Displays 71 (2022) 102101

Contents lists available at ScienceDirect

Displays

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

DeepRPN-BIQA: Deep architectures with region proposal network for natural-scene and screen-content blind image quality assessment

Mobeen ur Rehman [a,b,c], Imran Fareed Nizami [d,*], Muhammad Majid [e]

[a] Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Korea
[b] Institute of Avionics and Aeronautics (IAA), Air University, Islamabad 44000, Pakistan
[c] Department of Avionics Engineering, Air University Islamabad, Pakistan
[d] Department of Electrical Engineering, Bahria University, Islamabad, Pakistan
[e] Department of Computer Engineering, University of Engineering & Technology Taxila, Pakistan

ARTICLE INFO

MSC:
41A05
41A10
65D05
65D17

Keywords:
Screen content images
Visual saliency
Blind image quality assessment
Regional proposed networks

ABSTRACT

With the emerging use of technology and screen-oriented applications in our daily life, screen content images have gained the same importance as natural scene images. This results in many natural-scene and screen-content blind image quality assessment (BIQA) models to evaluate the perceptual quality without any prior information regarding the reference image. Recently, patch-based techniques for image quality assessment (IQA) have shown promising results. As per our knowledge, no IQA technique in literature is available that can be equally effective for both natural-scene and screen-content images. In this work, we have proposed a deep architecture with a region proposal network (RPN) for blind natural-scene and screen-content image quality assessment, named DeepRPN-BIQA. The proposed architecture computes visual saliency using RPN to extract important regions having a high contribution towards the image quality. Important regions are extracted by utilizing the texture and edges of images by sliding the network over the extracted feature map from deep architectures i.e., VGGNet and ResNet. The regions proposed (RP) that overlap more than 60% are merged into one proposal and are called the region of interest (ROI). The overlap between RPs is computed using anchors having 3 different scales and aspect ratios. A local quality score is computed over each ROI and the total quality score is computed by taking the average of all the local quality scores. Experimental results show that the DeepRPN-BIQA shows a high correlation between mean observer score and predicted quality score and performs better than other models for screen content images, synthetically distorted images, images taken in real-life conditions using mobile phone cameras, and large scale image quality assessment database.

1. Introduction

The advancement in technology has made sharing and usage of multimedia content, part of our daily life. Distortion can be introduced in multimedia content due to limitations in technology. The most common type of multimedia content used in daily life is natural-scene images acquired using camera or screen content images, which involve computer-related content [1]. The distorted natural-scene images can mainly be categorized into three categories *i.e.,* images distorted by a single type of distortion, images distorted by multiple distortions, and the images taken by cameras in the real world where the type of distortion is unknown [2]. Image quality assessment (IQA) has gained importance since human beings are stimulated by visual content and the quality of visual content can help in improving the user experience in many applications. IQA is used for benchmarking of image enhancement algorithms [3–6], evaluating the performance of steganography algorithms, and selecting network parameters so that the multimedia content is visually appealing to human beings [7–10].

IQA methods are broadly divided into subjective and objective IQA. Subjective IQA is the evaluation of image quality performed by human observers. Since human observers are the ultimate user of images, therefore, subjective IQA is taken as the standard [11]. Subjective IQA suffers from many drawbacks *i.e.,* large amount of time for the evaluation of image quality, prior knowledge about the image can affect the observer's ability to evaluate the image quality and subjective IQA is a tedious task. Objective IQA predicts the quality of images using objective metrics and computational models that correlates highly with the perceptual quality of images [12]. Objective IQA can be categorized into three main categories. Full reference image quality assessment (FR-IQA), in which the reference/pristine version of the image is available while predicting the quality of the distorted image. FR-IQA methods

* Corresponding author.
E-mail address: imnizami.buic@bahria.edu.pk (I.F. Nizami).

https://doi.org/10.1016/j.displa.2021.102101
Received 15 April 2021; Received in revised form 21 August 2021; Accepted 20 September 2021
Available online 27 October 2021

usually measure the dissimilarity between the pristine and distorted image using distance metrics [13–17]. In reduced reference image quality assessment (RR-IQA) methods partial information about the pristine image is known while predicting the quantitative measure for quality of distorted image [18–23]. Whereas in no-reference image quality assessment (NR-IQA) also known as blind image quality assessment (BIQA), no information regarding the original image is available while predicting the quantitative measure for quality of distorted image [24–29].

Visual saliency/region of interest (ROI)/visual attention is the most important method through which locations in an image can be determined towards which people are most interested. In recent times, visual saliency has been widely studied by researchers to investigate which areas of an image attract the attention of the human observer. Visual saliency is very closely related to image quality since, presence of distortion in an image can affect the saliency map [30]. In a patch-based IQA technique, visual saliency can help in determining the quality of a local patch and it can also be employed as a weighing function that reflects the importance of a local region for IQA [31]. The premise of using visual saliency is that the distortion occurring in an area that attracts the attention of the observer is more annoying than any other area. Saliency maps can help in IQA that process local visibility due to distortion with its corresponding saliency [32,33]. The perceptual quality of the salient region in an image tends to represent the perceptual quality of the whole image and hence, it should be helpful to incorporate the visual saliency in image quality metrics [32].

Most BIQA techniques already present in the literature are usually for a certain set of image content *i.e.,* real-life images taken from a camera or images affected by synthetic distortion, natural scene images (NSI), or screen content images (SCI). Literature shows many types of research which have employed visual saliency with machine learning-based techniques to assess the quality of an image [34–36]. In [34], a comparative perceptual quality assessment technique based on brain theory is proposed. The technique emulates the process of comparing two input stimuli to assess the quality of images. A technique for automatic region selection for assessment of sharpness in images taken using mobile devices is proposed in [35]. The technique uses local textures, depth information and inter picture differences to select optimal local regions. In [36], subjective quality assessment of 40,000 images and 120,000 patches with around 4 million subjective scores are collected. The technique builds a deep region based architecture for image quality assessment. Most deep-based BIQA techniques do not use deep CNN-based architecture to apply visual saliency. To the best of our knowledge very limited BIQA techniques have used deep-based CNN to apply visual saliency to extract the most relevant patches called ROI and compute local quality score over each patch for prediction of overall image quality score [37]. Moreover, no deep-based BIQA technique exists that performs better in predicting the quality score for images distorted by synthetic, real-life images, large data subjective IQA databases, SCI, and NSI. This paper proposes an end-to-end deep-based region proposed network (RPN) methodology for BIQA called DeepRPN-BIQA. The proposed methodology is based on CNN to extract the ROIs using an RPN. The resultant ROIs are given as input to a CNN to predict the quality score of the image. The major contributions of this work are as follows,

- An end-to-end deep CNN-based BIQA technique is proposed that uses visual saliency to extract region proposals, which are used to predict the quality of the image, and to the best of our knowledge, such a method of quality prediction is not used in literature.
- The proposed technique provides a unified approach to predict the image quality of real-life images taken from a camera or images affected by synthetic distortion, NSI, or SCI.

The rest of the paper is organized as follows. Section 2 discusses previously proposed deep neural network-based and screen content BIQA techniques. Section 3 describes the proposed methodology. Section 4 presents the experimental setup and results, and the conclusion is presented in Section 5.

2. Related work

2.1. Deep neural network based BIQA

Recently, deep convolutional neural networks have been used for IQA [38–41]. In [39] and [40], an end to end feature learning in CNN was used to predict the image quality using transfer learning. A shallow network with the input-patch size of 32 × 32 having a single convolutional layer followed by two fully connected layers was used [39]. The architecture in [39] was improved by the addition of more layers [40], where a twelve-layered CNN architecture was used with the same input image size for improved image quality assessment. In [42], a CNN approach was proposed, which uses supervised learning filters for the quality prediction of a single distorted image. In [43], authors have adopted a shallow CNN approach in combination with the regressor for enhancing the quality prediction of a distorted image. A similar combinational approach was used in [44] and [45] which uses a deep belief network for feature extraction and support vector machine (SVM) as a regressor for quality prediction.

An AlexNet [46] based architecture was proposed in [41] where deep-based features were extracted and then regressed using human allocated scores. In [47] ResNet [48] and AlexNet [46] architectures were retrained for image quality assessment and achieved impressive results. One of the most recent approaches in the literature proposed by Mai et al. [49] extracts features from multiple scaled input images having a constant aspect ratio. The features were extracted using multi-net in which all nets are pre-trained on VGGNet [50]. Further after score prediction on different scales, the scene aware aggregation layer was used to predict the aggregative score. In [51], a multi-task learning approach for BIQA was proposed that extracts natural scene statistics (NSS) based features and then performs the quality score prediction.

