INTRODUCTION
With the rapid development of social media and digital imaging equipment generates an increasing demand in terms of image acquisition, transmission, compression and so on. In the computer vision, image quality assessment (IQA) as a critical metric of video and image processing technology has become increasingly prominent. The accurate evaluation of the image quality and further optimization of the image-processing technology has highly practical value.

In general, IQA can be divided into subjective IQA methods and objective IQA methods. Since the subjective IQA methods adopt continuous double incentive quality assessment and need to experiment on test images repeatedly, but they are time-consuming, expensive, difficult to operate, and not suitable for real-time processing. Therefore, to develop effective objective IQA methods has become the focus of image quality research.

According to the availability of a reference image, objective IQA metrics can be classified as FR-IQA (full reference), RR-IQA (reduced reference) and NR-IQA (no reference). A variety of FR-IQA methods [1-4, 18] have been proposed to achieve high correlation with human perception. However, in most visual applications, the availability of the reference image is impossible or costs too much. Only the distorted image is given. So it is necessary to develop NR-IQA to automatically evaluate the image quality. NR-IQA is the most challenging category of objective IQA problem. Due to without ideal image, it is important for NR-IQA task to capture good quality features. NR-IQA research is more likely to reveal the principles of human visual perception.

Convolutional neural network (CNN) is a deep learning model based on supervised learning. In recent years, CNN has been responsible for major breakthrough in various computer vision tasks and attracted the attention of many researchers. CNN can learn good and complex approximations of the functional relationships between input and output. Its deep non-linear network structure shows the strong ability of feature representation. However, few studies have considered applying CNNs to the image quality assessment. Therefore we alternatively investigate the problem of NF-IQA based on the CNN framework.

No-reference Quality Assessment Based on Convolutional Neural Network

Lin Li
School of Computer Science and Technology, Wuhan University of Technology, Wuhan, Hubei, China
School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China

Shengsheng Yu
School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China

ABSTRACT: How to extract image features highly correlated with visual perception is still a challenging task in No-reference image quality assessment. We aim to test the feasibility of introducing deep learning into quality assessment algorithm. In this paper, we developed a novel convolution neural network IQF-CNN that is able to learn more discriminative image quality features, and applied the learned features to predict image quality. We also used the local luminance coefficients normalization and dropout technology to improve the IQF-CNN learning ability. The proposed method can accurately measure the five common image distortions on standard benchmarks, and also improvements over the previous state of the art NR-IQA works have been demonstrated.

Keywords: convolutional neural network; image quality assessment; feature learning; normalization
By improving the traditional convolutional neural network, we proposed a novel network IQF-CNN (CNN based on Image Quality Features) to automatically learn discriminative quality feature. By using deep learning image feature, we are able to accurately predict the quality of the distorted image, which is well correlated with subjective judgment of quality. Experiments on the LIVE and TID2008 standard dataset show that proposed method can successfully predict image quality across a range of distortion types.

In the following, we first review the recent published literature in Section 2. Section 3 presents the convolutional neural network IQF-CNN and Section 4 describes the learning process and algorithm implementation in detail. The performance evaluation and the results are given in Section 5. The paper is concluded in Section 6.

2 RELATED WORK

The basic idea of the NF-IQA algorithm is to first analyze the image, extract the distortion features, and then evaluate the image quality based on quantitative measures of these various types of distortion features.

In existing no-reference image quality assessment, there are two kinds of image feature extraction methods: natural scene statistics (NSS) based features and machine learning based features.

NSS are extracted in image transformation domain, using local descriptors to obtain statistical characteristics. Image distortion will have an impact on feature statistical model. Bovik et al. [5] first introduced NSS based features into NF-IQA. They used a NSS model for blindly measuring the quality of JPEG2000 compressed images. In [6], an algorithm based on DCT coefficients to extract image statistical features for JPEG distortion was proposed. DIIVINE [7] used wavelet subband decomposition to extract NSS features, and mapped these features to the image quality score through regression model. But DIIVINE is difficult to use in real time application because a large number of image features need to be computed. In [8], the image was firstly divided into equal size blocks, and then the local DCT coefficients were calculated in each block. The DCT coefficients were applied to the generalized Gaussian model for feature extraction. Finally Bayesian model was used to predict image quality. Mittal et al. [9] proposed NF-IQA model named BRISQUE, which is operated in the spatial domain and uses scene statistics of locally normalized coefficients to quantify the distorted image. However, NSS feature as handcrafted feature is not effective enough for some types of distortion (especially for hybrid distortion).

