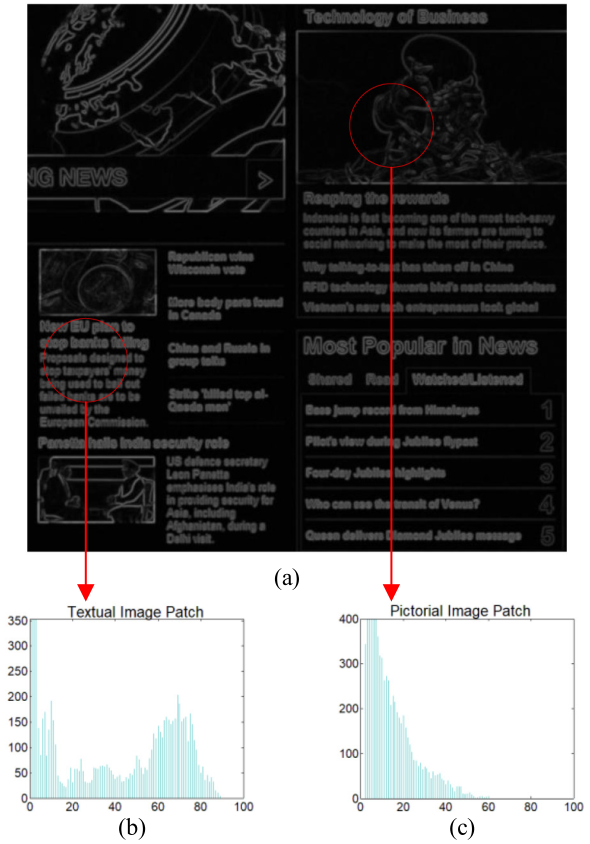


Fig. 5. (a) is the map with a smooth processing of a SCI distorted by Gaussian noise; (b) and (c) are of different pixel distributions of pictorial region and textual region in the smoothed image.

adaptive threshold. The pre-trained model is fine-tuned with selected training data to obtain a more effective and higher-performance model. The comparison experimental results are shown in Section 4.4.1, and verify the effectiveness of training with valid image patches.

3.3. A adaptive weighting method based on VLSD

The pooling-by-average approach is used to calculate global quality with local scores in most patch-level CNN-based models [25,27]. However, the average pooling on local score estimation does not consider the effect of spatially varying perceptual relevance of local quality. Especially for SCIs, HVS is sensitive to edge information, which means that image patches containing textual region owns higher weights than image patches containing pictorial regions for IQA. It is very difficult to distinguish two regions only from distorted images. Fang's FR-IQA model [19] obtains global quality by weighting local quality with gradient entropy. Using gradient entropy to distinguish textual and pictorial regions is useful for the FR-IQA model, but it is difficult for NR-IQA since entropy is sensitive to noise.

In our model, SCIs are divided into image patches to do data augmentation, and the contents of these image patches are quite different. The SCI patches are classified to two types including textual image patches and pictorial image patches since SCIs are composed of textual and pictorial regions. Considering the perceptual property of the HVS, an adaptive weighting method based on VLSD is proposed to measure local weights on local score fusion. The VLSD is utilized to evaluate weights of textual and pictorial image patches, and overcome the influence of distortions under different types and degrees. As shown in Fig. 5, Fig. 5(a) depicts map of a typical gaussian distorted SCI with a smooth processing of local standard deviation (LSD), and Fig. 5(b) and (c) depict two histograms of the textual region and pictorial region. LSD is applied on each image to indicate the structural complexity and reduce the impact of noise. The value of output feature map is calculated as,

$$LSD(i,j)=\sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i,j)-\mu(i,j))^2} \tag{9}$$

where $w = w_{(k,l)}|k=-K,\dots,K,l=-L,\dots,L$ is a 2D circularly-symmetric Gaussian weighting function, $I_{k,l}(i,j)$ is a pixel point, and $\mu(i,j)$ is the mean value of within a $K\times L$ local window centered at (i,j), $LSD(i,j)$ is the output pixel of the corresponding position. In our implementation, $K=L=3$. It can be observed that the noise of SCI is weakened after LSD processing, and plenty of thin lines are left on the image which is beneficial for extracting texture information in pictorial and textual regions in Fig. (a).