The distribution of mean opinion scores that shows the technical and aesthetic quality of images was predicted using a CNN [26] that can assist in the process of photo editing and enhancement. In [52], pre-trained DNN was used for feature extraction that can predict the quality score of images affected by generic distortions. Fine-tuning of DNN was then performed to assess the image quality. A combination of handcrafted features and CNN was used for BIQA by dividing the process into two steps *i.e.,* an objective distortion part, which learns to predict the error map, and a human visual system-related part that predicts the image quality score [53]. A bi-linear DNN was used to predict the image quality of authentically and synthetically distorted images.

In [54], multi-level representations were modeled utilizing a very deep neural network and extracting features at each layer to predict the quality score corresponding to each layer. The overall quality score is computed by averaging all the quality scores predicted at each layer. A deep clustering-based ensemble approach for BIQA that utilizes layer by layer characteristics to extract features of fixed length, which are divided into clusters using fuzzy C-means, and for each cluster a particular fitting function is learned using the differential mean opinion score (DMOS) for prediction of the quality score. In [55], multivariate Gaussian (MVG) distribution was used to assess the quality of images. The image was divided into patches and only those patches that have high contrast were considered. The MVG of pristine and the distorted images were compared to assess the local quality scores of each patch and, lastly, deep activation pooling was applied to suppress the less important scores and give higher weight-age to important scores. Low-level visual features and high-level semantic features were used in [56] to represent the information in the image. The image was pre-processed by normalizing local brightens, then the image was divided into non-overlapping image blocks. The blocks were used with CNN to predict the quality of images. In [57], multiple belief networks based on Boltzmann machines were utilized with multiple regression models for the prediction of image quality score. A CNN designed for IQA was proposed in [58] that uses a new database with one million images

and a pseudo mean opinion score was considered as the quality score of each image. An IQA-orientated CNN method was designed, in which the hierarchical degradation is considered. Multilevel features were extracted and hierarchical degradation concatenation was optimized to predict the quality score of the image using end-to-end CNN framework.

As human observers are the ultimate users of images, therefore the properties of the human visual system (HVS) can be utilized to form an effective metric to assess the perceptual quality of an image. Among the many properties of HVS, visual saliency seems to be the most trivial HVS characteristic for visual information processing [32]. Visual saliency ignores the unimportant part of the image, whereas it will selectively process the important part of the image. From the perspective of IQA, the distortions present in the salient parts of the image attract more attention from the human observer. In other words, the perceptual quality of the salient regions can represent the quality of the whole image [32]. Therefore, deep CNN-based BIQA techniques that do not use visual saliency may not perform well as the deep-based BIQA techniques that make use of visual saliency to extract ROI that represents the overall quality of the image.

2.2. Screen content images based BIQA

Recently, a lot of interest has been shown in the IQA of SCIs since they are more relevant to imaging and video applications [59]. In [59], SCI segmentation along with local and global perceptual feature representation was considered. Two types of local features were extracted for the sharp edge patches in the image *i.e.,* , the entropy of contrast features and the local phase coherence. Average feature pooling was performed to fuse the features. The global features were extracted using the BRISQUE IQA technique, which was combined with the local features to predict the quality score using support vector regression. Two sets of features were extracted to assess the quality of SCIs. The first set of features utilize an orientation selectivity mechanism to model the visual distortion in SCIs and the second set of features utilize the statistical distribution of the histogram of orientation. The extracted features were given as input to the support vector regression for prediction of image quality score [60]. In [61], a CNN was used to segment the textual, animated/graphic, and natural portion of the SCI. The textual and animated/graphic section of the SCI was considered as one class and, the graphics and natural portion of the SCI were considered as the second class. The overlapping of the graphics portion in both classes allows the graphics portion to be analyzed by both text-based and natural-image-based features. The technique was an end-to-end CNN-based approach where the CNN gives the SCI quality score as the output.

In [62], a Gabor feature-based model was used to assess the quality of SCIs, which was based on the fact that the imaginary part of the Gabor filter has odd symmetry and it can be easily used to detect edges. The local similarity of Gabor features along with two chrominance components were used in a feature pooling strategy to extract features for predicting the image quality score [63]. In [64], spread transform dither modulation watermarking scheme based on hybrid just noticeable distortion model was utilized in the YCrCb color space to assess the quality of SCI images. A dictionary of local and global quality features are extracted over image patches using k-means clustering and support vector regression (SVR) was used to assess the quality of images [65]. The difference of Gaussians at multiple scales was used to characterize edge information and edge similarity to compute the image quality score [66]. Image quality assessment of SCIs was performed using content-specific codebooks [67]. The features were automatically extracted by diving the image into textual and pictorial portions. Two code-books *i.e.,* for textual and pictorial content were formed separately. Each content portion was divided into patches and the final quality score was computed by aggregating the quality score from each patch. In [68], the quality score of SCIs was computed by comparing the macroscopic and microscopic structures of images. A deep neural network was introduced that uses three modules *i.e.,* producing pseudo-natural inputs using naturalization module, a series of pooling and convolutional layers for feature extraction and prediction, and two fully connected layers to form a prediction network [69]. In [70], two strategies for quality score prediction of SCIs were introduced using CNN *i.e.,* FR-IQA and NR-IQA. In the first strategy, a pseudo-reference image was generated to imitate an FR-IQA scenario on image patches. The quality score predicted using FR-IQA was utilized as the ground truth for the NR-IQA strategy. For the NR-IQA strategy, all the image patch qualities of one entire SCI were fused to obtain the SCI quality. This takes into account the diverse contents in different image patches. A dictionary-based on histogram representation of multi-scale local patterns was used to assess the quality of SCIs [71]. A pursuit algorithm was used to efficiently code the dictionary and an SVR model was used to predict the SCI quality score.

2.3. Visual saliency based image quality assessment

Visual saliency has been used to improve the performance of NR-IQA techniques. Previously the visual saliency models were frequently used for RR-IQA. The RR-IQA technique for NSIs and SCIs utilizes visual saliency as an image feature rather than using it as a weighing map. The image saliency was detected using an image signature, which is basically a binary image descriptor [72]. In [73], global and local saliency attributes were used to generate a saliency map. The final quality score was computed by considering the sharpness, edges, and saliency map of the image. In [29], the existing saliency model was applied to the image to select patches of interest, and then the CNN model was used to predict the quality of images. An NR-IQA technique utilizes CNN to extract features from image patches and local saliency to predict the image quality score [74]. The local saliency was computed by taking into account the difference between the saliency map of the local patch and its corresponding region of the saliency map for the whole image. The IQA technique in [75] uses contrast and visual saliency to characterize the image quality. The standard deviation of the contrast map and visual saliency maps were used to compute the image quality score. The FR-IQA technique in [75] uses an end-to-end CNN that uses visual saliency to predict the quality score of SCI images. FR-IQA approach to IQA using visual saliency was used to assess the quality of synthetically distorted images by giving extra attention to the center by increasing the sensitivity of the similarity maps in that region [76]. Speeded-up robust features, gradient map extracted from image patches, and visual saliency index were used with the euclidean distance to predict the image quality score [77].

In [78], the NR-IQA technique was proposed that assumes that visual artifacts cause visual saliency deviation in the image. The technique uses this visual deviation in the visual saliency map to assess the quality of an image. Visual saliency was used as a feature and then as a weighting function in predicting the quality score of synthetically distorted images [30]. In [79], an exhaustive search for the best visual saliency model among various IQA techniques was performed. Visual saliency for the quality assessment of SCIs was introduced in [80], where recognition and clarity of the salient areas, *i.e.,* text, and their surroundings were detected through a simple detector, motivated by the behaviors of “fixation” and “saccade".

3. Proposed methodology

The proposed deep convolutional neural network with region proposal network for blind image quality assessment, which we called DeepRPN-BIQA is shown in Fig. 1. The proposed methodology consists of three steps *i.e.,* CNN architecture for feature map extraction, extracting ROIs using RPN for extraction of the feature vector and computing the local quality score of each ROI, and then averaging to compute the quality score of the whole image. Each step is discussed in detail below.

