Letters

A Novel Patch Variance Biased Convolutional Neural Network for No-Reference Image Quality Assessment

Lai-Man Po, *Senior Member, IEEE*, Mengyang Liu, Wilson Y. F. Yuen, Yuming Li, Xuyuan Xu, Chang Zhou, Peter H. W. Wong, *Senior Member, IEEE*, Kin Wai Lau, and Hon-Tung Luk

***Abstract*— Deep convolutional neural networks (CNNs) have been successfully applied on no-reference image quality assessment (NR-IQA) with respect to human perception. Most of these methods deal with small image patches and use the average score of the test patches for predicting the whole image quality. We discovered that image patches from homogenous regions are unreliable for both neural network training and final image quality score estimation. In addition, image patches with complex structures have much higher chances of achieving better image quality prediction. Based on these findings, we enhanced the conventional CNN-based NR-IQA algorithm to avoid homogenous patches for the network training and quality score estimation. Moreover, we also use a variance-based weighting average to bias the final image quality score to the patches with complex structure. The experimental results show that this simple approach can achieve state-of-the-art performance compared with well-known NR-IQA algorithms.**

***Index Terms*— Deep learning, convolution neural network, no-reference image quality assessment.**

I. Introduction

During the last three decades, the volume of digital image data was growing explosively due to the rapid development of multimedia and networking technologies. Nowadays, every hour has a massive number of digital images generated that makes image quality assessment (IQA) become a popular area for both academic and industrial developments. According to the dependency of reference images, IQA methods are usually divided into 3 types: full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA) and no-reference IQA (NR-IQA). FR-IQA and RR-IQA metrics assume that the whole or partial information of the reference image is available, and do a comparison between reference image and tested image. PSNR, SSIM [1], FSIM [2], IFC [3] and VIF [4] are well-known FR-IQA algorithms. However, the reference image is not always available that makes NR-IQA more desirable for practical applications and many NR-IQA algorithms have been developed. The first generation of those algorithms are calibrated to some specific distortions such as JPEG [5], JPEG2000 [6] and H.264/AVC [7]. They are difficult to be generalized for other new distortion types. The second-generation NR-IQA algorithms focus on the investigation of natural scene statistics (NSS) and use handcrafted features that supposedly captures relevant factors affecting image quality. Well-known NSS-based algorithms are DIIVINE [8], BLIINDS-II [9] and BRISQUE [10].

In recent years, deep learning has been proved to perform well on a large variety of problems. The rise of deep learning had also revolutionized the NR-IQA development as a data driven approach, which learn discriminative features directly from raw image patches. Pioneer works of this approach are CORNIA [11] and CNN-NRIQA [12]. CORNIA aims at training image representation kernels directly from raw image pixels using unsupervised feature learning and CNN-NRIQA integrates feature learning and regression into one optimization process using Convolutional Neural Network (CNN).

Basically, Kang *et al.* [12] are the pioneers to apply CNN to NR-IQA. They proposed a very meaningful framework and achieved excellent results. This approach mainly deal with small image patches (such as 32×32) and the whole image quality score is the average predicted scores of the small test patches. However, Kang's CNN-NRIQA [12] network only contains one convolution layer, which is too shallow. Thus, complete image with size of 224×224 pixels were used in [13] and [14] to train deep CNNs with many layers for handling the small patch problem that cannot match with human perception. While Sun *et al.* [15] and Bosse *et al.* [16] applied existing deep CNNs to fine-tune the parameters. In addition, CNNs designed for small patches with weight adjustment for each patch were proposed in [17] and [18]. Recently, Cheng *et al.* [19] proposed a pre-saliency map (pre-SM) based NR-IQA method via CNN. They demonstrated that prediction error of the image patches in saliency regions is on average lower than that in homogenous regions using a fast saliency map (SM) model. Based on this result, pre-SM algorithm adaptively applies CNN computation on image patches and assigns higher weights for salient patches in the whole score estimation. It can achieve high accuracy with subjective quality. Liu *et al.* [20] proposed a RankIQA with use of Siamese network to train the CNN for achieving better image quality score prediction.