Secondly, the LSD distributions of typical textual and pictorial regions are respectively shown in Fig. 5(b) and (c). It can be seen that the histogram of the textual region (b) is relatively scattered compared to the pictorial region (c) whose histogram is relatively concentrated. Considering the differences of the histogram distributions, the variance of LSD is taken as the feature to describe the texture contents of image. The value of variance is calculated by,

$$VLSD=\frac{1}{N}\sum_{n=1}^{N}(LSD(i,j)-L\hat{S}D)^2 \tag{10}$$

where N is number of pixels in a image patch, $L\hat{S}D$ is the mean value of local deviation map, and $LSD(i,j)$ is the pixel point in the local deviation map which is calculated in Eq. (9). The reference SCI and the corresponding VLSD maps of three distortion types (GN, CC and JPEG) are shown in Fig. 6. To demonstrate the noise robustness of VLSD, the distorted SCIs with the slightest and most serious noise are as examples. As seen in Fig .6, two typical areas are marked with two different color boxes where yellow boxes represent pictorial regions and blue boxes represent textual regions. It can be observed that textual regions of SCI obtain bigger values compared with pictorial regions of SCI in all VLSD maps under different types and degrees of distortions. This shows VLSD can effectively distinguish pictorial and textual regions, and measure local weight of SCIs. Moreover, VLSD also owns strong noise robust ability. Compared with gradient entropy, the VLSD can better distinguish textual regions and pictorial regions, and is robust to distortion types and intensity. Thus the VLSD is employed to measure the importance of local regions in a large SCI.

Finally, scores predicted by the fine-tuned CNN and the corresponding VLSD of the image patches are obtained. A weighting method is applied to fuse quality of textual and pictorial image patches which is calculated as,

$$S=\frac{\sum_{n=1}^{N} s_n \times VLSD_n}{\sum_{n=1}^{N} VLSD_n} \tag{11}$$

where s_n and $VLSD_n$ are the score of the nth patch and its variance value calculated based on Eq. (10), N is the number of the patches of the test image, S is the final score of the test image. The proposed weighting method can measure the importance of different image patches, and surely improve the performance of the proposed model. The comparison experimental results are shown in Section Section 4.4.2, and verify the effectiveness of the weighting method based on VLSD.

3.4. Training

Before training our model, a grayscale processing and a data augmentation are applied by dividing large color images into gray image patches with size 32 × 32. The Tensorflow is used as the training toolbox, and two databases [17,18] are used to train and test our model. Both pre-training and fine-tuning steps adopt the Adam optimization algorithm [40] with a mini-batch of 64, and employ DMOSs as ground-truth of training. The penalty factor α of l_2 regularization is 1×10^{-5}. In the pre-training stage, the learning rate is changed from 1×10^{-4} to 1×10^{-13} at the interval of ten epochs. For fine-tuning, we fine-tune the pre-trained model with the same learning rate conditions. After training for two hundred epochs, the final model is obtained to predict visual quality.

4. Experimental results

4.1. Database and evaluation methodology

To verify the effectiveness of the proposed QODCNN, two screen content image databases including SIQAD [17] and SCID [18] are used to conduct the experiments. The SIQAD database has 20 reference screen content images and 980 distorted screen content images which contain seven types of distortion including GN, GB, MB, CC, JC, J2C and LSC, and each is with seven levels of distortions. The SCID database is used for cross-database experiments. Six distortion types (GN, GB, MB, CC, JC, and J2C) at five different levels are considered that are common in the two databases. This leaves us 1200 test SCIs in SCID.

In most cases, three typical performance evaluation criteria are adopted to evaluate the performance of IQA algorithms, the Pearson Linear Correlation Coefficient (PLCC), Spearman's Rank-order Correlation Coefficient (SRCC) and Root Mean Square Error (RMSE). The values of PLCC and SRCC are closer to 1, and the value of RMSE is smaller, indicating that the algorithm is more accurate. Given the ith image in the database (with N images in total), o_i and s_i are the objective and subjective scores, $\bar{o}$ and $\bar{s}$ are the mean values of o_i and s_i, e_i is the difference between the subjective and objective results. PLCC, SRCC and RMSE are defined as follows,

$$PLCC = \frac{\sum_{i=1}^{N}(o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{N}(o_i - \bar{o}) \times \sum_{i=1}^{N}(s_i - \bar{s})}} \tag{12}$$

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2-1)} \tag{13}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(o_n - s_n)^2}{N}} \tag{14}$$

A five-parameter mapping function [41] is employed to nonlinearly regress the quality scores into a common space as follows,

$$f(x) = \beta_1\left(\frac{1}{2} - \frac{1}{1+exp(\beta_2(x-\beta_3))}\right) + \beta_4 x + \beta_5 \tag{15}$$

where $(\beta_1, \ldots, \beta_5)$ are parameters to be compute with a curve fitting process.

4.2. Performance evaluation on SIQAD

For performance evaluation, the proposed QODCNN model is trained on database SIQAD. To distinguish the NR and FR models, the NR model is named as QODCNN-NR and the FR model is named as QODCNN-FR. For SIQAD, 980 distorted SCIs are randomly divided into two subsets according to the image content. Training set contains 784 distorted SCIs associated to 16 reference images (80% data for training) and the rest 196 distorted SCIs associated to 4 reference images are used as testing set (20% data for testing). The training–testing set partitions are randomly repeated 10 times, and the average performance of ten experiments is calculated as the overall performance.