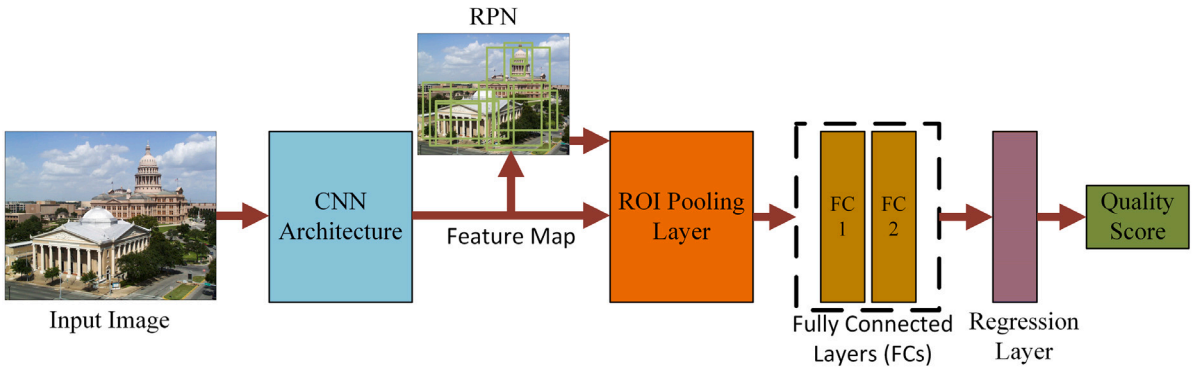

Fig. 1. Proposed architecture.

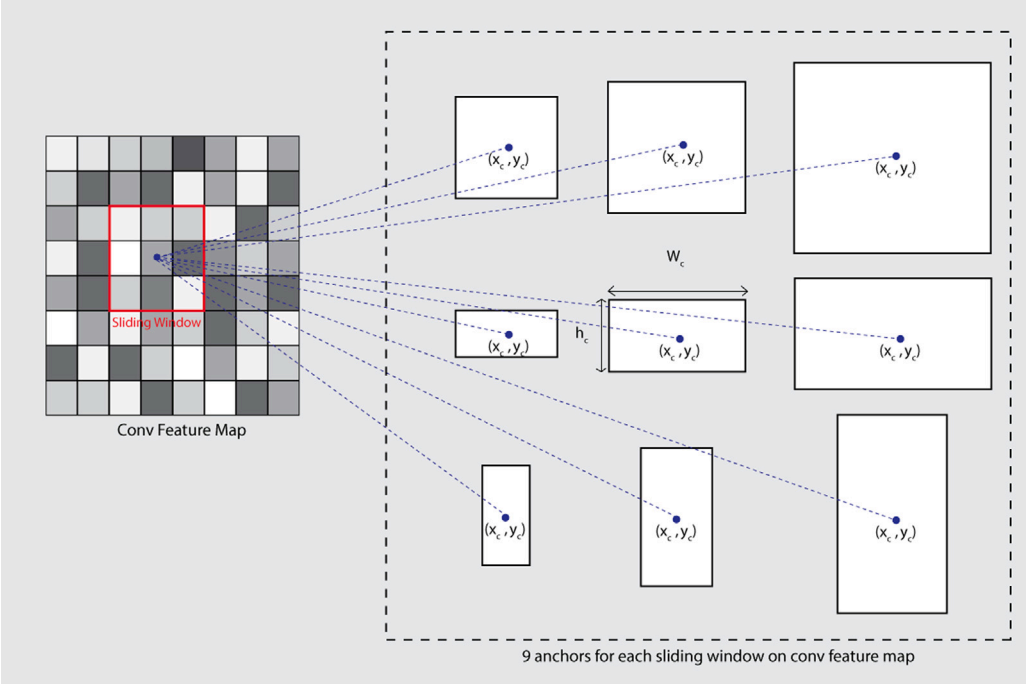

Fig. 2. Anchors at different scales and aspect ratios.

3.1. CNN for feature map extraction

A CNN architecture is used to extract the feature map from the input distorted image. The extracted feature map is given as an input to the RPN. Two pre-trained CNN's are used in this work to extract feature maps *i.e.,* VGGNet [50] and ResNet [48]. The description of each CNN is explained in detail as follows.

3.1.1. VGGNet

VGGNet has two different variants VGG-16 and VGG-19 with a varying number of weight layers [50]. VGGNet reduces the number of parameters utilized in convolutional layers, which ultimately improves the training time. The variant used in this work is VGG-16, which includes 16 convolutional layers with a convolutional kernel of size 3×3. Max pooling kernel of size 2×2 is used in VGG with the stride of two. The reason for choosing VGG-16 is because it has a lower training error.

As VGGNet uses small receptive fields, therefore, it requires more rectified linear units (ReLUs) that make the system to be eligible to have a more discriminative decision function. The architecture is trained on the ImageNet database which holds more than 1 million images [81]. The input to VGG-16 CNN is an RGB image of size 224×224, therefore, we resize the images and obtain a feature map using the VGG-16 CNN. The input images are resized using bilinear interpolation to a size of 224×224. The only pre-processing performed in VGGNet is to subtract the mean values of R, G, and B channels from each pixel. The convolutional stride in VGGNet is fixed to 1 pixel and the spatial padding is performed such that the spatial resolution is preserved after convolution. The max-pooling in VGGNet is performed over 2×2 pixels, with a stride of 2. The stack of convolutional layers is followed by three fully connected layers (FC). The first FC contains 4096 channels and the last FC layer has 1000 channels. The last layer is the soft-max layer. VGGNet performs better since three rectifier layers are used, which makes the decision function more discriminative and the number of parameters for L layers with C channels are given by L^2C^2. The training on the VGGNet is performed using a batch size of 256, momentum is set to 0.9 and the penalty multiplier is set to 5×10^{-4}. The dropout ratio is set to 0.5, the learning rate is set to 10^{-2} and then reduced by a factor of 10 when validation accuracy stops increasing.

3.1.2. ResNet

ResNet is based on residual functions and shortcut connections. The input image size of ResNet is 224×224. Residual representation solvers

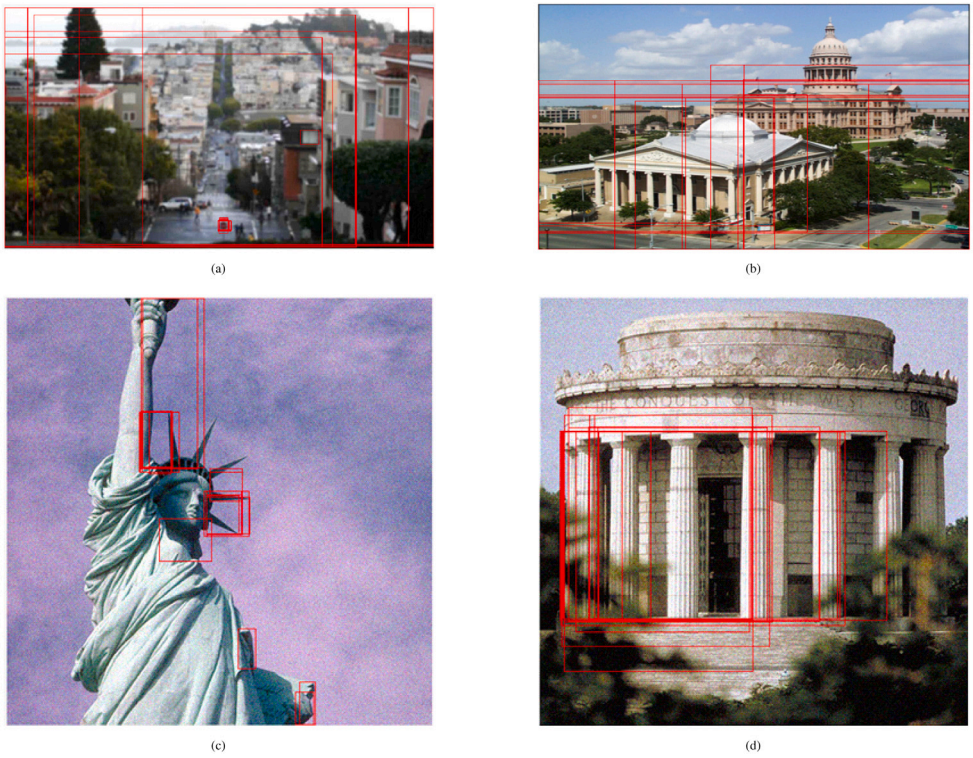

Fig. 3. The region proposals extracted using RPN for different input images.

