International Journal of Electrical and Computer Engineering (IJECE)
Vol. 9, No. 6, December 2019, pp. 5604~5614
ISSN: 2088-8708, DOI: 10.11591/ijece.v9i6.pp5604-5614

Improve of contrast-distorted image quality assessment based on convolutional neural networks

Ismail Taha Ahmed[1], Chen Soong Der[2], Norziana Jamil[3], Mohamad Afendee Mohamed[4]
[1]College of Computer Sciences and Information Technology, University of Anbar, Iraq
[2]College of Graduate Studies, Universiti Tenaga Nasional, Malaysia
[3]Center of Information and Network Security, Universiti Tenaga Nasional, Malaysia
[4]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Malaysia

Article Info

Article history:

Received Apr 5, 2019
Revised Jul 19, 2019
Accepted Jul 27, 2019

Keywords:

CDI
CNN
NR-IQA-CDI
NR-IQA-CDI-NonPreCNN

ABSTRACT

Many image quality assessment algorithms (IQAs) have been developed during the past decade. However, most of them are designed for images distorted by compression, noise and blurring. There are very few IQAs designed specifically for Contrast Distorted Images (CDI), e.g. Reduced-reference Image Quality Metric for Contrast-changed images (RIQMC) and NR-IQA for Contrast-Distorted Images (NR-IQA-CDI). The existing NR-IQA-CDI relies on features designed by human or handcrafted features because considerable level of skill, domain expertise and efforts are required to design good handcrafted features. Recently, there is great advancement in machine learning with the introduction of deep learning through Convolutional Neural Networks (CNN) which enable machine to learn good features from raw image automatically without any human intervention. Therefore, it is tempting to explore the ways to transform the existing NR-IQA-CDI from using handcrafted features to machine-crafted features using deep learning, specifically Convolutional Neural Networks (CNN). The results show that NR-IQA-CDI based on non-pre-trained CNN (NR-IQA-CDI-NonPreCNN) significantly outperforms those which are based on handcrafted features. In addition to showing best performance, NR-IQA-CDI-NonPreCNN also enjoys the advantage of zero human intervention in designing feature, making it the most attractive solution for NR-IQA-CDI.

Corresponding Author:

Norziana Jamil,
Center of Information and Network Security,
Universiti Tenaga Nasional,
Jalan Ikram-Uniten, Institut Latihan Sultan Ahmad Shah, 43000 Kajang, Malaysia.
Email: norziana@uniten.edu.my

1. INTRODUCTION

Image quality can be degraded due to various types of distortion such as noise, blurring, fast fading, blocking artifacts and contrast distortion. These distortions may occur during operations such as acquisition, compression, storage, transmission, display and post-processing. Contrast distortion is among the most common and fundamental distortion. Contrast-distorted image (CDI) is an image with low range of grayscale as shown in Figure 1. Contrast distortion may be caused by poor lighting condition and poor quality image acquisition device.

Many image quality assessment algorithms (IQAs) have been developed during the past decade. However, most of them are designed for images distorted by compression, noise and blurring. Such distortions cause structural change in image [1] which does not happen in contrast distortion. Hence, it is not suitable to use the above mentioned IQAs to assess contrast-distorted images (CDI).

(a) (b)

Figure 1. (a) Good contrast image; (b) Poor contrast image

There are very few IQAs designed specifically for CDI. The first IQA for CDI is a Reduced-Reference IQA (RR-IQA) called RIQMC [2]. The disadvantage of RIQMC is that it requires partial access to reference image, which is impractical in real-life application. Unlike distortion caused by image compression where the original image could be used as reference image, contrast distortion is caused by poor image acquisition conditions such as poor lighting or poor device so the original image itself is distorted and reference image is practically not available.

The first practical solution is proposed by Yaming et al. which is called No-Reference IQA for CDI (NR-IQA-CDI) [3]. It is develop based on the principles of Natural Scene Statistics (NSS) in that there are certain regularities in the statistics of natural scenes which could be missing from the statistics distorted images. The five features used in NR-IQA-CDI are the global spatial statistics of an image including the mean, standard deviation, entropy, kurtosis and skewness. Unfortunately, the performance of NR-IQA-CDI are not encouraging in two of the three test image databases, TID2013 and CSIQ, where the Pearson Linear Correlation Coefficients are only around 0.57 and 0.76, respectively.

The existing NR-IQA-CDI relies on features designed by human or handcrafted features because considerable level of skill, domain expertise and efforts are required to design good handcrafted features. Recently, there is great advancement in machine learning with the introduction of deep learning through Convolutional Neural Networks (CNN) which enable machine to learn good features from raw image automatically without any human intervention. Therefore, it is tempting to explore the ways to transform the existing NR-IQA-CDI from using handcrafted features to machine-crafted features using deep learning, specifically Convolutional Neural Networks (CNN).

The evaluation results indicated that NR-IQA-CDI based on machine-crafted features generally performed better than NR-IQA-CDI based on handcrafted features while enjoying the advantage of requiring zero human intervention in identifying the features. In the next section (Section 2), Brief overview on CNN for NR-IQA. Section 3 Designing NR-IQA-CDI Based on Non-Pre-Trained CNN Models are described. Section 4 describes the performance evaluation and Section 5 concludes the current work.

2. CONVOLUTIONAL NEURAL NETWORKS (CNNs) FOR NR-IQA

Artificial neural networks has been the most popular tools for machine learning [4], which in more general sense for deep learning. Among several deep learning architectures, stacked denoising autoencoders [5], deep belief networks [6-7], and convolutional neural networks [8-13] are three of the most popular architectures utilized for different type of applications. Convolutional neural networks (CNNs) are a special kind of deep learning method, CNNs run much faster on GPUs, such as NVidia‘s Tesla K80 processor, and has achieved state of the art performance on various computer vision tasks, such as object detection, recognition, retrieval, annotation, image classification, and segmentation [14-16].