In this letter, we first report an interesting discovery of image patches with low variances are not reliable for quality score estimation in CNN-based NR-IQA. We use the distribution of quality score prediction errors against patch variances for demonstrating very low-variance patches are not robust and we should bias to the patches with high variances. Based on this finding, we enhanced the conventional CNN-based NR-IQA network to avoid the use of very low-variance image patches for both network training and quality score prediction. The remainder of the letter is organized as follows. Section II discusses the main problem of the small patch approach and presents the prediction error characteristic of image patches in terms of patch variances. Section III presents the proposed patch variance biased CNN-based NR-IQA algorithm. Experimental results are presented in section IV. Finally, conclusion is given in section V.

Manuscript received March 18, 2018; revised October 22, 2018; accepted January 1, 2019. Date of publication January 7, 2019; date of current version April 3, 2019. This work was supported by a grant from the Innovation and Technology Fund (ITF) of Hong Kong Government in the City University of Hong Kong under Project 9440172. This paper was recommended by Associate Editor B. Li. *(Corresponding author: Lai-Man Po.)*

L.-M. Po, M. Liu, and C. Zhou are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: eelmpo@cityu.edu.hk).

W. Y. F. Yuen, P. H. W. Wong, K. W. Lau, and H.-T. Luk are with TFI Digital Media Ltd., Hong Kong.

Y. Li is with Minieye Company, Shenzhen, China.

X. Xu is with Tencent Video, Tencent Holdings Ltd., Shenzhen 518057, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2019.2891159

	1	2		3	4		5	6		7	8		9	10		11	12
Input 32x32x3	Conv3-32	Conv3-32	Maxpool 2	Conv3-64	Conv3-64	Maxpool 2	Conv3-128	Conv3-128	Maxpool 2	Conv3-256	Conv3-256	Maxpool 2	Conv3-512	Conv3-512	Maxpool 2	FC-512	FC-1

Fig. 1. Image patch based CNN architecture for NR-IQA.

II. Prediction Characteristic of Image Patches

One of the main ideas of CNN-based NR-IQA is to use small image patches to train the network. During the training, patches are randomly sampled from the training images with quality labels set to the score of the source image. Moreover, each patch is treated as an independent sample [12]. After training, the network is used to estimate individual patch's score that is scanned from the tested image. The whole image quality score is based on the average predicted scores of the patches from the tested image. Thus, the accuracy of the final score is highly depended on the individual patch quality score estimation.

The main drawback of using small image patches is that not every patch has the same amount of information. Normally, homogenous patches from flat or smooth regions have relatively lower accuracy as compared with non-homogenous patches that consist of strong edges or complex textures. It is because homogenous regions of high and low quality images could be very similar in most of the real-world images. Thus, very similar homogenous patches could come from high and low quality images. As a result, similar homogenous patches have relatively higher chances to be assigned with very different quality labels during the CNN training process. These low-quality training data will confuse the network during the training process. In addition, the trained network is also unreliable for homogenous patch quality score estimation.

With the above hypothesis, we propose to use patch variance as the homogenous indicator for performing image patch quality score prediction error analysis. The patch variance is defined as the average pixel value variances of the patch in RGB color channels, which can be expressed as

$$var_{ave}(P) = \frac{1}{3}\left[var_R\,(P) + var_G\,(P) + var_B\,(P)\right] \tag{1}$$

where $var_R(P)$, $var_G(P)$ and $var_B(P)$ are variances of pixel values in RGB color channels, respectively. The reason to use patch variance as homogenous indicator is mainly due to homogenous patches normally have very low variances in pixel values. To start the analysis, we trained a CNN-based NR-IQA network with use of LIVE database [20] and the network architecture is shown in Figure 1. This network is very similar to Bosse's CNN in [16] with use of 3×3 convolutional kernels only. Basically, the network consists of 12 convolutional layers with max-pooling between every 2 convolutional layers. Except the last fully-connected layer, all layers are activated by ReLU activation function. Zero-padding and 3×3 kernels are used in all convolutional layers. Dropout regularization is added after layer 11 with 0.5 ratio and MAE loss function is used with 0.0001 learning rate for ADAM optimizer.