In the second stage of our proposed QODCNN model, the proportion of data selection influences the performance of the fine-tuned model. Here, considering the difference of local quality within a large image, 10% to 80% of image patches of each image in training set of the second stage are selected by adjusting the threshold of Eq. (7) for training, and all of the image patches in testing set are utilized to evaluate the performance. Experiments are repeated for different training–testing sets of the first stage and different proportion of training data. For different proportions of training data, the average performance of ten results is calculated as the overall performance.

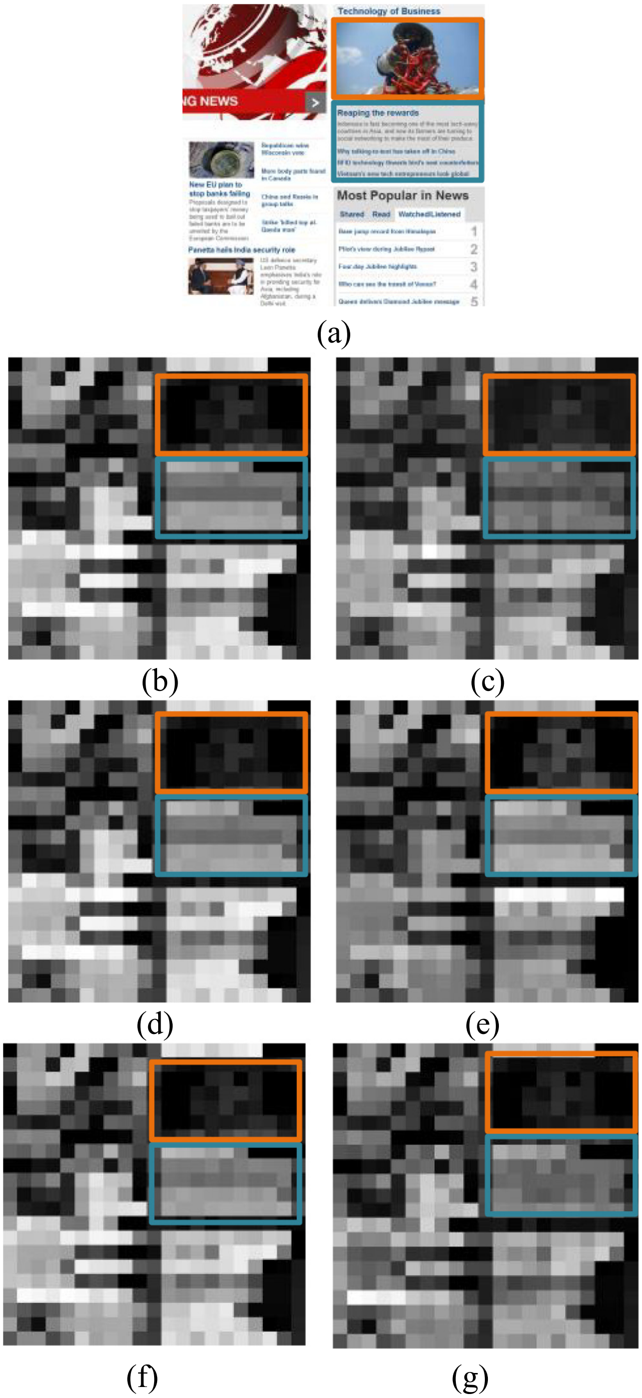


Fig. 6. The visual samples of VLSD maps. (a) is a reference SCI and (b–g) are VLSD maps of six distorted SCIs. (b, c), (d, e) and (f, g) are VLSD map pairs with the slightest and most serious noise of GN, CC and JPEG in SIQAD.

4.2.1. Full-reference image quality assessment

The results of QODCNN-FR models are shown in Table 2, compared with other state-of-the-art FR models: PSNR, SSIM [1], GMSD [6], SPQA [17], ESIM [18], SQMS [20],GFM [21], SFUW [19], MDOGS [22] and CNN-SQE [33]. Specially the last seven models are designed for SCIs. It can be seen from Table 2 that FR methods designed for SCIs achieve higher performance than FR metrics designed for natural images. The reason is that these SCI models (SPQA, ESIM, SQMS, GFM, SFUW, MDOGS and CNN-SQE) consider the correlation between HVS

Table 2
Experimental results of proposed and other existing FR and NR methods on SIQAD database.