The fundamental difference between convolutional neural network (CNN) and conventional machine learning is that, rather than using hand-crafted features, such as SIFT [17] and HoG, CNN can automatically learn features from data (images) and acquire scores from the output of it [18]. Figure 2 shows the difference between machine learning and deep learning.

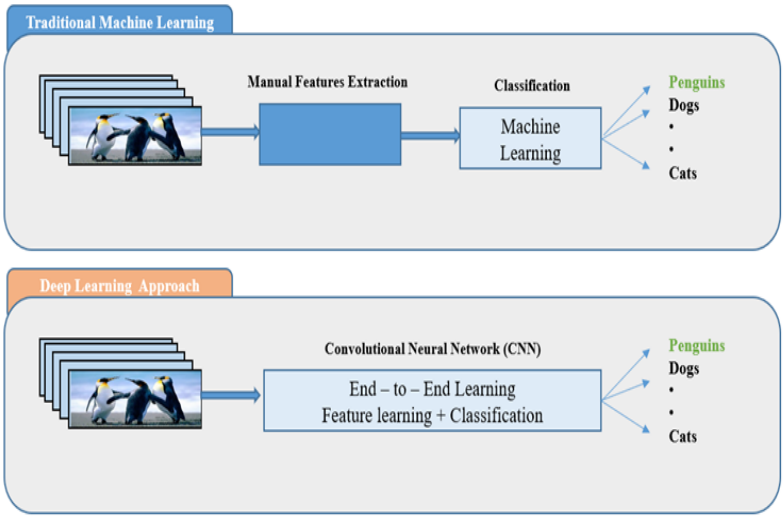


Figure 2. The difference between CNN and machine learning

Generally, a CNN architecture comprises different layer types such as convolutional layer, Rectified Linear Unit (ReLU) layer, cross channels normalization layer, pooling layer, fully connected layer, dropout layer, SoftMax layer, and output classification layer. Each layer obtains the data from the previous layer. Then, the data is transformed and passed to the subsequent layer. CNN architecture varies in terms of the number of outputs per layer, the size and type of spatial pooling, the number of layers, and the size of the convolutional filters. In general, CNNs are trained in a supervised pattern using the standard backpropagation [19]. Figure 3 shows the typical architecture of a CNN model.

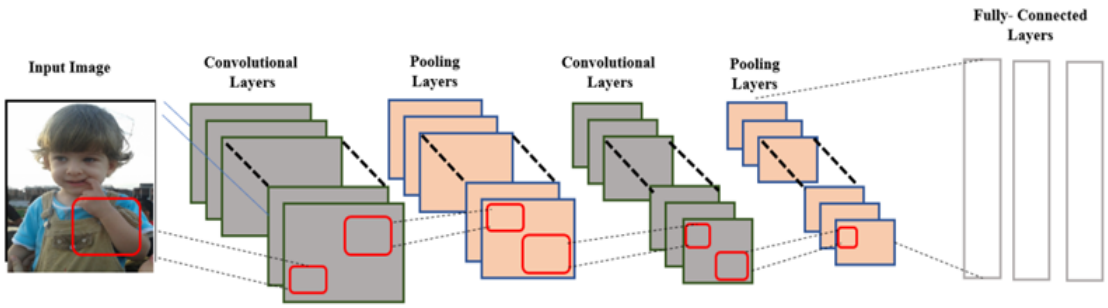


Figure 3. The typical architecture of convolutional neural networks (CNNs)

The application of CNN in IQA was first proposed by Kang et al. [20]. They treated image patches as input and employed CNN to predict the image quality. As a result, CNN could predict the quality score on small image patches in an accurate manner. Also, instead of using handcrafted features, it could merge the feature learning and regression processes into a single optimization process.

In order to eliminate the need for manual feature extraction, deep learning is performed to learn the features from raw data (images) automatically. For example, NR-IQA learns important features automatically from raw images [21]. Most of the conventional NR-IQA depends on two main steps: feature extraction and score prediction, while in NR-IQA based on machine-crafted features, learning and feature extraction are integrated into one single step.

3. DESIGNING NR-IQA-CDI BASED ON NON-PRE-TRAINED CNN MODELS

The proposed NR-IQA-CDI based on non-pre-trained CNN (NR-IQA-CDI-NonPreCNN) is trained from scratch is illustrated in Figure 4. The details of the network architecture and the training procedures are as presented in the following two sections.

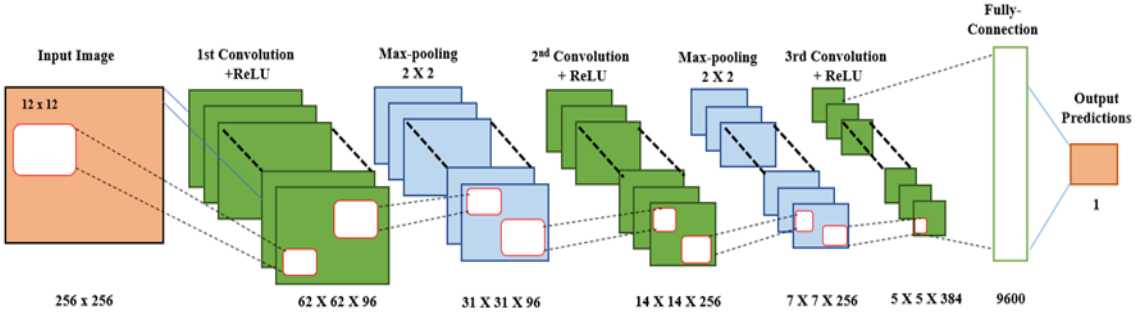


Figure 4. The diagram of NR-IQA-CDI-nonpreCNN based on CNN architecture

3.1. Network architecture

Instead of feeding the entire image as an input to the proposed network, grayscale image is decomposed into non-overlapping patches of size N x M, where N > H, image height and M > W, image width. These non-overlapping patches of the image are given to network as input since this would increase number of training samples tremendously and could help to improve the performance. Experiments were conducted on various patch size (32, 64, 128, 256). The experiments showed that the optimum patch size is 64 x 64. Each image patch is considered as an independent image sample during the training step, and it is labeled with the quality value of the corresponding source image.