offer the advantage of converging to the solution much faster than standard solvers, whereas highway networks based on a short connection with a gated shortcut address the problem of exploding and vanishing gradients. ResNet has a larger number of layers in comparison to VGGNet which increases the depth of the network. ResNet has a total of 34 parameter layers. 33 of these layers are convolutional layers and one is a fully connected layer. Short connections turn the ResNet network into a residual network. Identity shortcuts can be used when input and output are of the same dimensions. When the dimensions are increased, two options are considered *i.e.,* in the first option identity mapping is performed using zero paddings for increasing dimensions and in the second option, a projection shortcut is used to match dimensions. When the shortcuts go across feature maps of the same size, they are performed with a stride of 1 and when the short connections go across feature maps of different sizes then the short connections are performed using a stride size of 2. Batch normalization after each convolution layer and before activation is performed. The learning starts from 0.1 and is divided by 10 when a solution does not improve anymore. The model has trained up to 6×10^4 iterations and the weight decay of 0.0001 is used, with a momentum of 0.9. No dropout layer is used in ResNet. Increasing the depth of architecture up to a certain extent increases system performance. The increase in layers demands efficient weight assignment especially to the earlier layers, so that problem of vanishing gradient can be addressed. Another issue related to larger networks is performance degradation. Large networks involve various parameters to be optimized, that is where the residual network comes into action, it constructs modules in the network known as residual models. Residual neural networks utilize skip connections over the number of layers and use activation from the previous layer till the time weights of the adjacent layer are learned. So basically during the training process of the network few layers are skipped, which helps the network to learn residual rather than learning the true output. Therefore, using backpropagation weight learning is done using,

$$\Delta w^{l-k,l} = -\tau a^{l-k} \delta^{l}, \tag{1}$$

where τ is the learning rate of the network, which is always less than zero, a^l is the activation of the neuron at layer l, and δ^l is the calculated error signal of the neuron at layer l. Once, the feature map is extracted, it is given as input to the RPN.

3.2. Region proposal network (RPN)

RPN based on a fast recurrent convolutional neural network (RCNN) is used to extract ROIs. A fast RCNN has three main advantages over the RCNN. Firstly, training in RCNN is a multistage process that includes fine-tuning the object proposals using log loss, then fitting SVM on the convolution network features and lastly, bounding box regressors are learned. Secondly, training is expensive in terms of space and time *i.e.,* for SVM and bounding box regressor training features are extracted on each image and written to the disk, which requires a lot of time and a large amount of storage space. Thirdly, object detection is slow as test features are extracted over each object proposal in each test image. RPN generates different object proposals based on texture and edges. A network is slid over the extracted feature map to get the region proposals. The spatial window of size $m \times m$ of the last layer convolutional map is fully connected to the network, and the value is kept to 3. The 9 anchors are created for each sliding window, having anchor center (x_c, y_c) with 3 different scales of heights and weights. A value q is computed for all the anchors to determine the overlap of an anchor with the ground truth and is defined as,

$$q = \begin{cases} 1, & \text{if } IoU > 0.7 \\ -1, & \text{if } IoU < 0.3, \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where IoU is the intersection over the union between the anchor and ground truth box μ_{GT} and is calculated as,

$$IoU = \frac{\mu_{GT} \cap Anchor}{\mu_{GT} \cup Anchor}. \tag{3}$$

Fig. 2 shows the anchors used in the proposed approach. RPN generates a large number of proposals that may overlap with each

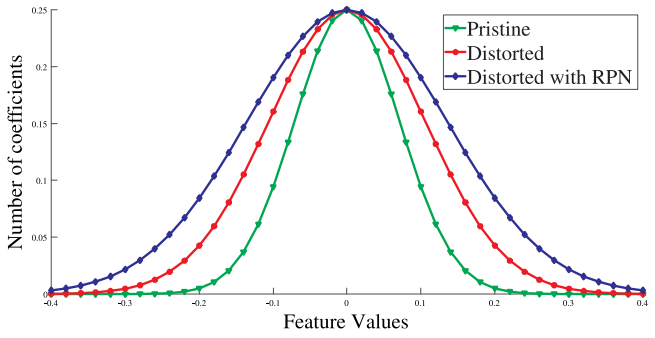

Fig. 4. Distribution of feature vector of (a) pristine image (b) distorted image (c) ROI using RPN.

other. Non-maximum suppression (NMS) is engaged to eliminate the overlapping, in which overlapping regions are merged if they have high IoU. A score is assigned to the remaining proposals, based on which the relevant proposals are selected. For BIQA, we have used a maximum of 4 proposals, which are ranked top. For RPN training, PASCAL 2011 [82] dataset along with 2000 annotated images are used. The 2000 additional images are also used that include images from natural IQA datasets and screen content IQA datasets. Moreover, images enriched with texture and edges were also taken into account for training RPN, so that performance can be improved.

For training purpose labeling of each region, the proposal is conducted to let the system learn that either region is selected as ROI or not. A threshold is selected and if the proposal and ground truth have an IoU greater than the threshold, it is assigned a positive training proposal. On the other hand, another threshold is set on the minimum side, where if IoU gets lower than that proposal would be a negative sample. A multi-task loss function is utilized as proposed in faster RCNN [83] which is defined as,

$$L(a_i, b_i) = \frac{1}{N_c}\sum_i L_c(a_i, a_i^*) + \lambda\frac{1}{N_r}\sum_i a_i^* L_r(b_i, b_i^*), \tag{4}$$

where the number of anchors is represented by i, predicted probability for an anchor to be selected is represented by b_i, ith anchor's labeled ground truth is represented by b_i^*, a_i is a vector having predicted bounding box coordinates, while vector with ground truth bounding box coordinates is represented by a_i^*.

For normalization purpose two factors N_c and N_r are used along with λ which is a balancing parameter. The values of N_c and N_r are set to 256 and 2400 respectively. Further, the value of λ is kept as 10 to have equal weightage. The loss function is computed to determine whether the bounding box carries an object or not. Any bounding box with a larger number of edges will be represented as a bounding box having an object. Since edges hold substantial information that can be used to evaluate the quality of an image. Fig. 3 shows images from IQA databases and screen content images and the region proposals extracted using the RPN. Fig. 4 shows the distribution of feature vector mapped to a normal Gaussian distribution. It can be observed that the distance between the distribution of pristine and distorted images increases when ROI using RPN is utilized.

3.3. Feature vector and regression

Once region proposals are extracted, then the corresponding feature map extraction is performed. ROIs are extracted from the feature map and a feature vector of fixed length is formed using a pooling layer. This fixed-length feature map is then given as input to fully connected layers. The purpose to pass feature vectors from fully connected layers is to avoid the over-fitting problem. Fully connected layers reduce the feature vector length and keep features that help in predicting the quality score of images, which correlate with the quality of images as perceived by human observers. This process is carried out for feature vectors of all ROIs. The feature vector extracted by fully connected layers is then used with regression layers that assign a quality score to each ROI. The average of the quality scores of all ROIs is obtained by using averaging layer, which gives a final quality score of the input image.

4. Experimental results

4.1. IQA databases and experimental setup

The proposed architecture is developed for both natural and screen content images. Seven different databases are taken into account out of which five are of natural images, and the other two are of screen content images. Moreover, out of the five natural-scene image databases, four databases namely LIVE [84], TID2013 [85], CSIQ [86] and KADID-10k [101] are of the single distorted images and one CLIVE [87] is of the multi distorted images. LIVE database has five different distortion types with a total of 29 reference images. The database provides 779 images along with their mean observer scores. The TID2013 dataset has 25 reference images, which are used to develop a dataset of 3000 images with the help of 24 distortion types. CSIQ has a total of six different distortion types that used over 30 pristine images to produce a total of 886 images. KADID-10k is the largest labeled database with 10,000 distorted images in total. It has 81 pristine images, which are distorted using 25 different distortion types used on 5 different levels. CLIVE is collected using multiple mobile devices along with various time frames in both day and night. In total CLIVE dataset contains 1162 images.

The two-screen content image databases used are SIQAD [102] and SCID [103]. SIQAD has a total of 20 pristine screen content images, which are used to create 980 distorted images with the help of 7 distortion types at 7 different levels. A single stimulus scaling paradigm is used to collect the mean observer score. SCID contains 40 pristine images. In total database contains 1800 distorted images prepared using nine distortion types with five different levels.

The database is divided into two non-overlapping sets of training and testing. 80% images from the dataset are used for training and 20% images are used for testing. The images present in the training set are not present in the testing set. Training and testing are performed over 100 iterations, randomly selecting non-overlapping sets for training and testing to nullify any bias due to the selection of images.