Machine learning explores the study and construction of algorithms that can learn from and make predictions on the data. In recent years, Ye et al. [10] proposed unsupervised feature learning method, using a clustering algorithm on Gabor feature vectors to create the visual codebook; the code words are used as the input to a regression model for the estimation of quality scores. However, the use of Gabor features and a large codebook makes this approach highly computationally expensive. To improve their previous work, they developed another method called CORNIA [11]. The method uses a codebook based approach to learn features automatically. The codebook is constructed by K-means clustering on local features extracted from unlabeled training images. Test image is encoded in the learned codebook, to extract its features. CORNIA has achieved higher prediction precision in each dataset.

Based on above discussions, we can know that machine learning based approach is a better candidate for NR-IQA algorithms. But the above methods do not take full advantage of HVS characteristics and human cognitive processing to extract quality features, and affect the accuracy of image quality evaluation. Instead of extracting handcrafted features, one of the promises of deep learning is to extract hierarchical feature learned by unsupervised or supervised algorithms. However, little research attempted to apply deep learning for NF-IQA.

Our study is focused on introducing deep learning into the active feature learning and putting forward an image quality evaluation framework based on convolutional neural network. In this framework, we sample patches from the normalized image, and take image patches as input to the network model to learn discriminant feature. The output of this neural network is a value corresponding to the image quality level. We integrated image feature learning and quality assessment into the IQF-CNN framework.

3 IQF-CNN ARCHITECTURE

The network architecture used in our paper is a multi-layer deep structure presented in Figure 1. IQF-CNN is comprised of two convolutional layers intertwined with the subsampling layer, and then followed by two fully connected layers. The input is 32×32 fixed-size image patch. The first layer is a convolutional layer of 8 filters with the size of 3×3. The input convolves with filters at the stride of one pixel, resulted in a map with the size of 30×30 and 8 channels. The following layer is a pooling layer with the size of 2×2, which is to maximally pool the feature maps of the first convolutional layer without overlapping. The second convolutional layer has 32 kernels each with the size of 3×3, which can produce 32 response maps. The following pooling layer reduces each feature map to a maximum and minimum value and produces 64 outputs. The 64 values are set as the input to the fully connected layers. The last two fully connected layers are of 128 and 512 nodes respectively.
In IQF-CNN model, the input image is performed with normalization preprocessing. Ruderman [12] observed that MSCN (mean subtraction and contrast normalization) coefficients strongly tend towards a unit normal Gaussian characteristic for natural images. Figure 2 shows distributions of MSCN coefficients for a natural ideal image and for its various distorted versions. Normalized luminance values have similar statistical properties. Such an operation can be used to model the contrast-gain masking process in early human vision. So we utilize improved local luminance coefficient normalization in our model. The specific process is as follows. First, convert the image from the RGB to Lab color space because the Lab color is closer to human vision. Its L-component closely matches human brightness perception. Second, divide the image into non-overlapping patches with size $N \times N$ (here $N = 3$). Suppose the luminance value of a pixel at location $(i, j)$ is $L(i, j)$, the normalized value $\hat{L}(i, j)$ is computed as follows:

$$\hat{L}(i, j) = \frac{L(i, j) - \mu(i, j)}{\sigma(i, j) + 1}$$  \hspace{1cm} (1)

$$\mu(i, j) = \frac{1}{(2N+1)^2} \sum_{m=-N}^{m=N} \sum_{n=-N}^{n=N} L(i + m, j + n)$$  \hspace{1cm} (2)

$$\sigma(i, j) = \frac{1}{(2N+1)^2} \sum_{m=-N}^{m=N} \sum_{n=-N}^{n=N} [L(i + m, j + n) - \mu(i, j)]^2$$  \hspace{1cm} (3)

After being normalized, divide the image into patches with the size of $32 \times 32$, and set the patches as the input to the IQF-CNN. All the convolution layers are followed with pooling layers for reducing the dimensions of the filtering response. In particular, each feature map of the second convolutional layer will be sampled with two extreme valves, which is similar to CORNIA algorithm. In traditional convolution neural networks, pooling is performed in a small neighborhood (for example: $3 \times 3$ cell). By integrating the features in a small neighborhood, a new representative feature can be obtained to maintain the invariance including rotation, translation, scaling and so on. This operation is very helpful for object detection and recognition, but is not suitable for image quality assessment tasks. Image distortion is usually locally homogenous, so the same level of distortion occurs in the entire image patch. For pooling operation in IQA algorithm, it is not necessary to consider the spatial location information. To reduce computational cost, our network model directly subsample the whole feature map into one maximum and one minimum value.