There is no golden rule in designing CNN model in terms of the number of layers and the size of filters. This work started with designing the network with 3 convolutional layers. Each layer contains 96, 256, and 384 filters of size 12x12, 5x5, 3x3 respectively. This work was also tested on various settings of number of layers and filter size and the results are as presented in the coming section 4, Figures 7 and 8. The ReLU (Rectified Linear Unit) activation functions were embedded as well. The first two convolutional layers were subjected to 2x2 max pooling while no pooling was applied to the third layer. The three convolutional layers were attached to a fully-connected layer containing 9600 hidden units and a linear regression layer (for image quality score prediction).

The network layers are organized as follows: Input patch 256×256, Conv 96x62x62− Max 31x31x96− Conv 256x14x14− Max 7x7x256− Conv 384x5x5− FC 5x5x384=9600− Output (FC-1). (Conv denotes the convolution layer; max denotes the max-pooling layer; FC denotes the full-connected layer.) Figure 4 shows the architecture of proposed (NR-IQA-CDI-NonPreCNN), which is based on Convolutional Neural Networks with 8 layers.

3.2. Training the CNN

Since the input size (image size) of CNN is fixed, the input images for all databases were resized to 512x512. During the training phase, the quality label of the whole image was assigned to all patches of the same image. The proposed networks were trained repeatedly by performing backpropagation over several epochs. Here, one epoch is defined as the period during in which each sample from the training set has been used once. While fixing the learning rate as 0.0001, all models were trained for 150 epochs. Upon inserting the training image as input, the forward propagation step (consisting of convolution, ReLU and pooling operations in the fc layer) was performed to identify the output probability of each class. A laptop (Intel (R) Core (TM) 2 Duo CPU, 8G RAM memory and NVDIA GTX 950M GPU with a MATLAB R2017a platform) was used to perform the experiment.

4. PERFORMANCE EVALUATION

In this section, the performance evaluation of the proposed NR-IQA-CDI based on non-pre-trained CNN (NR-IQA-CDI-NonPreCNN) is trained from scratch is presented. The presentation begins with the details of the evaluation methodology such as the test image databases, performance metrics and evaluation procedures. This is followed by discussions on the evaluation results and conclusions from the performance evaluation.

4.1. Evaluation methodology

The test image databases used for the evaluation are similar to those used to evaluate the existing NR-IQA-CDI for fair comparison. They are CSIQ database [22], TID2013 database [23] and CID2013 database [2]. The test images used include only the contrast distorted images in the three databases without

the reference images. A total of 116, 250, and 400 distorted images are selected from CSIQ [22], TID2013 [23], and CID2013 [2], respectively. Subjective scores are represented by either mean opinion score (MOS) or differential mean opinion score (DMOS).

Cross-validation was used in the performance evaluation. It is a model validation method to evaluate how well the performance of the model could generalize to an independent data set. k-fold cross-validation (k-fold cv) was chosen for this work. This method allows performance evaluation with many different combinations of data set to minimize bias. In this method, data are divided into k subsets and performance evaluation is repeated for k times. During the k times of evaluation, each of the k subsets is used for testing for one time while the rest used for training. The final evaluation result is the average results of the k times of evaluation.

It is well-known that the performance of a machine learning model tends to improve with the increase number of training data. Therefore, k-fold cv with higher k tends to show better performance. In this work, the k-fold cv was repeated for k range from 2 to 10 to reduce bias due to the size of training data. However, only 10 train-test iterations are conducted as the training of CNN is very time consuming.

In order to evaluate the performance of IQA, the performance metrics such as (1) Spearman Rank-Order Correlation Coefficient (SROCC), (2) Pearson's (Linear) Correlation Coefficient (PLCC) and (3) Root Mean Square Error (RMSE) between the predicted objective scores and the subjective Mean Opinion Scores (MOS) were employed. Conditions such as SROCC~1, PLCC~1 and RMSE~0 indicate good performance in terms of correlation with human perception.

4.2. Evaluation of NR-IQA-CDI-nonpreCNN

The evaluation was repeated with various patch sizes M x N such as 256 x 256, 128 x 128, 64 x 64, and 32 x 32 as well as without patches (the image was resized before input to CNN) to determine the best setting. Tables 1-5 show the average result of assessment using patch size 256 x 256, 128 x 128, 64 x 64, 32 x 32 and without using patch, respectively.