4.2. Performance analysis

Table 1 shows the performance of the proposed DeepRPN-BIQA with 15 state-of-the-art BIQA techniques over LIVE, TID2013, CSIQ, and CLIVE subjective IQA databases. Nine of these techniques extract hand-crafted features, whereas six use deep neural network-based features for BIQA. It can be observed from the results that the proposed DeepRPN-BIQA using ResNet for the extraction of RPNs and ROIs is ranked top over all the four databases. On the LIVE database the proposed method with ResNet performs best with the SRCC score of 0.9871, NSSADNN [51] is ranked second with the SRCC score of 0.9860, and DIQA is ranked third with the SRCC score of 0.9750. The proposed methodology with ResNet is again ranked top over TID2013 database with the SRCC score of 0.9487, M3 [93] is ranked second with the SRCC score of 0.9369 and BRISQUE [91] is ranked third with the SRCC score of 0.9336 on the CSIQ database, the proposed methodology with ResNet performs best with the SRCC score of 0.9449 whereas, the proposed methodology with VGGNet is ranked second with the SRCC score of 0.9375 and Pseudo [94] is ranked third with the SRCC score of 0.9309. On the CLIVE database, the proposed methodology with ResNet and VGGNet are ranked first and second with the SRCC score of 0.8427 and 0.8261 respectively. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet

Table 1
Overall performance comparison of the proposed NR-IQA technique in terms of median SRCC, PCC and RMSE for LIVE, TID2013, CSIQ and CLIVE databases.

IQA Technique	LIVE [84]			TID2013 [85]			CSIQ [86]			CLIVE [87]		
	SRCC	PCC	RMSE	SRCC	PCC	RMSE	SRCC	PCC	RMSE	SRCC	PCC	RMSE
BIQI [88]	0.8042	0.8280	15.388	0.8415	0.8598	0.7872	0.7598	0.8353	0.1542	0.4621	0.4490	19.981
DIIVINE [89]	0.9003	0.8943	12.329	0.8923	0.8867	0.6714	0.8697	0.9010	0.1249	0.5902	0.6075	18.742
BLIINDS-II [90]	0.9304	0.9361	9.5185	0.9046	0.9197	0.6117	0.9003	0.9282	0.1028	0.4618	0.4473	20.011
BRISQUE [91]	0.9393	0.9430	8.7214	0.9336	0.9360	0.5442	0.9085	0.9356	0.0980	0.6089	0.6088	17.742
NSS [92]	0.9470	0.9500	8.7141	0.9260	0.9200	0.5321	0.9050	0.9250	0.1002	0.6121	0.6290	17.241
M3 [93]	0.9511	0.9468	8.0444	0.9369	0.9406	0.5377	0.9243	0.9457	0.0909	0.5064	0.5532	20.573
Pseudo [94]	0.9355	0.9364	8.7251	0.8890	0.8854	0.5672	0.9309	0.9315	0.0783	0.5955	0.5987	18.5512
FSI-RR [95]	0.8826	0.8821	12.8720	0.5798	0.6111	0.9945	0.9175	0.9265	0.1011	0.5810	0.5834	18.1911
MSDD [96]	0.9472	0.9488	8.7719	0.9324	0.9416	0.4687	0.9105	0.9227	0.1080	0.4871	0.4862	19.6253
MGDNN [97]	0.951	0.949	–	–	–	–	–	–	–	–	–	–
DIQaM-NR [25]	0.960	0.972	–	0.835	0.855	–	–	–	–	0.606	0.601	–
BIECON [98]	0.958	0.932	–	0.721	0.765	–	0.825	0.838	–	0.595	0.613	–
MEON [99]	0.951	0.953	–	0.811	0.828	–	0.839	0.850	–	0.688	0.693	–
DIQA [100]	0.975	0.977	–	0.825	0.850	–	0.884	0.915	–	0.703	0.704	–
NSSADNN [51]	**0.986**	0.984	–	0.844	0.910	–	0.893	0.927	–	0.745	0.813	–
VGGNet	0.9189	0.9144	9.4172	0.8999	0.8924	0.5964	0.9193	0.9014	0.1063	0.7257	0.7132	18.8374
ResNet	0.9256	0.9312	8.9256	0.9190	0.9133	0.5159	0.9014	0.9032	0.0986	0.7391	0.7330	17.2182
Proposed using VGGNet	0.9716	0.9742	5.1258	0.9226	0.9183	0.5066	0.9375	0.9340	0.0493	0.8261	0.8193	8.701
Proposed using ResNet	0.9835	**0.9872**	**4.3682**	**0.9487**	**0.9461**	**0.4794**	**0.9449**	**0.9504**	**0.0142**	**0.8427**	**0.8408**	**8.394**

and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures. It can be observed that the proposed technique with ResNet is ranked top over all the four subjective IQA databases whereas, the proposed technique with VGGNet is ranked among the top four in three out of four databases.

Table 2 shows the performance of the proposed methodology in comparison with 11 state-of-the-art BIQA techniques for cross-database validation *i.e.,* when training is performed on LIVE database and testing is performed on the other three databases. The first column in the Table 2 shows the IQA technique used. It can be observed that the proposed methodology using ResNet shows the highest SRCC score when training is performed on LIVE database and testing is performed on the CSIQ and CLIVE database *i.e.,* 0.9518 and 0.4726 respectively, whereas the proposed methodology using VGGNet shows the best performance when training is performed on the LIVE database and testing is performed on the TID2013 database with an SRCC score of 0.9635. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

Table 3 shows the performance of the proposed methodology in comparison with 9 state-of-the-art BIQA techniques for cross-database validation *i.e.,* when training is performed on the CSIQ database and testing is performed on the other three databases. The first column in the Table 3 shows the IQA technique used. It can be observed that the proposed methodology using ResNet shows the highest SRCC score when training is performed on the CSIQ database and testing is performed on the LIVE, TID2013, and CLIVE database *i.e.,* 0.963, 0.9176 and 0.4307 respectively. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

Table 4 shows the performance of the proposed methodology in comparison with 10 state-of-the-art BIQA techniques for cross-database validation *i.e.,* when training is performed on the TID2013 database and testing is performed on the other three databases. The first column in the Table 4 shows the IQA technique used. It can be observed that the proposed methodology using ResNet shows the highest SRCC score when training is performed on the TID2013 database and testing is performed on the LIVE, CSIQ, and CLIVE database *i.e.,* 0.9525, 0.9177 and 0.6098 respectively. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

Table 5 shows the performance of the proposed methodology in comparison with 8 state-of-the-art BIQA techniques for cross-database validation *i.e.,* when training is performed on the CLIVE database and testing is performed on the other three databases. The first column

Table 2
Overall cross database SRCC scores for each NR-IQA technique using LIVE database for training and the TID2013, CSIQ and CLIVE databse for testing.

IQA technique	Testing database		
	CSIQ [86]	TID2013 [85]	CLIVE [87]
BIQI [88]	0.7805	0.8174	0.2621
BLIINDS-II [90]	0.8878	0.9036	0.2618
BRISQUE [91]	0.8993	0.9030	0.4089
DIIVINE [89]	0.8571	0.8579	0.3902
M3 [93]	0.9108	0.9214	0.3064
Pseudo [94]	0.9003	0.8580	0.3955
FSI-RR [95]	0.8870	0.5493	0.3810
MSDD [96]	0.8794	0.9017	0.2871
DIQaM-NR [25]	0.9080	0.8670	–
DIQA [100]	0.9150	0.9220	–
NSSADNN [51]	0.9330	0.9450	–
VGGNet	0.9053	0.9086	0.3778
ResNet	0.9089	0.9127	0.3804
Proposed using VGGNet	0.9472	**0.9635**	0.4013
Proposed using ResNet	**0.9518**	0.9572	**0.4726**

Table 3
Overall cross database SRCC scores for each NR-IQA technique using CSIQ database for training and the LIVE, TID2013 and CLIVE databse for testing.