We randomly choose 80% images of LIVE database for training and the remaining 20% for testing. In addition, we densely sampled 70,650 image patches from the test set for quality score prediction error analysis. The scatter plot of these patches in terms of image quality score prediction errors against image patch variances is shown in Fig. 2. It can be easily observed that the prediction errors for patches with very low variances are not reliable for quality score estimation as their prediction errors are widely spread out. In contrast, the prediction errors distribution for patches with high variances are distributed on the much lower prediction error regions.

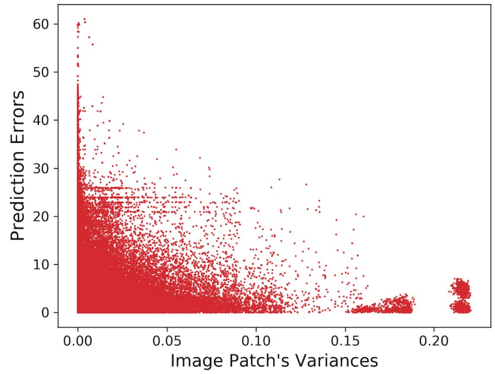


Fig. 2. Scatter plot of image patch quality score prediction errors against image patch variances.

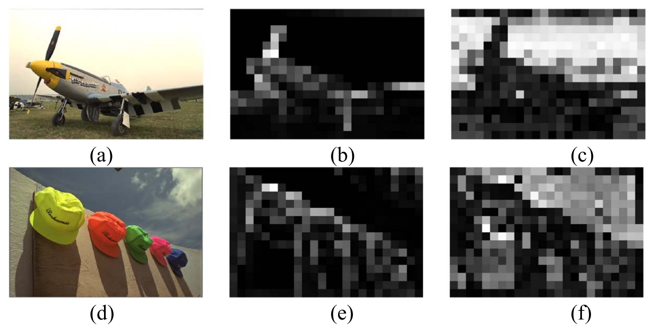


Fig. 3. Image examples and their corresponding patch variances and quality score prediction errors. (a)(d) Test images, (b)(e) Patch's variances, and (c)(f) Prediction errors, where brighter pixels indicate higher variances or errors.

To further visualize this phenomenon, two images from the test set with their correspondent patch variances and quality score prediction errors are shown in Figure 3. It can be found that the homogenous regions of these two images are having very low patch variances, which corresponding to the dark regions of Fig. 3(b) and 3(e). While most of these dark regions are corresponding to bright regions in Fig. 3(c) and 3(f), which indicate that homogenous regions create relatively higher prediction errors. These examples further demonstrate that the image patches with very low patch variances are not reliable for estimating the whole image quality score. In addition, we should bias to the patches with higher variances for calculating the overall image quality score.

III. Variance Biased Convolutional Neural Network

Based on the findings in section II, we should avoid to use homogenous patches during the CNN training and quality score estimation. To enhance the performance of the conventional CNN-based NR-IQA, we propose to use a variance threshold to avoid using these low-quality data in both CNN training and image quality estimation. In addition, we also propose to use variance-based weighting for calculating the whole image quality score.