Table 1. The average PLCC, SROCC and RMSE across 10 train-test rounds for three databases using NR-IQA-CDI-nonpreCNN of patch size 256 x 256

K	CSIQ			TID2013			CID2013			All Databased		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
2	0.7629	0.7098	0.8862	0.7940	0.7689	0.7183	0.8476	0.8647	0.3246	0.8969	0.9427	0.6878
3	0.7610	0.7587	0.8974	0.8182	0.8041	0.6398	0.8839	0.8876	0.3023	0.9083	0.9148	0.8410
4	0.8276	0.8819	0.8501	0.8102	0.7884	0.7501	0.9179	0.9332	0.2433	0.8893	0.8791	0.9793
5	0.8682	0.9097	0.7956	0.8176	0.7853	0.6236	0.9198	0.9429	0.2398	0.9184	0.9310	0.7208
6	0.8893	0.8663	0.7103	0.8219	0.7797	0.5879	0.9155	0.9353	0.2311	0.8891	0.9197	0.8510
7	0.9282	0.9120	0.6014	0.8299	0.7899	0.6217	0.9285	0.9421	0.2229	0.8595	0.8829	0.9326
8	0.8997	0.9229	0.7068	0.8202	0.7931	0.6072	0.9299	0.9472	0.2142	0.9076	0.9206	0.8638
9	0.9513	0.9293	0.5103	0.8319	0.7932	0.6163	0.9276	0.9395	0.2119	0.9085	0.8901	0.9429
10	0.9351	0.9384	0.6892	0.8358	0.7989	0.5909	0.9392	0.9528	0.2024	0.9354	0.8587	0.9692
Av	0.8692	0.8698	0.7385	0.8199	0.7890	0.6395	0.9122	0.9272	0.2436	0.9014	0.9044	0.8653

Table 2. The average PLCC, SROCC and RMSE across 10 train-test rounds for three databases using NR-IQA-CDI-nonpreCNN of patch size 128 x 128

K	CSIQ			TID2013			CID2013			All Databased		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
2	0.7612	0.7589	0.8398	0.8189	0.7679	0.6892	0.8892	0.8921	0.3041	0.9089	0.9579	0.6709
3	0.7926	0.7996	0.8418	0.8174	0.8093	0.6302	0.9125	0.9127	0.2959	0.9181	0.9282	0.8311
4	0.8469	0.8988	0.8123	0.8426	0.8182	0.6686	0.9295	0.9397	0.2312	0.9013	0.8938	0.9532
5	0.8784	0.9321	0.7502	0.8584	0.8367	0.5611	0.9312	0.9558	0.2317	0.9294	0.9476	0.7045
6	0.9116	0.8892	0.6879	0.8493	0.8265	0.5723	0.9289	0.9464	0.2219	0.9012	0.9324	0.8439
7	0.9328	0.9269	0.5792	0.8494	0.8326	0.5639	0.9372	0.9553	0.2034	0.8719	0.8993	0.9256
8	0.9179	0.9394	0.6779	0.8568	0.8398	0.5612	0.9385	0.9592	0.2012	0.9278	0.9391	0.8571
9	0.9723	0.9416	0.5026	0.8538	0.8356	0.5579	0.9369	0.9489	0.2049	0.9283	0.9089	0.9336
10	0.9417	0.9519	0.6463	0.8684	0.8497	0.5231	0.9448	0.9656	0.1961	0.9521	0.8698	0.9541
Av	0.8839	0.8931	0.7042	0.84611	0.8240	0.5919	0.9276	0.9417	0.2322	0.9154	0.9196	0.8526

Table 3. The average PLCC, SROCC and RMSE across 10 train-test rounds for three databases using NR-IQA-CDI-nonpreCNN of patch size 64 x 64

K	CSIQ			TID2013			CID2013			All Databased		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
2	0.7841	0.7828	0.8212	0.8294	0.7893	0.6732	0.9075	0.9068	0.2808	0.9258	0.9711	0.6371
3	0.8062	0.8095	0.8337	0.8281	0.8278	0.6088	0.9325	0.931	0.2711	0.9372	0.9476	0.8161
4	0.8626	0.9174	0.8011	0.8496	0.8348	0.6379	0.9497	0.9594	0.2102	0.9245	0.9182	0.9329
5	0.8979	0.9391	0.731	0.8674	0.8523	0.5443	0.9512	0.9721	0.2113	0.9491	0.9653	0.6784
6	0.9253	0.9098	0.6725	0.8583	0.8417	0.5532	0.9489	0.9625	0.2019	0.9282	0.9527	0.8126
7	0.9441	0.9411	0.5509	0.8581	0.8474	0.5439	0.9565	0.9686	0.1932	0.9026	0.9288	0.8979
8	0.9332	0.9532	0.6513	0.8678	0.8582	0.5437	0.9577	0.9714	0.1939	0.9487	0.9529	0.8274
9	0.9798	0.9561	0.4924	0.8685	0.8525	0.5326	0.9562	0.9667	0.1943	0.9495	0.9298	0.9017
10	0.9602	0.9687	0.6187	0.8859	0.8612	0.501	0.9671	0.9798	0.1794	0.9712	0.9028	0.9268
Av	0.8992	0.9086	0.6858	0.8570	0.8405	0.5709	0.9474	0.9575	0.2151	0.9374	0.9410	0.8256

Table 4. The average PLCC, SROCC and RMSE across 10 train-test rounds for three databases using NR-IQA-CDI-nonpreCNN of patch size 32 x 32