Neurons in the last two fully connected layers have full connections to all activations in the previous layer. We use the ReLUs (Rectified Linear Units) to replace the traditional sigmoid function as the activation function. Key advantages of ReLUs are sparse activation and biological plausibility, which make it closer to the human brain activation model. In addition, compared to other functions the usage of ReLUs is preferable especially for deep neural network, because it results in the neural network training times faster [13], without making a significant difference to generalization accuracy. It is worth to note that we use a linear activation function rather than ReLUs function in the first two layers of IQA-CNN. Because ReLUs allows only non-negative signal, we selected identity transform to avoid blocking the most negative value generated by min pooling.
4 NR-IQA ALGORITHM

Similar quality images are of the same or similar hidden image quality features. According to such a hypothesis, the machine learning based on NR-IQA method obtain distinguishing quality features by automatic learning, and directly use learned features to estimate the quality score for the distorted image without corresponding ideal versions.

Here, we adopt this strategy for developing image quality assessment metric IQF-CNN. Our algorithm is distinct in two ways: (1) we use luminance coefficient normalization as pre-processing to help extract features closely related to the degradations; (2) we leverage the trained multilayer IQF-CNN to learn hidden quality features used for quality estimation. The flowchart of the proposed method is illustrated in Figure 3.

The proposed algorithm is directly applied to process the image domain, which can realize the discriminant feature learning and image quality prediction in a general network framework. The algorithm can be summarized as the following four key steps and two main processing stages.

Step 1: preprocess the selective images in the dataset;
Step 2: randomly select a sample patch with the size of $32 \times 32$ form the normalized images and generate training samples;
Step 3: use supervised learning method to training IQF-CNN model and optimize model parameters;
Step 4: give a test image, image quality assessment is provided directly by IQF-CNN model.

The algorithm mainly consists of training and testing stages.

The first stage is the training process. We trained our network on $32 \times 32$ patches instead of the whole image. One reason is that training samples are gathered from the LIVE dataset, and distortion of the images on the LIVE dataset is often locally homogeneous. Another reason is that we have a much larger number of training samples by using image patches. According to the image ground truth score, we assigned a quality score to each patch. Let $x_n$ denote the input patch; $y_n$ denote the patch quality score.

$(x_n, y_n)$ is a training sample for IQA-CNN. $f(\omega, x_n)$ is prediction function learned by network. We applied object function similar to the regression function in Support Vector Regression (SVR) $J(\omega; x_n, y_n) = \frac{1}{N} \sum_{n=1}^{N} \|f(\omega, x_n) - y_n\|^2$, and use stochastic gradient descent and backpropagation method to minimize object function to obtain network parameter. The number of SGD iterations is 45 in training process. In the fully connected layer, the probability that a hidden node being set to zero is 0.5. In order to guarantee that the model parameters are the highest linear correlation coefficients, we also randomly selected six reference images and their distorted images (where the distortions are shared by LIVE dataset and TID2008 dataset) from TID2008 database as validation set and then executed cross validation to obtain the optimal network parameter estimation.

The second stage is the testing. Given a test image $i$, we first performed luminance coefficient normalization, and then sample non-overlapping patches from the normalized image. By using the trained IQF-CN, we averaged the predicted patch scores to obtain the visual quality score of the image.

5 EXPERIMENTS

5.1 Dataset

To evaluate the accuracy and robustness of the proposed IQF-CNN algorithm, during the training stage, we have conducted experiments on standard benchmarks, including the datasets LIVE [14] and TID2008 [15].

LIVE: contain 982 distorted images which are derived from 29 reference images. There are five types of the distortion: JP2k compression (JP2K), JPEG compression (JPEG), White Gaussian (WN), Gaussian blur (BLUR) and Fast Fading (FF). Differential Mean Opinion Scores (DMOS) is a subjective measurement of the image quality. Each of the distorted images has an associated DMOS result, roughly in the range of $[0, 100]$. Higher DMOS indicates lower quality.

TID2008: A total of 1700 distorted images derived from 25 reference images contain 17 different types of
distortion. Each image quality result is provided by a Mean Opinion Score (MOS) value in the range of [0, 9]. Higher MOS indicates higher quality.