	Method	PLCC	SRCC	RMSE
Full-Reference	PSNR	0.5869	0.5608	11.5859
	SSIM [1]	0.7561	0.7566	9.3676
	GMSD [6]	0.7259	0.7305	9.4684
	SPQA [17]	0.8584	0.8416	7.3421
	ESIM [18]	0.8788	0.8632	6.8310
	SQMS [20]	0.8872	0.8803	6.6039
	GFM [21]	0.8828	0.8735	6.7234
	SFUW [19]	0.8910	0.8800	6.4990
	MDOGS [22]	0.8839	0.8822	6.6951
	CNN-SQE [33]	**0.9040**	**0.8940**	**6.1150**
	QODCNN-FR	**0.9142**	**0.9066**	**5.8015**
No-Reference	BLINDS-II [11]	0.7255	0.6813	9.4991
	BRISQUE [12]	0.7708	0.7237	8.1342
	BLIQUP-SCI [23]	0.7705	0.7990	10.0213
	NRLT [24]	0.8442	0.8202	7.5957
	CNN-Kang [25]	0.8487	0.8091	7.4472
	WaDIQaM-NR [32]	0.8594	0.8522	7.0570
	PICNN [35]	**0.896**	**0.897**	**6.790**
	QODCNN-NR	**0.9008**	**0.8888**	**6.2258**

and local area consisting of pictorial and textual regions compared to FR-IQA models of natural images.

Among all FR metrics, QODCNN-FR can obtain the best performance and achieve a great improvement. The SRCC value of QODCNN-FR model is 2.44% higher than the MDOGS model, 1.26% higher than the CNN-SQE model, and the PLCC value of QODCNN-FR model is 2.32% higher than the SFUW model, 1.02% higher than the CNN-SQE model. In addition, our FR models fully utilize strong extracting features ability and generalization ability of CNN while CNN-SQE model only uses CNN to distinguish pictorial and textual regions which is mainly based on traditional ways.

4.2.2. No-reference image quality assessment

To demonstrate the excellent performance of our proposed QODCNN-NR models, it is compared with the above excellent FR-IQA models and the following state-of-the-art NR perceptual quality evaluation methods: BLINDS-II [11], BRISQUE [12], BLIQUP-SCI [23], NRLT [24], CNN-Kang [25], WaDIQaM-NR [32] and PICNN [35]. Among these NR-IQA approaches, the BLIQUP-SCI, NRLT and PICNN are designed for IQA of SCIs. In addition, CNN-SQE, CNN-Kang, WaDIQaM-NR and PICNN are CNN-based methods. For NR models, NRLT shows excellent performance among traditional NR methods, utilizing the global scope statistical luminance and texture features.

Our proposed QODCNN-NR model achieves the best performance on visual quality evaluation of SCIs compared with traditional FR and NR methods. Compared with CNN-based methods, our NR model obtains better performance than CNN-Kang model, WaDIQaM-NR model and PICNN model, and shows very close performance with CNN-SQE FR model. The PLCC value of our model is 5.66% higher than the NRLT model of NR-IQA, 1.69% higher than the MDOGS model of FR-IQA, 5.21% higher than the CNN-Kang NR model, 4.14% higher than the WaDIQaM-NR model and 0.48% higher than the PICNN model. Although the performance of PICNN is closed to the performance of our QODCNN-NR for that both two models are designed for SCIs. The proposed QODCNN-NR achieves better generalization ability from the results of cross-database experiments in Table 4. The main reason is that our method fully utilizes the strong feature extracting ability of CNN, and considers the divergence of local quality scores within a large SCI. In addition, our method fuses local scores to obtain visual quality of images by using an adaptive and effective weighting method.

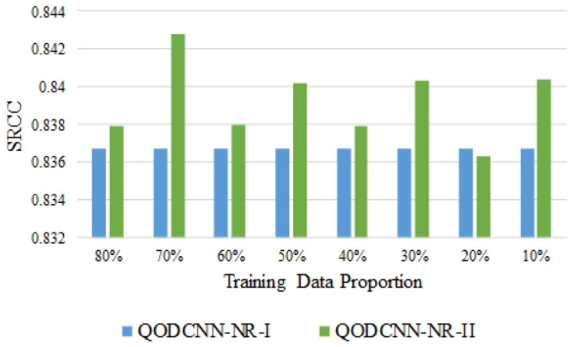


Fig. 7. Comparison of prediction performance under different proportion training data.

4.2.3. Performance comparison on individual distortion

The performances of our models and other FR models on each individual distortion type are shown in Table 3. From Table 3, it can be observed that GMSD, ESIM, SFUW and MDOGS show better performance on individual distortion type compared to other traditional methods. The main reason is that these approaches consider the edge and gradient information for the visual quality prediction of SCIs. The CNN-based learning models demonstrate superiority compared to traditional methods, our FR model demonstrates excellent performance for all distortion types, and even the QODCNN-NR model outperforms traditional FR approaches and achieves competitive performance for most distortion types. Especially, the proposed NR and FR models show outstanding performance on the GN, CC, JC and LSC.

4.3. Cross-database evaluation

To verify the generalization of proposed learning models, the CNN models are trained on SIQAD and tested on SCID. Considering that two databases contain different distortion types, experimental results of our models are given on 6 common distortion types consisting of GN, GB, MB, CC, JC and J2C. All distorted SCIs of six distortion types in SIQAD are used as training set. In the procedure of testing, the common practice of Mittal et al. [12] and Ye et al. [42] is employed, and 80% distorted SCIs associated with 32 reference SCIs of SCID are randomly chosen to evaluate the parameters of nonlinear function (Eq. (15)). The rest 20% distorted SCIs are utilized for testing. This operation is repeated with 1000 times, and the median performance is reported.