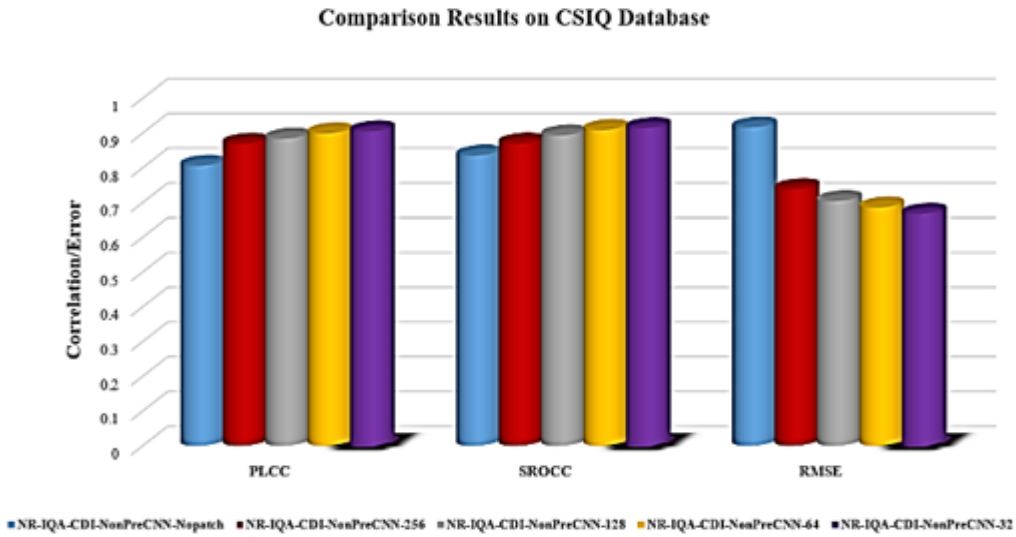
K	CSIQ			TID2013			CID2013			All Databased		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
2	0.7892	0.7878	0.8123	0.8395	0.7992	0.6626	0.9179	0.9175	0.2732	0.9368	0.9795	0.6296
3	0.8087	0.8169	0.8167	0.8389	0.8367	0.6007	0.9364	0.9397	0.2627	0.9464	0.9569	0.8083
4	0.8694	0.9182	0.7812	0.8586	0.8412	0.6251	0.9558	0.9624	0.2019	0.9327	0.9288	0.9241
5	0.9121	0.9427	0.7119	0.8693	0.8623	0.5246	0.9569	0.9781	0.2036	0.9576	0.9722	0.6702
6	0.9284	0.9166	0.6525	0.8679	0.8517	0.5344	0.9547	0.9692	0.1921	0.9381	0.9619	0.8063
7	0.9467	0.9472	0.5309	0.8668	0.8537	0.5258	0.9631	0.9726	0.1814	0.9121	0.9379	0.8909
8	0.9389	0.9565	0.6211	0.8734	0.8698	0.5247	0.9655	0.9749	0.1807	0.9587	0.9624	0.8178
9	0.9831	0.9596	0.4834	0.8772	0.8589	0.5129	0.9649	0.9742	0.1832	0.9595	0.9397	0.8935
10	0.9679	0.9827	0.6007	0.8938	0.8746	0.4965	0.9742	0.9895	0.1765	0.9816	0.9183	0.9142
Av	0.9049	0.9142	0.6678	0.8650	0.8497	0.5563	0.9543	0.9642	0.2061	0.9470	0.9508	0.8172

Table 5. The average PLCC, SROCC and RMSE across 10 train-test rounds for three databases using NR-IQA-CDI-nonpreCNN without patches

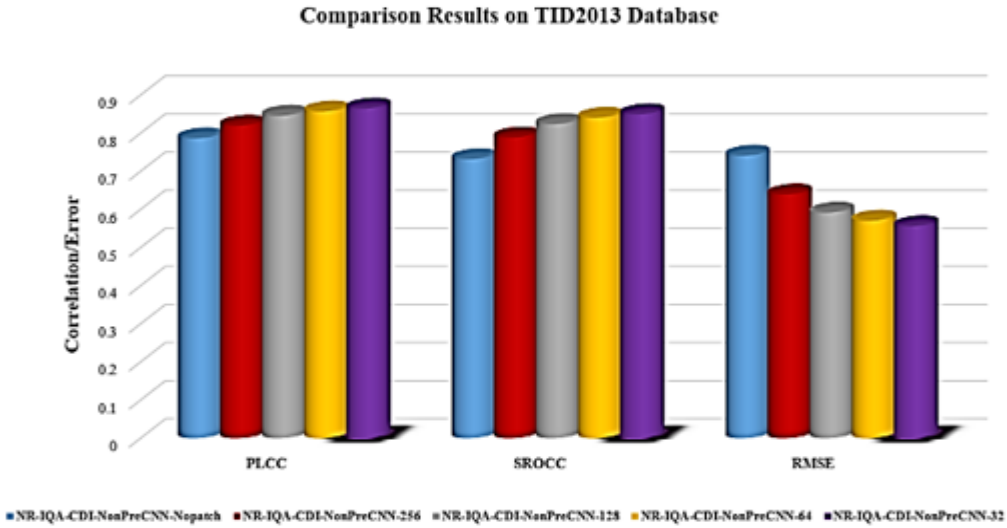
K	CSIQ			TID2013			CID2013			All Databased		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
2	0.5948	0.6204	1.4775	0.7717	0.7177	0.7720	0.7852	0.8371	0.4208	0.9375	0.9393	0.7031
3	0.6183	0.7109	1.4815	0.7904	0.7423	0.7411	0.8618	0.9018	0.3352	0.8845	0.8962	0.8778
4	0.7788	0.8395	0.9850	0.7886	0.7385	0.7358	0.8804	0.9087	0.3036	0.8340	0.8474	1.0409
5	0.8391	0.8807	0.8518	0.7842	0.7283	0.7300	0.8771	0.9105	0.3072	0.9094	0.9165	0.7560
6	0.8581	0.8489	0.7725	0.7851	0.7203	0.7402	0.8755	0.9066	0.3157	0.8777	0.8967	0.8997
7	0.8867	0.8905	0.6422	0.7880	0.7337	0.7420	0.8946	0.9150	0.2920	0.8490	0.8678	1.0530
8	0.8716	0.9044	0.7328	0.7855	0.7344	0.7432	0.8777	0.9110	0.3134	0.8906	0.9041	0.8922
9	0.9244	0.9152	0.5854	0.7942	0.7365	0.7276	0.8971	0.9213	0.3011	0.8537	0.8685	0.9646
10	0.8813	0.9066	0.7167	0.7907	0.7360	0.7378	0.8927	0.9156	0.2878	0.7291	0.7946	1.2726
Av	0.8059	0.8352	0.9161	0.7864	0.7319	0.7410	0.8713	0.9030	0.3196	0.8628	0.8812	0.9399

Figure 5 (a)-(c) presents the comparison of the results of different patch sizes over the three databases in form of bar chart. Notice that the performance clearly improved with the decrease of patch size. The NR-IQA-CDI-NonPreCNN of patch size 32 x 32 outperforms the rest across all databases and all three performance metrics.