IQA technique	Testing database		
	LIVE [84]	TID2013 [85]	CLIVE [87]
BIQI [88]	0.4538	0.6987	0.2122
BLIINDS-II [90]	0.9365	0.8015	0.2115
BRISQUE [91]	0.9311	0.8996	0.3589
DIIVINE [89]	0.8475	0.8233	0.3402
M3 [93]	0.9459	0.9061	0.2562
Pseudo [94]	0.8955	0.9001	0.3451
FSI-RR [95]	0.8526	0.8872	0.3310
MSDD [96]	0.9172	0.8785	0.2361
DIQA [100]	0.9260	0.9230	–
VGGNet	0.9194	0.9037	0.3146
ResNet	0.9258	0.9082	0.3404
Proposed using VGGNet	0.9568	0.9104	0.4281
Proposed using ResNet	**0.9634**	**0.9176**	**0.4307**

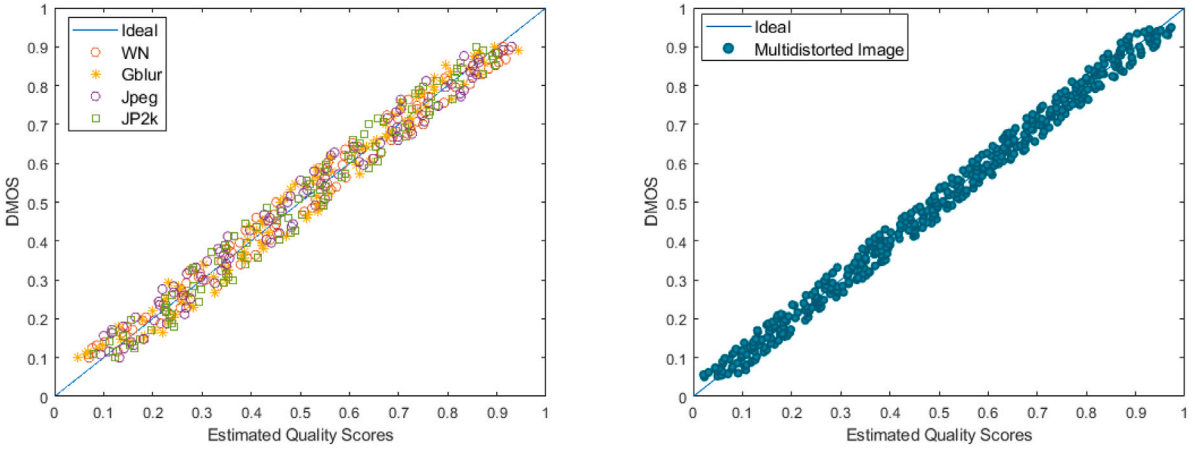

Fig. 5. Scatter plots for the databases of (a) single distorted natural scene images (b) multi distorted natural scene images.

Table 4
Overall cross database SRCC scores for each NR-IQA technique using TID2013 database for training and the LIVE, CSIQ and CLIVE databse for testing.

IQA Technique	Testing database		
	LIVE [84]	CSIQ [86]	CLIVE [87]
BIQI [88]	0.7621	0.8019	0.4026
BLIINDS-II [90]	0.9379	0.8757	0.5313
BRISQUE [91]	0.9278	0.8675	0.4021
DIIVINE [89]	0.8648	0.8491	0.4219
M3 [93]	0.9336	0.8383	0.5524
Pseudo [94]	0.8955	0.8803	0.4463
FSI-RR [95]	0.8421	0.8672	0.5351
MSDD [96]	0.9068	0.8609	0.5309
DIQaM-NR [25]	–	0.7170	–
DIQA [100]	0.9040	0.8770	–
VGGNet	0.9037	0.8425	0.5303
ResNet	0.9162	0.8834	0.5427
Proposed using VGGNet	0.9460	0.9153	0.5432
Proposed using ResNet	**0.9525**	**0.9177**	**0.6098**

Table 5
Overall cross database SRCC scores for each NR-IQA technique using CLIVE database for training and the LIVE, CSIQ and TID2013 databse for testing.

IQA Technique	Testing database		
	LIVE [84]	CSIQ [86]	TID2013 [85]
BIQI [88]	0.3619	0.3127	0.5013
BLIINDS-II [90]	0.3620	0.3125	0.6215
BRISQUE [91]	0.5069	0.4591	0.5123
DIIVINE [89]	0.4892	0.4406	0.5217
M3 [93]	0.4061	0.3179	0.6525
Pseudo [94]	0.4959	0.4441	0.5464
FSI-RR [95]	0.4831	0.4314	0.6353
MSDD [96]	0.3893	0.3362	0.6312
VGGNet	0.7738	0.7672	0.7325
ResNet	0.7864	0.7853	0.7765
Proposed using VGGNet	0.8226	0.7963	0.7840
Proposed using ResNet	**0.8621**	**0.8342**	**0.8168**

in the Table 5 shows the IQA technique used. It can be observed that the proposed methodology using ResNet shows the highest SRCC score when training is performed on the CLIVE database and testing is performed on the LIVE, CSIQ, and TID2013 database *i.e.,* 0.8621, 0.8342 and 0.8168 respectively. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

Table 6
Overall performance comparison of the proposed DeepRPN-BIQA technique in terms of median SRCC, PCC and KROCC for KADID-10k database.

IQA Technique	KADID-10k [101]		
	PCC	SRCC	KROCC
BIQI [88]	0.4600	0.4310	0.299
DIIVINE [89]	0.5320	0.4890	0.3410
BLIINDS-II [90]	0.5590	0.5270	0.375
BRISQUE [91]	0.5540	0.5190	0.368
CORNIA [104]	0.5800	0.5410	0.384
HOSA [105]	0.6530	0.6090	0.438
SSEQ [106]	0.4630	0.4240	0.295
InceptionResNetV2 [101]	0.7340	0.7310	0.546
VGGNet	0.7063	0.7095	0.5037
ResNet	0.7298	0.7266	0.0.532
Proposed using VGGNet	0.7410	0.7560	0. 568
Proposed using ResNet	**0.8460**	**0.8380**	**0.672**

Table 7
Overall performance comparison of the proposed NR-IQA technique with state of the art techniques in terms of median SRCC, PCC and RMSE for screen content databases.

IQA Technique	SIQAD [102]			SCID [103]		
	SRCC	PCC	RMSE	SRCC	PCC	RMSE
BQMS [107]	0.7366	0.7558	5.406	–	–	–
NRLT [108]	0.8202	0.8442	7.5957	0.6454	0.6625	10.6452
PICNN [69]	0.897	0.896	6.790	0.822	0.827	8.031
GFM [62]	0.8735	0.8820	6.7237	0.8759	0.8760	6.8313
TFSR [109]	0.8354	0.8618	7.4910	0.7840	0.8017	8.8041
VGGNet	0.882	0.894	5.927	0.874	0.877	7.892
ResNet	0.891	0.886	5.583	0.887	0.885	6.374
Proposed using VGGNet	0.925	0.913	5.319	0.896	0.891	7.175
Proposed using ResNet	**0.932**	**0.937**	**4.662**	**0.924**	**0.918**	**5.630**

The CLIVE database consists of authentically distorted images. The IQA of the authentically distorted images is the most challenging task since we do not know the type and level of distortion affecting the image. Table 1 shows that when training and testing are performed on authentically distorted images shows good performance. The synthetically distorted images are different from authentically distorted images. Therefore, cross-database validation i.e., training is performed on one database and testing is performed on another database does not achieve good results as shown in Tables 2, 3, 4, and 5. Considering the aforementioned reason the proposed method performs best among the well-established CNN architectures such as ResNet or VGGNet on the CLIVE database. In terms of hit count that shows the number of times, each BIQA technique is ranked among the top three in terms of the highest SRCC score, the proposed methodologies based on ResNet and

VGGNet show the highest hit count of 12, which is much higher than the hit count of 4 obtained by the existing BIQA techniques.

Table 6 shows the performance of the proposed methodology with 8 state-of-the-art BIQA techniques on the KADID-10k database. it can be observed that the proposed methodology with ResNet shows the best performance with the SRCC score of 0.8460 and the proposed methodology with VGGNet is ranked second with the SRCC score of 0.7410 and Inception ResNetV2 is ranked third. The significance of the results on the KADID-10k database is that it is has a large number of images for IQA and is specifically designed for IQA using deep CNNs. The better performance of the proposed methodology shows that visual saliency using deep CNN helps to improve the performance of BIQA techniques when large datasets are under consideration. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

Table 7 shows the performance of the proposed methodology in comparison to 6 state-of-the-art screen content image quality assessment techniques. It can be observed that the proposed technique with ResNet shows the best performance on the SIQAD as well as SCID databases with the SRCC score of 0.9320 and 0.9240 respectively, whereas the proposed technique with VGGNet is ranked second with the SRCC score of 0.9250 and 0.8960 on the SIQAD and SCID databases respectively. PICNN [69] is ranked third with the SRCC score of 0.8970 on the SIQAD and SRCC score of 0.8220 on the SCID database. The proposed DeepRPN-BIQA shows better performance on NSI as well as SCI IQA databases in comparison to state-of-the-art techniques. The results of simple VGGNet and ResNet for the ablation study are compared with the VGGNet and ResNet being used in the proposed methodology. The proposed methodology has shown significant improvement in the performance of both architectures.