The cross-data performance of our proposed models including NR and FR is compared with the following FR methods and NR CNN-based model: PSNR, SSIM [1], GMSD [6], VSI [5], ESIM [18] and PICNN [35]. From Table 4, it can be observed that our models achieve better performance than PSNR, SSIM, GMSD, VSI and PICNN, and our FR model obtains similar performance with ESIM designed for SCIs. Comparing two CNN-based NR models including the proposed QODCNN-NR model and PICNN, it can be find that our NR model obtains significant performance improvement which means the proposed model owns stronger generalization ability.

4.4. Performance analysis

In order to fully demonstrate the effectiveness of our optimization approaches including training with valid image patches in Section 3.3 and weighting local score based on VLSD in Section 3.3, the pre-trained models using average weighting method of NR-IQA and FR-IQA in the first training stage are named as QODCNN-NR-I and QODCNN-FR-I. The fine-tuning models using average weighting method of second stage are denoted as QODCNN-NR-II and QODCNN-FR-II, and the fine-tuning models employing VLSD weighting method are named as QODCNN-NR-III and QODCNN-FR-III.

Table 3
Experimental results of our proposed models and other state-of-the-art FR models on different distortion types on SIQAD.

Method	Distortions	PSNR	SSIM [1]	GMSD [6]	SPQA [17]	ESIM [18]	SQMS [20]	SFUW [19]	MDOGS [22]	CNN-SQE [33]	QODCNN-NR	QODCNN-FR
PLCC	GN	**0.905**	0.881	0.899	0.892	0.899	0.900	0.887	0.898	–	**0.913**	**0.918**
	GB	0.860	0.901	0.910	0.906	**0.923**	0.912	**0.923**	0.920	–	**0.925**	**0.934**
	MB	0.704	0.806	0.844	0.831	**0.889**	0.867	0.878	0.842	–	**0.889**	**0.907**
	CC	0.753	0.744	0.783	0.799	0.764	0.803	**0.829**	0.801	–	**0.837**	**0.866**
	JC	0.770	0.749	0.775	0.770	**0.800**	0.786	0.757	0.789	–	**0.830**	**0.848**
	J2C	0.789	0.775	**0.851**	0.825	0.789	0.826	0.815	**0.861**	–	0.818	**0.857**
	LSC	0.781	0.731	**0.856**	0.796	0.792	0.813	0.759	0.832	–	**0.867**	**0.897**
	Overall	0.587	0.756	0.726	0.858	0.879	0.887	0.891	0.884	**0.904**	**0.901**	**0.914**
SRCC	GN	0.879	0.870	0.886	0.882	0.876	0.886	0.869	0.888	**0.893**	**0.905**	**0.907**
	GB	0.858	0.892	0.912	0.902	**0.924**	0.915	0.917	0.919	**0.924**	0.916	**0.921**
	MB	0.713	0.804	0.844	0.826	**0.894**	0.869	0.874	0.835	**0.904**	0.871	**0.895**
	CC	0.683	0.641	0.544	0.615	0.611	0.695	**0.722**	0.664	0.665	**0.700**	**0.778**
	JC	0.757	0.758	0.771	0.767	0.799	0.789	0.750	0.786	**0.847**	**0.815**	**0.829**
	J2C	0.775	0.760	0.844	0.815	0.783	0.819	0.812	**0.862**	**0.862**	0.795	**0.835**
	LSC	0.793	0.737	0.859	0.800	0.796	0.829	0.754	0.851	**0.887**	**0.882**	**0.898**
	Overall	0.561	0.757	0.731	0.842	0.863	0.880	0.880	0.882	**0.894**	**0.889**	**0.907**
RMSE	GN	**6.338**	7.068	6.521	6.739	6.827	6.921	6.876	6.558	–	**6.150**	**5.963**
	GB	7.738	6.570	6.310	6.430	5.827	6.611	**5.592**	5.964	–	**5.772**	**5.454**
	MB	9.229	7.697	6.982	7.222	**5.964**	7.204	6.236	7.012	–	**5.762**	**5.251**
	CC	8.282	8.412	7.828	7.618	8.114	7.743	**7.048**	7.528	–	**6.939**	**6.381**
	JC	6.000	6.230	5.941	6.000	**5.640**	5.983	6.143	5.779	–	**5.460**	**5.141**
	J2C	6.382	6.591	**5.459**	5.871	6.388	6.050	6.023	**5.293**	–	6.000	**5.286**
	LSC	5.330	5.825	**4.411**	5.166	5.215	5.104	5.555	4.738	–	**4.338**	**3.857**
	Overall	11.590	9.368	9.642	7.342	6.831	7.297	6.499	6.695	**6.115**	**6.226**	**5.801**

Table 4
Cross-database evaluation (both SRCC and plcc) of our proposed models and other FR models on SCID.