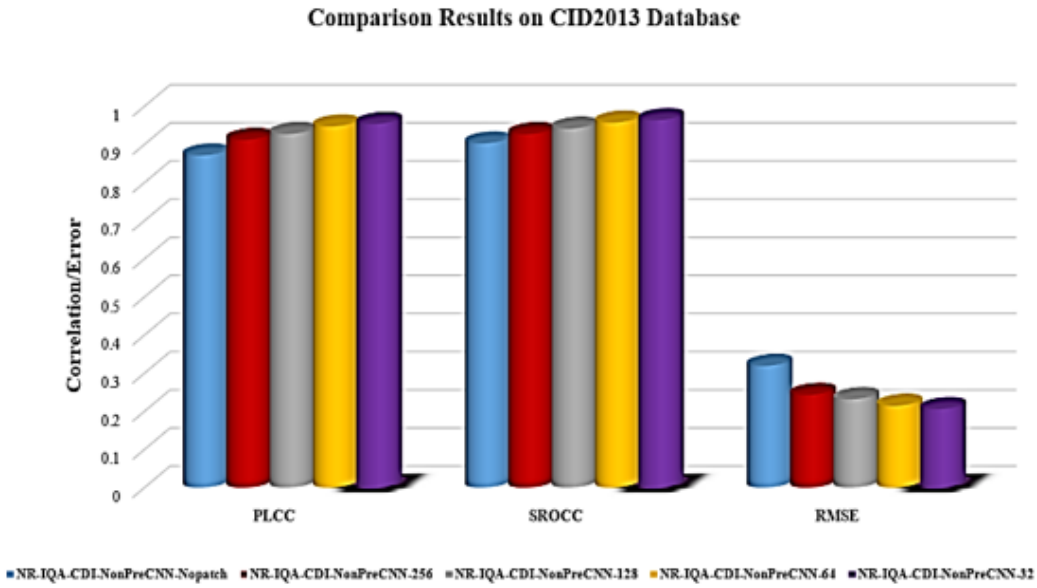
It is apparent that increasing the number of training samples by dividing the image into patches could affect the performance of CNN-based NR-IQA-CDI. For example, there were only 400 images in CID2013 database but there are about 2400 image patches of patch size 256 x 256 used, and about 153600 image patches when patch 32 x 32 was used as shown in Table 6. The average results of k-fold cross-validation (k= 2 to 10) are summarized in Table 7 where the patch size ranges from 32 to 256 pixels. It is observed that the proposed method shows better performance with increased number of patches as shown in Figure 6. It can also be observed that the performance of NR-IQA-CDI-NonPreCNN of small patch size 32 x 32 were not significantly improved as compared to those of patch size 64 x 64 to justify the exponential increase in the number of training samples. Therefore, the patch size of 64 x 64 was chosen as the optimum patch size.



(a)



(b)



(c)

Figure 5. Comparison of SROCC, PLCC, and RMSE of NR-IQA-CDI-nonpreCNN of patch size 256 x 256, 128 x 128, 64 x 64, 32 x 32, and without patch on CSIQ, TID2013, and CID2013 databases (a, b, and c)

Table 6. The patch numbers across different patch sizes for IQA databases

PATCH SIZE	CSIQ PATCH NO	TID2013 PATCH NO	CID2013 PATCH NO	ALL DATABASES PATCH NO
RESIZED	116	250	400	766
256	464	1000	2400	3864
128	1856	3000	9600	14456
64	7424	12000	38400	57824
32	29696	48000	153600	231296

Table 7. Performance evaluation of PLCC, SROCC, and RMSE on different patches number

Patch Size	CSIQ			TID2013			CID2013			All Databases		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Resized	0.8059	0.8352	0.9161	0.7864	0.7319	0.7410	0.8713	0.9030	0.3196	0.8628	0.8812	0.9399
256	0.8692	0.8698	0.7385	0.8199	0.7890	0.6395	0.9122	0.9272	0.2436	0.9014	0.9044	0.8653
128	0.8839	0.8931	0.7042	0.8461	0.8240	0.5919	0.9276	0.9417	0.2322	0.9154	0.9196	0.8526
64	0.8992	0.9086	0.6858	0.8570	0.8405	0.5709	0.9474	0.9575	0.2151	0.9374	0.9410	0.8256
32	0.9049	0.9142	0.6678	0.8650	0.8497	0.5563	0.9543	0.9642	0.2061	0.9470	0.9508	0.8172