The proposed network is trained and tested on each database separately i.e., a separate model is trained for each database for testing for the results in TABLE 1. For TABLE 2 training is performed on LIVE database and testing in performed on CSIQ, CLIVE, TID2013 databases. In TABLE 3, training is performed on CSIQ database and testing is performed on the LIVE, CLIVE and TID2013 databases. For TABLE 4, training is performed on the TID2013 database and testing is performed on the LIVE, CLIVE and CSIQ databases. In TABLE 5, training is performed on CLIVE database and testing is performed on the LIVE, CSIQ and TID2013 databases. For the results shown in TABlE 7 for the screen content databases the model needs to be retrained and testing and training is performed on the screen content images of the respective databases. In the TABLE 6 for the KADID database the model is trained and tested on the images from the KADID database

Fig. 5 shows the scatter plot between the MOS and predicted quality score. It can be observed that the predicted quality score correlates with the perceptual quality of images for all the databases. The proposed DeepRPN-BIQA performs better as compared to other deep CNN-based techniques because it makes use of visual saliency to extract region proposals that are most relevant for picture quality. The region proposals are used to extract ROIs that are most noticeable to HVS in an image. As human observers are the ultimate users of images and visual saliency tries to mimic the HVS therefore, visual saliency will selectively process the important parts of the image. From the perspective of IQA, the distortions present in the image will attract more attention of the human observer and the visual saliency will select regions concerning image quality. Therefore, the perceptual quality of the ROIs extracted using RPNs represents the quality of the whole image, and the proposed DeepRPN-BIQA performs better in comparison to state-of-the-art deep-based BIQA techniques.

5. Conclusion

BIQA is a challenging task due to the absence of reference images. This work proposes an end-to-end DNN based BIQA technique for IQA that makes use of visual saliency to extract ROIs that are most noticeable to the HVS and represent the overall perceptual quality of the image. The proposed methodology suggests proposals based on RPNs to extract ROIs, which are most relevant for IQA. A local quality score for each ROI is computed and the overall image quality score is taken as the average of all the quality scores over all the patches. The proposed methodology shows better performance on synthetically distorted, images taken in real-world conditions using mobile phone cameras and screen content images.

CRediT authorship contribution statement

Mobeen ur Rehman: Performed the simulation and wrote the initial draft of the manuscript. **Imran Fareed Nizami:** Conceived the idea, Analyzed the results, Refined the manuscript for submission. **Muhammad Majid:** Conceived the idea, Analyzed the results.

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] D. Liu, Y. Wang, Z. Chen, Joint foveation-depth just-noticeable-difference model for virtual reality environment, J. Vis. Commun. Image Represent. 56 (2018) 73–82.
[2] D. Ghadiyaram, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (2015) 372–387.
[3] B. Gupta, M. Tiwari, Color retinal image enhancement using luminosity and quantile based contrast enhancement, Multidimens. Syst. Signal Process. 30 (2019) 1829–1837.
[4] X. Kuang, X. Sui, Y. Liu, Q. Chen, G. Gu, Single infrared image enhancement using a deep convolutional neural network, Neurocomputing 332 (2019) 119–128.
[5] E. Jung, N. Yang, D. Cremers, Multi-frame gan: Image enhancement for stereo visual odometry in low light, in: Conference on Robot Learning, 2020, pp. 651–660.
[6] W. Kim, R. Lee, M. Park, S.H. Lee, Low-light image enhancement based on maximal diffusion values, IEEE Access 7 (2019) 129150–129163.
[7] Z. Zhou, Y. Mu, Q.J. Wu, Coverless image steganography using partial-duplicate image retrieval, Soft Comput. 23 (2019) 4927–4938.
[8] Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research, Neurocomputing 335 (2019) 299–326.
[9] A.A. Abd EL-Latif, B. Abd-El-Atty, S.E. Venegas-Andraca, A novel image steganography technique based on quantum substitution boxes, Opt. Laser Technol. 116 (2019) 92–102.
[10] Z. Qu, Z. Cheng, W. Liu, X. Wang, A novel quantum image steganography algorithm based on exploiting modification direction, Multimedia Tools Appl. 78 (2019) 7981–8001.
[11] No-reference image quality assessment using bag-of-features with feature selection, Multimedia Tools Appl. (2020) 1–26.
[12] New feature selection algorithms for no-reference image quality assessment, Appl. Intell. 48 (2018) 3482–3501.
[13] W. Sun, Q. Liao, J.H. Xue, F. Zhou, Spsim: A superpixel-based similarity index for full-reference image quality assessment, IEEE Trans. Image Process. 27 (2018) 4232–4244.
[14] A. Saha, Q.J. Wu, Full-reference image quality assessment by combining global and local distortion measures, Signal Process. 128 (2016) 186–197.
[15] Z. Tang, Y. Zheng, K. Gu, K. Liao, W. Wang, M. Yu, Full-reference image quality assessment by combining features in spatial and frequency domains, IEEE Trans. Broadcast. 65 (2018) 138–151.
[16] Y. Wen, Y. Li, X. Zhang, W. Shi, L. Wang, J. Chen, A weighted full-reference image quality assessment based on visual saliency, J. Vis. Commun. Image Represent. 43 (2017) 119–126.
[17] Z. Shi, J. Zhang, Q. Cao, K. Pang, T. Luo, Full-reference image quality assessment based on image segmentation with edge feature, Signal Process. 145 (2018) 99–105.

[18] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimed. 20 (2017) 379–391.

[19] Y. Zhang, Reduced-reference image quality assessment based on distortion families of local perceived sharpness, Signal Process., Image Commun. 55 (2017) 130–145.

[20] D. Liu, F. Li, H. Song, Regularity of spectral residual for reduced reference image quality assessment, IET Image Process. 11 (2017) 1135–1141.

[21] K. Rahul, Fqi: feature-based reduced-reference image quality assessment method for screen content images, IET Image Process. 13 (2019) 1170–1180.

[22] E. Kalatehjari, F. Yaghmaee, A new reduced-reference image quality assessment based on the svd signal projection, Multimedia Tools Appl. 77 (2018) 25053–25076.

[23] Y. Fang, J. Liu, Y. Zhang, W. Lin, Z. Guo, Reduced-reference quality assessment of image super-resolution by energy change and texture variation, J. Vis. Commun. Image Represent. 60 (2019) 140–148.

[24] Y. Shi, W. Guo, Y. Niu, J. Zhan, No-reference stereoscopic image quality assessment using a multi-task cnn and registered distortion representation, Pattern Recognit. 100 (2020) 107168.

[25] S. Bosse, D. Maniry, K.R. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (2017) 206–219.

[26] H. Talebi, P. Milanfar, Nima: Neural image assessment, IEEE Trans. Image Process. 27 (2018) 3998–4011.

[27] T.J. Liu, K.H. Liu, No-reference image quality assessment by wide-perceptual-domain scorer ensemble method, IEEE Trans. Image Process. 27 (2017) 1138–1151.

[28] C. Fan, Y. Zhang, L. Feng, Q. Jiang, No reference image quality assessment based on multi-expert convolutional neural networks, IEEE Access 6 (2018) 8934–8943.

[29] S. Jia, Y. Zhang, Saliency-based deep convolutional neural network for no-reference image quality assessment, Multimedia Tools Appl. 77 (2018) 14859–14872.

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

[31] Y. Liu, J. Yang, Q. Meng, Z. Lv, Z. Song, Z. Gao, Stereoscopic image quality assessment method based on binocular combination saliency model, Signal Process. 125 (2016) 237–248.

[32] X. Wang, L. Ma, S. Kwong, Y. Zhou, Quaternion representation based visual saliency for stereoscopic image quality assessment, Signal Process. 145 (2018) 202–213.

[33] W. Zhang, Y. Tian, X. Zha, H. Liu, Benchmarking state-of-the-art visual saliency models for image quality assessment, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 1090–1094.

[34] G. Zhai, Y. Zhu, X. Min, Comparative perceptual assessment of visual signals using free energy features, IEEE Trans. Multimed. (2020).

[35] Q. Lu, G. Zhai, W. Zhu, Y. Zhu, X. Min, X.P. Zhang, H. Yang, Automatic region selection for objective sharpness assessment of mobile device photos, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 106–110.

[36] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3575–3585.

[37] S. Yang, Q. Jiang, W. Lin, Y. Wang, Sgdnet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1383–1391.

[38] W. Xue, L. Zhang, X. Mou, Learning without human scores for blind image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 995–1002.

[39] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1733–1740.

[40] S. Bosse, D. Maniry, T. Wiegand, W. Samek, A deep neural network for image quality assessment, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3773–3777.

[41] S. Bianco, L. Celona, P. Napoletano, R. Schettini, On the use of deep learning for blind image quality assessment, Signal Image Video Process. 12 (2018) 355–362.

[42] L. Kang, P. Ye, Y. Li, D. Doermann, Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 2791–2795.

[43] A color intensity invariant low-level feature optimization framework for image quality assessment, Signal Image Video Process. 10 (2016) 1169–1176.