Method	Distortions	PSNR	SSIM [1]	GMSD [6]	VSI [5]	ESIM [18]	PICNN [35]	QODCNN-NR	QODCNN-FR
PLCC	GN	0.955	0.936	0.954	**0.958**	**0.956**	–	0.949	**0.960**
	GB	0.778	**0.871**	0.797	0.836	**0.870**	–	0.845	**0.866**
	MB	0.763	**0.880**	0.834	0.827	**0.882**	–	0.812	**0.849**
	CC	0.755	0.708	**0.811**	**0.878**	0.791	–	0.752	**0.817**
	JC	0.839	0.859	0.935	0.915	**0.942**	–	**0.935**	**0.942**
	J2C	0.918	0.859	**0.943**	**0.946**	**0.946**	–	0.890	0.940
	Overall	0.716	0.747	**0.851**	0.697	–	0.827	**0.849**	**0.882**
SRCC	GN	0.944	0.917	0.934	**0.946**	**0.946**	–	**0.947**	0.938
	GB	0.776	**0.870**	0.799	0.822	**0.870**	–	0.829	**0.856**
	MB	0.756	**0.859**	0.815	0.801	**0.861**	–	0.803	**0.828**
	CC	**0.732**	0.679	**0.715**	**0.816**	0.618	–	0.569	0.687
	JC	0.833	0.850	**0.934**	0.914	**0.946**	–	0.929	**0.935**
	J2C	0.907	0.846	**0.928**	**0.931**	**0.936**	–	0.865	0.916
	Overall	0.673	0.716	**0.843**	0.668	–	0.822	**0.848**	**0.876**

Table 5
Performance of QODCNN-NR-II with different proportion training data.

Index	QODCNN -NR-I	QODCNN-NR-II							
		80%	70%	60%	50%	40%	30%	20%	10%
PLCC	0.8849	0.8854	**0.8908**	0.8882	0.8858	0.8863	0.8877	0.8853	0.8830
SRCC	0.8650	0.8656	**0.8706**	0.8695	0.8671	0.8678	0.8694	0.8640	0.8647
RMSE	6.6930	6.6697	6.5924	**6.5907**	6.6426	6.6525	6.5936	6.6901	6.7400

4.4.1. *Influence of training data selection proportion*

to choose the appropriate proportion of training data, the proportions from 80% to 10% are taken into account for NR-IQA. The experimental results on SIQAD are reported in Table 5, and the cross-data experimental results are shown in Fig. 7. From Table 5 and Fig. 7, it can be seen that QODCNN-NR-II achieves better performance than QODCNN-NR-I except when the proportions of 20% are applied. Especially, QODCNN-NR-II obtains the best performance when 70% training data is employed to fine-tuned pretrained model. Considering that the local quality distribution is similar between FR and NR models, 70% of the training data is employed for the second stage of FR-IQA models.

4.4.2. *Effectiveness of two optimizations*

To show the advantage of our two optimizations including training with valid image patches and weighting local scores based on VLSD, the comparison of prediction performance between variations of our IQA model are shown in Figs. 8 and 9. The pre-trained models employing VLSD weighting are named as QODCNN-NR-I-W and QODCNN-FR-I-W. It can be observed that the performance is improved when two proposed optimizations are employed for both FR and NR models on two databases. In addition, the model with two optimizations obtains the best performance among variations of our IQA model. Therefore, experimental results show that the two optimizations are effective for visual perceptual prediction of SCIs. In addition, NR model has more performance improvement compared with FR model. We consider that FR CNN-based model more precisely predicts image quality for utilizing reference information. Two optimizations are proposed to improve accuracy of the model. When the accuracy of the model is higher, the amplitude of performance improvement is smaller.

4.4.3. *Reduce overfitting*

One of the important problems in machine learning is overfitting. In our model, two ways are adopted to solve this problem. Firstly, data

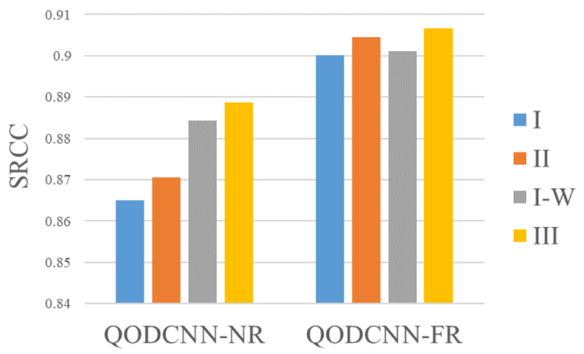


Fig. 8. Comparison of prediction performance between variations of our IQA model on SIQAD, including NR and FR models.