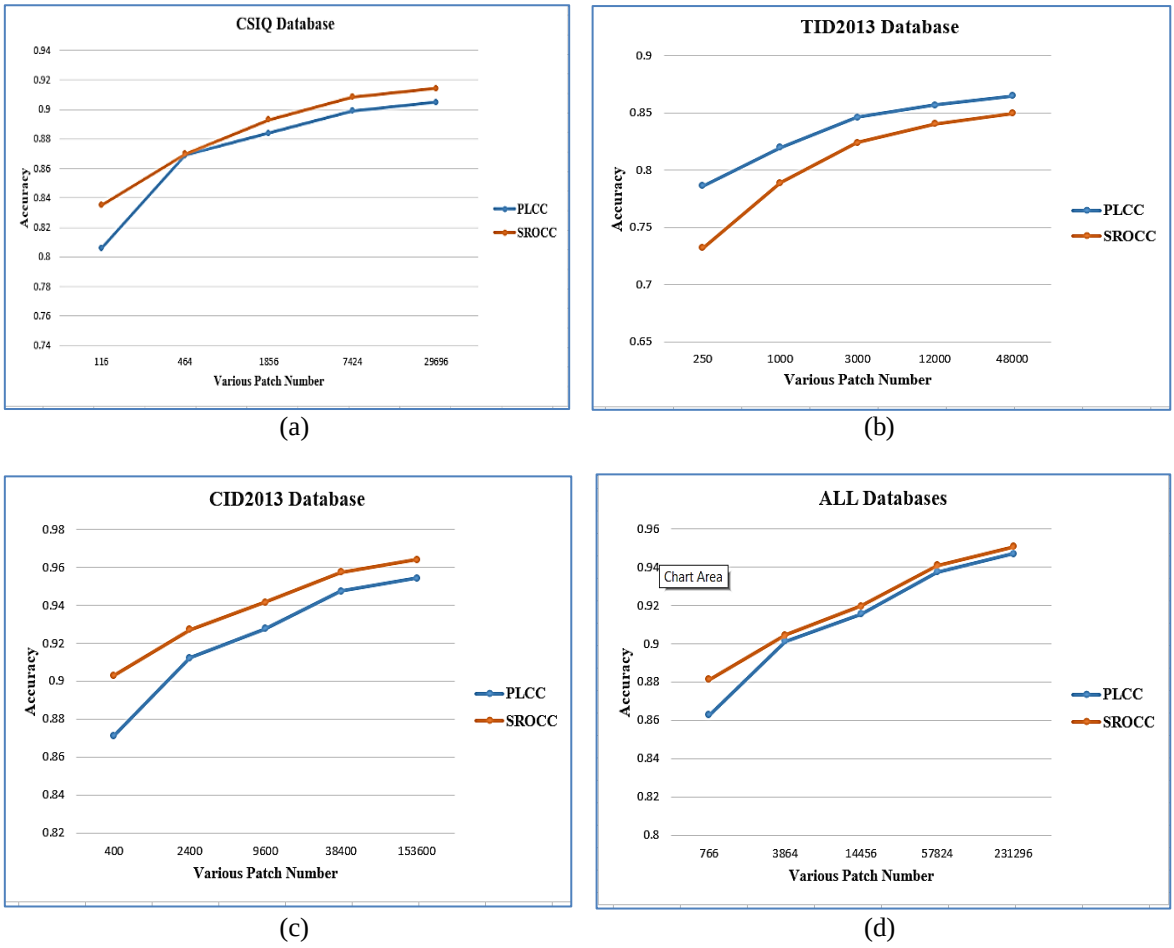


Figure 6. PLCC and SROCC results with various patches number
(a: CSIQ, b: TID2013, c: CID2013 DB, d: all databases)

Figure 7 shows the variation of performance with respect to the number of convolution kernels. In general, the use of more kernels would lead to better performance. In the case of CSIQ, the kernel number of the three convolution layers in setting 1-4 are (8, 16, 32), (16, 32, 64), (32, 64,128), and (96,256,384), respectively. It can be concluded that the performance could be improved by increasing the number of kernel. Similar conclusion also applies to databases TID2013 and CID2013.

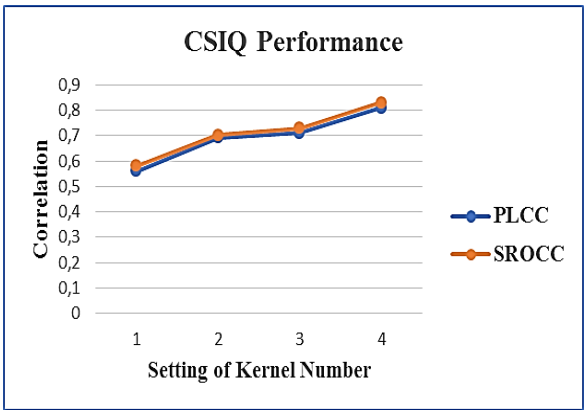


Figure 7. The SROCC and LCC with various number of kernels (8, 16, 32), (16, 32, 64), (32, 64,128), and (96,256,384) which correspond to 1,2,3,4, respectively

Generally, higher layers could extract better features. However, high-level features are not necessarily better than the low-level ones. From Figure 8, performance increased as the number of convolution layers was increased from one to three. However, the performance started to decrease when the numbers of layers were further increased. Similar results were found TID2013 and CID2013. Therefore, CNN with three convolution layers was optimal for NR-IQA-CDI.

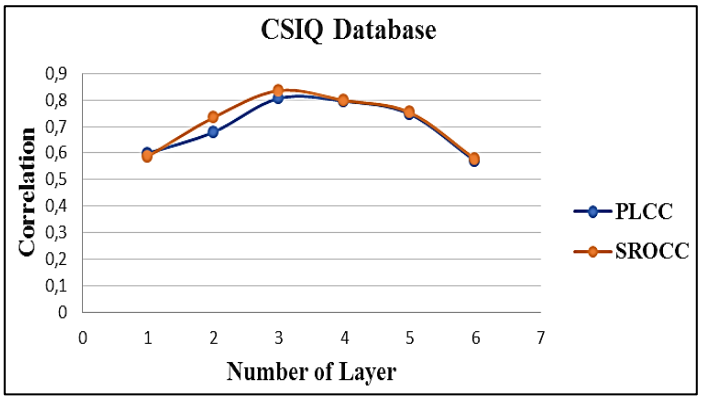


Figure 8. Proposed CNN prediction accuracy for CSIQ database across various layers

4.3. Cross-database evaluation

The cross-database evaluation method used to identify the generalization ability of a NR-IQA-CDI across different databases has been introduced in the current work. In this evaluation, all images in one of the databases are used for training while the images in the other two databases are used for testing. For consistency, the MOS and DMOS scores have been rescaled (0-10). The results of the cross-dataset test have been reported in Table 8. It may be observed that results when our method trained on CSIQ is decreased compared to being trained on CID2013. Therefore, larger training set would lead to better generalization. Hence, the generalization capability of a deep neural network is dependent on the size and diversity of the training set.