[44] D. Ghadiyaram, Blind image quality assessment on real distorted images using deep belief nets, in: 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2014, pp. 946–950.

[45] D. Ghadiyaram, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (2015) 372–387.

[46] A. Krizhevsky, I. Sutskever, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[47] H. Zeng, L. Zhang, A.C. Bovik, A probabilistic quality representation approach to deep blind image quality prediction, 2017, arXiv preprint arXiv:1708.08190.

[48] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[49] L. Mai, H. Jin, F. Liu, Composition-preserving deep photo aesthetics assessment, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 497–506.

[50] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[51] B. Yan, B. Bare, W. Tan, Naturalness-aware deep no-reference image quality assessment, IEEE Trans. Multimed. (2019).

[52] S. Bianco, L. Celona, P. Napoletano, R. Schettini, On the use of deep learning for blind image quality assessment, Signal Image Video Process. 12 (2018) 355–362.

[53] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, Blind image quality assessment using a deep bilinear convolutional neural network, IEEE Trans. Circuits Syst. Video Technol. (2018).

[54] F. Gao, J. Yu, S. Zhu, Q. Huang, Q. Tian, Blind image quality prediction by exploiting multi-level deep representations, Pattern Recognit. 81 (2018) 432–442.

[55] Z. Zhang, H. Wang, S. Liu, Deep activation pooling for blind image quality assessment, Appl. Sci. 8 (2018) 478.

[56] Research on the parallelization of image quality analysis algorithm based on deep learning, J. Vis. Commun. Image Represent. (2019) 102709.

[57] O. Alaql, C.C. Lu, No-reference image quality metric based on multiple deep belief networks, IET Image Process. 13 (2019) 1321–1327.

[58] J. Wu, J. Ma, F. Liang, W. Dong, G. Shi, W. Lin, End-to-end blind image quality prediction with cascaded deep neural network, IEEE Trans. Image Process. (2020).

[59] L. Zheng, L. Shen, J. Chen, P. An, J. Luo, No-reference quality assessment for screen content images based on hybrid region features fusion, IEEE Trans. Multimed. 21 (2019) 2057–2070.

[60] N. Lu, G. Li, Blind quality assessment for screen content images by orientation selectivity mechanism, Signal Process. 145 (2018) 225–232.

[61] Y. Zhang, Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation, IEEE Trans. Image Process. 27 (2018) 5113–5128.

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

[63] X. Min, G. Zhai, K. Gu, Y. Liu, X. Yang, Blind image quality estimation via distortion aggravation, IEEE Trans. Broadcast. 64 (2018) 508–517.

[64] W. Wan, J. Wang, J. Li, J. Sun, H. Zhang, J. Liu, Hybrid jnd model-guided watermarking method for screen content images, Multimedia Tools Appl. 79 (2020) 4907–4930.

[65] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.W. Wu, T. Luo, Local and global feature learning for blind quality evaluation of screen content and natural scene images, IEEE Trans. Image Process. 27 (2018) 2086–2095.

[66] Y. Fu, H. Zeng, L. Ma, Z. Ni, J. Zhu, K.K. Ma, Screen content image quality assessment using multi-scale difference of gaussian, IEEE Trans. Circuits Syst. Video Technol. 28 (2018) 2428–2432.

[67] Y. Bai, M. Yu, Q. Jiang, G. Jiang, Z. Zhu, Learning content-specific codebooks for blind quality assessment of screen content images, Signal Process. 161 (2019) 248–258.

[68] Z. Xia, K. Gu, S. Wang, H. Liu, S. Kwong, Toward accurate quality estimation of screen content pictures with very sparse reference information, IEEE Trans. Ind. Electron. 67 (2019) 2251–2261.

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

[70] X. Jiang, L. Shen, Q. Ding, L. Zheng, P. An, Screen content image quality assessment based on convolutional neural networks, J. Vis. Commun. Image Represent. 67 (2020) 102745.

[71] W. Zhou, L. Yu, Y. Zhou, W. Qiu, J. Xiang, Z. Zhai, Blind screen content image quality measurement based on sparse feature learning, Signal Image Video Process. 13 (2019) 525–530.

[72] X. Min, K. Gu, G. Zhai, M. Hu, X. Yang, Saliency-induced reduced-reference quality index for natural scene and screen content images, Signal Process. 145 (2018) 127–136.

[73] M. Banitalebi-Dehkordi, M. Khademi, A. Ebrahimi-Moghadam, H. Hadizadeh, An image quality assessment algorithm based on saliency and sparsity, Multimedia Tools Appl. 78 (2019) 11507–11526.

[74] X. Wang, X. Liang, B. Yang, F.W. Li, No-reference synthetic image quality assessment with convolutional neural network and local image saliency, Comput. Vis. Media 5 (2019) 193–208.

[75] H. Jia, L. Zhang, T. Wang, Contrast and visual saliency similarity-induced index for assessing image quality, IEEE Access 6 (2018) 65885–65893.

[76] M. Layek, A. Uddin, T.P. Le, T. Chung, E.N. Huh, et al., Center-emphasized visual saliency and a contrast-based full reference image quality index, Symmetry 11 (2019) 296.
[77] M. Oszust, No-reference quality assessment of noisy images with local features and visual saliency models, Inform. Sci. 482 (2019) 334–349.
[78] W. Zhang, W. Zou, F. Yang, Linking visual saliency deviation to image quality degradation: A saliency deviation-based image quality index, Signal Process., Image Commun. 75 (2019) 168–177.
[79] W. Zhang, A. Borji, Z. Wang, P. Le Callet, H. Liu, The application of visual saliency models in objective image quality assessment: A statistical evaluation, IEEE Trans. Neural Netw. Learn. Syst. 27 (2015) 1266–1278.
[80] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, W. Zhang, Saliency-guided quality assessment of screen content images, IEEE Trans. Multimed. 18 (2016) 1098–1110.
[81] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252.
[82] M. Everingham, L. Van Gool, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.
[83] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[84] A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (2006) 3440–3451.
[85] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Peculiarities, results and perspectives, Signal Process., Image Commun. 30 (2015) (2013) 57–77.
[86] Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (2010) 011006.
[87] D. Ghadiyaram, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (2016) 372–387.
[88] A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (2010) 513–516.
[89] Blind image quality assessment: From natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (2011) 3350–3364.
[90] Blind image quality assessment: A natural scene statistics approach in the dct domain, IEEE Trans. Image Process. 21 (2012) 3339–3352.
[91] A. Mittal, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (2012) 4695–4708.
[92] Y. Zhang, J. Wu, X. Xie, L. Li, G. Shi, Blind image quality assessment with improved natural scene statistics model, Digit. Signal Process. 57 (2016) 56–65.
[93] W. Xue, X. Mou, L. Zhang, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, IEEE Trans. Image Process. 23 (2014) 4850–4862.
[94] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, C.W. Chen, Blind quality assessment based on pseudo-reference image, IEEE Trans. Multimed. 20 (2018) 2049–2062.
[95] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimed. 20 (2018) 379–391.
[96] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, H. Sun, Optimizing multistage discriminative dictionaries for blind image quality assessment, IEEE Trans. Multimed. 20 (2018) 2035–2048.
[97] Y. Lv, G. Jiang, M. Yu, H. Xu, F. Shao, S. Liu, Difference of gaussian statistical features based blind image quality assessment: A deep learning approach, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 2344–2348.
[98] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE J. Sel. Top. Sign. Proces. 11 (2016) 206–220.
[99] 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 (2017) 1202–1213.
[100] J. Kim, Deep cnn-based blind image quality predictor, IEEE Trans. Neural Netw. Learn. Syst. 30 (2018) 11–24.
[101] H. Lin, V. Hosu, D. Saupe, Kadid-10k: A large-scale artificially distorted iqa database, in: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2019, pp. 1–3.
[102] H. Yang, Y. Fang, W. Lin, Perceptual quality assessment of screen content images, IEEE Trans. Image Process. 24 (2015) 4408–4421.
[103] 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 (2017) 4818–4831.
[104] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 1098–1105.
[105] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind image quality assessment based on high order statistics aggregation, IEEE Trans. Image Process. 25 (2016) 4444–4457.
[106] L. Liu, B. Liu, H. Huang, No-reference image quality assessment based on spatial and spectral entropies, Signal Process., Image Commun. 29 (2014) 856–863.
[107] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, Learning a blind quality evaluation engine of screen content images, Neurocomputing 196 (2016) 140–149.
[108] 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 (2017) 1600–1610.
[109] J. Yang, J. Liu, B. Jiang, W. Lu, No reference quality evaluation for screen content images considering texture feature based on sparse representation, Signal Process. 153 (2018) 336–347.