With the calculation of the correlation coefficient between the objective and subjective quality values, the image quality algorithm can be objectively evaluated. Two quantitative indexes are mainly used to evaluate the performance of our algorithm:

Linear Correlation Coefficient (LCC): measures the linear dependence between the objective values and subjective values. The bigger the LCC is, the better the performance of the algorithm is.

\[
LCC = \frac{\sqrt{N} \sum_{i=1}^{N} y_i \sum_{j=1}^{N} \hat{y}_j - \left( \sum_{i=1}^{N} y_i \right) \left( \sum_{j=1}^{N} \hat{y}_j \right)}{\sqrt{N} \sum_{i=1}^{N} y_i^2 - \left( \sum_{i=1}^{N} y_i \right)^2 \times \sqrt{N} \sum_{i=1}^{N} \hat{y}_i^2 - \left( \sum_{j=1}^{N} \hat{y}_j \right)^2}
\]  

(4)

Where \( N \) is the number of samples, \( y \) is the subjective evaluation value (such as DMOS), and \( \hat{y} \) is the objective evaluation value.

Spearman’s Rank Order Correlation Coefficient (SROCC): measures the prediction monotonicity of an IQA metric. The closer the value of SROCC is to 1, the more accurate the algorithm will be.

\[
SROCC = 1 - \frac{6 \sum D^2}{N(N^2-1)}
\]  

(5)

Where \( D = |y - \hat{y}| \) is the degrade difference.

5.2 Network parameter analysis

Some network parameters are involved in the design of IQF-CNN network. This section analyzes the influence of the different settings of these parameters on the performance of the network.

We first analyzed the influence of the different sizes of the input image. The whole image quality evaluation is based on the averaging of the sample patches, so it is necessary to analyze the effect of sampling strategy on the network performance. Table 1 shows that the network performance is only slightly improved with the increase of the patch size from 16 × 16 to 48 × 48. However, the large size of the patch will not only lead to more computing time, but also reduce the spatial quality of the resolution. Therefore, when selecting the size of the input image, the minimum patch that can produce a better performance. So we choose the patch with the size of 32 × 32 for IQF-CNN network.

Table 1. SROCC and LCC with different input sizes.

| SIZE | 16  | 24  | 32  | 40  | 48  |
|------|-----|-----|-----|-----|-----|
| SROCC| 0.946 | 0.947 | 0.949 | 0.949 | 0.950 |
| LCC  | 0.944 | 0.946 | 0.948 | 0.949 | 0.949 |

In addition, we examined the kernel size involved in the IQF-CNN design. We fixed the network structure, and the network is trained and tested with using different size of convolution kernel. Table 2 shows the performance of the network with different convolution kernel size. It can be seen that the IQF-CNN network is not particularly sensitive to the convolution kernel size.

Table 2. SROCC and LCC with different size of convolution kernel.

| SIZE | 3×3 | 5×5 | 7×7 |
|------|-----|-----|-----|
| SROCC| 0.949 | 0.947 | 0.946 |
| LCC  | 0.948 | 0.947 | 0.945 |

We further studied how to choose the appropriate number of kernels to achieve the highest accuracy. Figure 4 shows that the performance of the quality evaluation varies with the number of network convolution kernels. The number of convolution kernels affects the performance of the network to a large extent. In general, using more convolution kernel will get better performance. Under the condition that the structure of IQF-CNN network is not changed, the first layer kernels number remains the same and the second convolutional layer has different number of kernels. When the total number of network kernel exceeds 40, the SROCC/LCC will not be increased and even slightly reduced. Therefore, in our network, we selected the second convolutional layer which has 32 kernels.

Figure 4. SROCC and LCC with different number of convolution kernels.

5.3 Evaluation on LIVE

Table 3 and Table 4 show the comparison of the algorithms and other NR-IQA algorithms which can achieve better performance. It can be seen that our method can better evaluate five types of distortion, especially for JPEG, JP2K and BLUR distortion. From the perspective of comprehensive evaluation, our algorithms based on the IQF-CNN are better than some classical no-reference image quality assessment algorithms IQF-CNN is a deep neural network containing two convolution layers and two fully connected layers, and trained by a great deal of experimental data, so
our algorithm can get better image quality evaluation. There is no further study on the effect of changing the size of training set on evaluation results. If it needs to increase the number of training data samples, we can apply the method proposed by Krizhevsky[16] in 2012, which indicated that the data capacity can be expanded by image transformation and change of the intensity of the RGB channel.