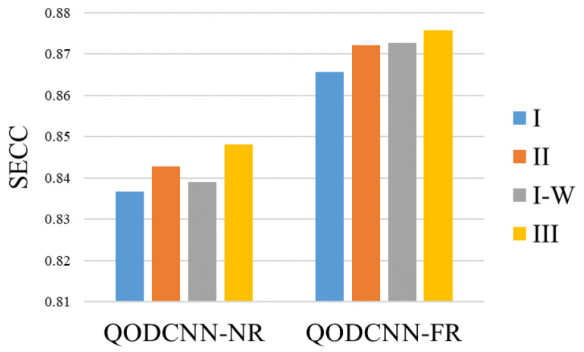


Fig. 9. Comparison of prediction cross-database performance between variations of our IQA model on SCID, including NR and FR models.

Table 6
Performance comparison between the proposed QODCNN-NR-I, QODCNN-NR-II and QODCNN-NR-FSIM. We conduct this test on SIQAD database. We highlight the best results with boldface.

	PLCC	SRCC	RMSE
QODCNN-NR-I	0.8849	0.8650	6.6930
QODCNN-NR-FSIM	0.8198	0.8166	8.0447
QODCNN-NR-II	**0.8908**	**0.8706**	**6.5924**

augmentation is employed to generate large data by cropping image to image patches with small size. Then in our model architecture, BN layers are added to improve learning ability of neural networks, L_2 regularization is used to generate sparse model, and both of them are effective approaches to reduce overfitting problem. Two regularization ways are tested to be valid and certainly helps to improve generalization ability. Our experiments on SIQAD and SCID verifies that our models achieve excellent performance on visual quality evaluation of SCIs.

4.5. Ablation study

4.5.1. Comparison of data selection method and FR-IQA-based method

The SCI contains more complex content compared to natural image. As a result, using FR-IQA method to predict quality of image patch shows poor performance. We have made an ablation experiments to verify the effectiveness of data selection method. Here, training a deep model with the predicted local scores by FSIM is named as QODCNN-NR-FSIM. The results are shown in Table 6. It can be observed that the QODCNN-NR-II owns better performance than QODCNN-NR-FSIM. In addition, the performance of QODCNN-NR-FSIM is worse than QODCNN-NR-I.

Table 7
Performance comparison between the proposed QODCNN-NR-III and QODCNN-NR-G. We conduct this test on SIQAD database. We highlight the best results with boldface.

	PLCC	SRCC	RMSE
QODCNN-NR-G	0.8976	0.8766	6.4268
QODCNN-NR-III	**0.9000**	**0.8888**	**6.2258**

4.5.2. Comparison of different weighting methods.

The biggest difficulty on distinguishing textual and pictorial regions in distorted images is the variety of noise types and intensities. In the previous IQA methods of SCI, they use the reference SCIs to distinguish textual and pictorial regions. Fang's FR-IQA model [19] obtains global quality by weighting local quality with gradient entropy. Using gradient entropy to distinguish textual and pictorial regions is useful for the FR-IQA model, but it is difficult for NR-IQA. The reason is that local gradient entropy is sensitive to noise. The VLSD is utilized to measure the importance of image patches, whose biggest advantage is owning strong noise robust ability. Here, we make ablation experiments to compare the performances of local gradient entropy and VLSD. The proposed model adopting the local gradient entropy is named as QODCNN-NR-G. As shown in Table 7, QODCNN-NR-III achieves better performance than QODCNN-NR-G, which verifies the high-performance of VLSD

5. Conclusion

In this paper, a neural network-based model is presented for full-reference and no-reference image quality assessment of SCIs with two optimizations. The OQDCNN model consists of three steps. In the first step, an effective CNN model is proposed to predict visual quality of SCIs for both FR and NR by employing concatenate layer to control input of reference information. Then, the Euclidean distance between DMOSs and predicted scores is employed to select valid data, and pre-trained model is fine-tuned with these data is utilized to optimize the model. For third step, local weights are measured using a noise robust index VLSD and applied to fuse local scores for obtaining image visual quality. Experimental results on SIQAD demonstrate that the proposed FR and NR models achieve the best performance compared to the state-of-the-art approaches. In addition, cross-data experimental results on SCID illustrate the strong generalization ability of our models and the efficiency of two optimizations including training with valid image patches and weighting local scores based on VLSD.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[2] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: Proc. IEEE Asilomar Conf. Signals, Syst, Comput, 2003, pp. 1398–1402.

[3] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.

[4] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.

[5] L. Zhang, Y. Shen, H. Li, VSI: A visual saliency-induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.

[6] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695.

[7] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, IEEE Trans. Image Process. 21 (4) (2012) 1500–1512.

[8] J. Wu, G. Shi, W. Lin, X. Wang, Reduced-reference image quality assessment with orientation selectivity based visual pattern, in: IEEE China Summit and International Conference on Signal and Information Processing, 2015, pp. 663–666.

[9] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, W. Gao, Reduced-reference quality assessment of screen content images, IEEE Trans. Circuits Syst. Video Technol. 28 (1) (2018) 1–14.