Table 8. Cross database evaluation

Trained on:	CID2013				TID2013				CSIQ			
Tested on:	CSIQ		TID2013		CID2013		CSIQ		CID2013		TID2013	
NR-IQA-CDI-NonPreCNN	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
	0.8169	0.8372	0.7935	0.7782	0.8273	0.8302	0.8092	0.8163	0.8028	0.8172	0.6584	0.6419

4.4. NR-IQA-CDI-nonpreCNN vs NR-IQA-CDI based on handcrafted features

The proposed NR-IQA-CDI-NonPreCNN is compared against NR-IQA-CDI based on handcrafted features. The comparison results shown in Table 9. The best score for each performance metric and database is highlighted by bolding the numbers.

Table 9. PLCC, SROCC, and RMSE comparison among enhanced versions of NR-IQA-CDI

Methods	CSIQ			TID2013			CID2013		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Existing NR-IQA-CDI	0.7623	0.7657	0.1102	0.5739	0.4943	0.8146	0.8797	0.8798	0.2981
NR-IQA-CDI-NonPreCNN	0.8992	0.9086	0.0686	0.8570	0.8405	0.5709	0.9474	0.9575	0.2151

The results show that NR-IQA-CDI based on non-pre-trained CNN, NR-IQA-CDI-NonPreCNN significantly outperforms those which are based on handcrafted features. In addition to showing best performance, NR-IQA-CDI-NonPreCNN also enjoys the advantage of zero human intervention in designing feature, making it the most attractive solution for NR-IQA-CDI.

5. CONCLUSION

In this paper, a study on transforming the existing NR-IQA-CDI using machine-crafted features based on CNN has been presented. It was able to accurately predict the quality score from an image and integrate the feature learning and regression into one optimization process. Our method does not require any reference image, and any handcrafted features and directly learns discrimination features from raw image pixels to achieve much better performance. The evaluation results indicated that NR-IQA-CDI based on machine-crafted features generally performed better than NR-IQA-CDI based on handcrafted features while enjoying the advantage of requiring zero human intervention in identifying the features.

ACKNOWLEDGEMENTS

This research is funded by 2018 Smart Fund of Universiti Tenaga Nasional.

REFERENCES

[1] Z. Wang, *et al*., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, 2004.
[2] K. Gu, *et al*., "Subjective and objective quality assessment for images with contrast change," *2013 IEEE International Conference on Image Processing*, Melbourne, VIC, 2013, pp. 383-387.
[3] Y. Fang, *et al*., "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, pp. 838-842, 2015.
[4] S. Juma, *et al*., "Machine learning techniques for intrusion detection system: A review," *Journal of Theoretical and Applied Information Technology*, vol. 72, pp. 422-429, 2015.
[5] P. Vincent, *et al*., "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal Of Machine Learning Research*, vol. 11, pp. 3371-3408, 2010.
[6] Y. Bengio, *et al*., "Greedy layer-wise training of deep networks," In *Advances in neural information processing systems*, pp. 153-160, 2006.
[7] S. Narejo, *et al*., "EEG Based Eye State Classification using Deep Belief Network and Stacked AutoEncoder," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6(6), pp. 3131-3141, 2016.
[8] D. Graupe, *et al*., "Applications of neural networks to medical signal processing," *Proceedings of the 27th IEEE Conference on Decision and Control*, 1988, pp. 343-347.
[9] N. Jamil, *et al*., "Can Convolution Neural Network (CNN) Triumph in Ear Recognition of Uniform Illumination Invariant?" *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 11(2), pp. 558-566, 2018.
[10] J. Gu, *et al*., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354-377, 2018.
[11] S. Ji, *et al*., "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 221-231, 2013.
[12] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, pp. 2352-2449, 2017.
[13] M. Nimbarte and K. Bhoyar, "Age Invariant Face Recognition using Convolutional Neural Network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8(4), pp. 2126-2138, 2018.
[14] Y. Li, *et al*., "No-reference image quality assessment with deep convolutional neural networks," *2016 IEEE International Conference on Digital Signal Processing*, 2016, pp. 685-689.
[15] Y. LeCun, *et al*., "Deep learning," *Nature 521*, Google Scholar, vol. 7553, pp. 436, 2015.

[16] S. Razavian, *et al*., "CNN features off-the-shelf: an astounding baseline for recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.
[17] D. Wei and Y Li, "No-Reference Image Quality Assessment Based on SIFT Feature Points," *International Journal of Simulation--Systems, Science & Technology*, vol. 17, 2016.
[18] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal Of Selected Topics In Signal Processing*, vol. 11, pp. 206-220, 2017.
[19] S. Bianco, *et al*., "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, pp. 355-362, 2018.
[20] L. Kang, *et al*., "Convolutional neural networks for no-reference image quality assessment," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733-1740.
[21] S. A. Amirshahi, *et al*., "Image quality assessment by comparing cnn features between images," *Journal of Imaging Science and Technology*, vol. 60, pp. 60410-1, 2016.
[22] E. C. Larson and D. M Chandler, "Categorical image quality (CSIQ) database," 2010.
[23] N. Ponomarenko, *et al*, "Color image database TID2013: Peculiarities and preliminary results," *Visual Information Processing (EUVIP), 2013 4th European Workshop on*, pp. 106-111, 2013.