Table 3. SROCC of different methods on LIVE.

| Method       | JP2K | JPEG | WN   | BLUR | FF   | ALL   |
|--------------|------|------|------|------|------|-------|
| DIIVINE[7]   | 0.913| 0.910| 0.984| 0.921| 0.863| 0.916 |
| BLIINDS-II[8]| 0.929| 0.900| 0.946| 0.923| 0.889| 0.921 |
| BRISQUE[9]   | 0.914| 0.913| 0.979| 0.951| 0.877| 0.932 |
| NSS-TS[17]   | 0.931| 0.915| 0.971| 0.939| 0.935| 0.930 |
| CORNIA[11]   | 0.924| 0.939| 0.969| 0.959| 0.906| 0.941 |
| Our method   | 0.932| 0.947| 0.976| 0.961| 0.913| 0.949 |

Table 4. SROCC of different methods on LIVE.

| Method       | JP2K | JPEG | WN   | BLUR | FF   | ALL   |
|--------------|------|------|------|------|------|-------|
| DIIVINE[7]   | 0.922| 0.921| 0.988| 0.923| 0.888| 0.917 |
| BLIINDS-II[8]| 0.934| 0.920| 0.936| 0.927| 0.896| 0.908 |
| BRISQUE[9]   | 0.922| 0.935| 0.966| 0.945| 0.903| 0.924 |
| NSS-TS[17]   | 0.947| 0.933| 0.963| 0.950| 0.942| 0.926 |
| CORNIA[11]   | 0.949| 0.964| 0.977| 0.954| 0.917| 0.941 |
| Our method   | 0.952| 0.971| 0.984| 0.953| 0.932| 0.948 |

Figure 5. Quality assessment results on the LIVE.

Figure 6. Quality assessment results on the TID2008.

6 CONCLUSION

In this paper, deep learning is introduced into the automatic learning of image quality features. An improved convolutional neural network IQF-CNN is proposed and applied to the quality evaluation of no reference image. In the preprocessing stage, the image is normalized by the local luminance coefficient, and the distinguishing feature is extracted from the transform domain. Then the patch and patch’s quality score are used to train the IQF-CNN network to get the objective quality evaluation model. The experimental results show that the proposed algorithm can accurately evaluate several kinds of image distortions; the overall performance is better than other classical evaluation methods. At the same time, the influence of various network parameters on image quality evaluation is analyzed, and verified on the LIVE dataset. Our work provides research reference for the further study of image quality assessment.

REFERENCES

[1] Z. Wang, A. Bovik, H. Sheikh & E. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600–612.
[2] H.R. Sheikh, M.F. Sabir & A.C. Bovik. 2006. A statistical evaluation of recent full reference image
quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11): 3441–3451.

[3] L. Zhang, D. Zhang, X. Mou & D. Zhang. 2011. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8): 2378–2386.

[4] W. Xue, L. Zhang, X. Mou & A. Bovik. 2014. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(8): 684–695.

[5] H.R. Sheikh, A.C. Bovik & L. Cormack. 2005. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11): 1918–1927.

[6] T. Brandão & M.P. Queluz. 2008. No-reference image quality assessment based on DCT domain statistics. *Signal Process*, 88(4): 822–833.

[7] M. Anush Krishna & B. Alan Conrad. 2011. Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12): 3350-3364.

[8] M.A. Saad & A.C. Bovik, C. Christophe. 2012. Blind image quality assessment: a natural scene statistics approach in the DCT domain. *Image Processing IEEE Transactions*, 21(8): 3339-3352.

[9] A. Mittal, A. Moorthy & A. Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695-4708.

[10] P. Ye, D. Doerrmann. 2012. No-reference image quality assessment using visual codebooks. *IEEE Transactions on Image Processing*, 21(7): 3129-38.

[11] P. Ye. 2012. Unsupervised feature learning framework for no-reference image quality assessment. *IEEE Conference on Computer Vision and Pattern Recognition*, 157(10): 1098-1105.

[12] D.L. Ruderman. 2010. The statistics of natural images. *Network Computation in Neural Systems*, 5(4): 517-548.

[13] A. Krizhevsky, I. Sutskever & G.E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25(2): 1097–1105.

[14] H.R. Sheikh, Z. Wang, L. Cormack & A.C. Bovik. LIVE image quality assessment database release 2. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli & F. Battisti. 2009. TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radio Electronics*, 10: 30-45.

[15] A. Krizhevsky, I. Sutskever & G.E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NIPS)*, 25(2): 1097-1105.

[16] Gao X., Gao F., Tao D. & Li X. 2013. Universal blind image quality assessment statistics and multiple kernel learning. *IEEE Transactions on Neural Network and Learning System*, 12(24): 2013-2026.

[17] Z. Wang & Q. Li. 2011. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5): 1185–1198.