[10] L. Ma, S. Li, K.N. Ngan, Reduced-reference video quality assessment of compressed video sequences, IEEE Trans. Circuits Syst. Video Technol. 22 (10) (2012) 1441–1456.

[11] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.

[12] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.

[13] Y. Zhang, A.K. Moorthy, D.M. Chandler, A.C. Bovik, C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes, Signal Process., Image Commun. 29 (7) (2014) 725–747.

[14] Q. Wu, Z. Wang, H. Li, A highly efficient method for blind image quality assessment, in: IEEE International Conference on Image Processing, 2015, pp. 339–343.

[15] A. Mittal, R. Soundararajan, A.C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Process. Lett. 22 (3) (2013) 209–212.

[16] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, IEEE Trans. Image Process. 24 (8) (2015) 2579–2591.

[17] H. Yang, Y. Fang, W. Lin, Perceptual quality assessment of screen content images, IEEE Trans. Image Process. 24 (11) (2015) 4408–4421.

[18] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, K.K. Ma, ESIM: Edge similarity for screen content image quality assessment, IEEE Trans. Image Process. 26 (10) (2017) 4818–4831.

[19] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, Z. Guo, Objective quality assessment of screen content images by uncertainty weighting, IEEE Trans. Image Process. 26 (4) (2017) 2016–2027.

[20] K. Gu, et al., Saliency-guided quality assessment of screen content images, IEEE Trans. Multimed. 18 (6) (2016) 1098–1110.

[21] Z. Ni, H. Zeng, L. Ma, J. Hou, J. Chen, K. Ma, A gabor feature-based quality assessment model for the screen content images, IEEE Trans. Image Process. 27 (9) (2018) 4516–4528.

[22] Y. Fu, H. Zeng, L. Ma, Z. Ni, J. Zhu, K. Ma, Screen content image quality assessment using multi-scale difference of Gaussian, IEEE Trans. Circuits Syst. Video Technol. (2018) http://dx.doi.org/10.1109/TCSVT.2018.2854176,2018.

[23] F. Shao, Y. Gao, F. Li, G. Jiang, Toward a blind quality predictor for screen content images, IEEE Trans. Syst. Man Cybern.: Syst. (2018) http://dx.doi.org/10.1109/TSMC.2017.2676180.

[24] Y. Fang, J. Yan, L. Li, J. Wu, W. Lin, No reference quality assessment for screen content images with both local and global feature representation, IEEE Trans. Image Process. 27 (4) (2018) 1600–1610.

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

[26] Y. Li, X. Ye, Y. Li, Image quality assessment using deep convolutional networks, AIP Adv. 7 (12) (2017) 125324.

[27] B. Bare, K. Li, B. Yan, An accurate deep convolutional neural networks model for no-reference image quality assessment, in: IEEE International Conference on Multimedia and Expo, 2017, pp. 1356–1361.

[28] H. Wang, L. Zuo, J. Fu, Distortion recognition for image quality assessment with convolutional neural network, in: IEEE International Conference on Multimedia and Expo, 2016, pp. 1–6.

[29] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, A.C. Bovik, Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment, IEEE Signal Process. Mag. 34 (6) (2017) 130–141.

[30] J. Kim, A.D. Nguyen, S. Lee, Deep CNN-based blind image quality predictor, IEEE Trans. Neural Netw. Learn. Syst. (2018) http://dx.doi.org/10.1109/TNNLS.2018.2829819.

[31] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks, IEEE Trans. Image Process. 27 (3) (2018) 1202–1213.

[32] S. Bosse, D. Maniry, K.R. Mller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2018) 206–219.

[33] Y. Zhang, D.M. Chandler, X. Mou, Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation, IEEE Trans. Image Process. (2018) http://dx.doi.org/10.1109/TIP.2018.2851390.

[34] L. Zuo, H. Wang, J. Fu, Screen content image quality assessment via convolutional neural network, in: IEEE International Conference on Image Processing, 2016, pp. 2082–2086.

[35] J. Chen, L. Shen, L. Zheng, X. Jiang, Naturalization module in neural networks for screen content image quality assessment, IEEE Signal Process. Lett. 25 (11) (2018) 1685–1689.

[36] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE J. Sel. Top. Signal Process. 11 (1) (2017) 206–220.

[37] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, 2015, arXiv preprint arXiv:1512.03385.

[38] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference On Machine Learning, 2015.

[39] Vinod Nair, Geoffrey E. Hinton, Rectified linear units improve restricted boltzmann machines, in: International Conference On Machine Learning, 2010.

[40] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, CoRR, abs/1412.6980.

[41] Methodology for the subjective assessment of the quality of television pictures, document rec. ITU-R BT, 2012, pp. 500–511.

[42] P. Ye, J. Kumar, D. Doermann, Beyond human opinion scores: Blind image quality assessment based on synthetic scores, in:IEEE Conference Computer Vision and Pattern Recognition, 2014, pp. 4241–4248.