A. Patch Sampling With Variance Threshold in CNN Training

To avoid homogenous patches used in the CNN training, we have to modify the patch sampling strategy. The most straightforward way

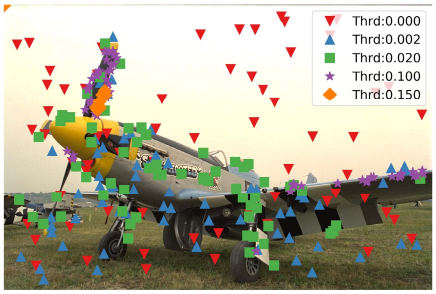


Fig. 4. Variance threshold based patch sampling examples with use of different threshold values.

is to use a patch variance threshold (T_{var}) to determine whether the sampled patch is good or not for the training. The pseudo code of this patch sampling strategy for CNN training is shown in Algorithm 1. This algorithm is used to randomly generate $M \cdot N$ patches $P_{i,j}$ with patch variances greater than or equal to T_{var} for each epoch of the training. In which the patches corresponding quality score $Q_{i,j}$ are assigned with the ground truth score of the source image I_i such that we can generate N patch/score pairs $[P_{i,j}, Q_{i,j}]$ from each training image with total $M \cdot N$ pairs. In realization of this algorithm, the patch variance is defined as the average $m \times m$ block variances of the patches in RGB color channels, where m is predefined as 32. In addition, the pixel's intensive values are normalized to [0, 1]. For each epoch of the training, a new set of training data are resampled to maintain the data richness. It is because the LIVE database is relatively small with only 29 reference images and 779 distorted images. Thus, N is set to 32 in our experiments for balancing the number of patches in each epoch and the data richness.

Algorithm 1 Patch Sampling Based on Variance Threshold

Input: Images from training set with M images
  $\{I_1, I_2, \ldots, I_M\}$
while i $[1, M]$ **do**
  $j = 0$;
  while $j < N$ **do**
    Randomly sample a $m \times m$ patch P from I_i;
    Calculate the patch variance of P:
      $var_{ave}(P) = [var_R(P) + var_G(P) + var_B(P)]/3$
      if $var_{ave}(P) \geq T_{var}$ **then:**
        Keep this patch as $[P_{i,j}, Q_{i,j}]$;
        $j = j + 1$;
      end if
  end while
end while
Shuffle the $M \cdot N$ patch/score pairs;
Feed them to the CNN model for one epoch of training;

Figure 4 shows the sampling results of the proposed patch sampling strategy using different variance thresholds. Each point of this figure represents a sampled patch's center. The color and shape of these points are used to representing different threshold values. The red-triangle points represent the patch sampling without threshold as the threshold is set to zero. Thus, these points are uniformly distributed on the whole image, in which quite a lot of patches are sampled from homogenous regions. When we increase the variance thresholds, the sampled patches are trends to locate on the regions with more complex structures (non-homogenous regions). However, we found that we cannot use too large threshold values, which will make the selected patches only concentrated on some special regions as the purple-star and orange-diamond markers in Fig. 4. These patches may ruin the CNN model performance as the training data may only come from very limited areas of the training images. Then, too much information may be lost during the patch sampling process. Thus, small variance threshold should be used and threshold selection analysis is provided in the experiments section.

B. Adaptive Stride Scan for Test Patch Generation

After we can train a CNN without using the unreliable data from homogenous regions, we should also avoid using homogenous patches from the tested image for quality score estimation. However, this may not be as straightforward as it may seem by just using the same variance threshold T_{var} to generate the patches from the tested image. It is because we may not obtain sufficient patches from some tested images with large portion of homogenous regions using a fixed sampling stride to scan the image for generating the test patches. Thus, we propose an adaptive stride method for generating the test patches with use of an initial stride (S_{init}) and a minimum of number of test patches (N_{min}). The idea is simple, by scanning the tested image with the initial stride to generate n patches from the tested image based on the same T_{var} that are used in the CNN training. After that we check whether n is large enough for the quality estimation. If n is greater or equal to N_{min}, then we can start the quality score estimation using these n patches. However, if n is lower than N_{min}, then we can scan the tested image again with stride reduce by half such that we can generate more patches from the tested image. Repeat this patch generation process until we can obtain n greater or equal to N_{min} or stride is reduced to 1. With appropriate selection of the parameters T_{var}, S_{init}, N_{min}, it is always possible to generate sufficient patches for quality score calculation. It is because most of the real-world images consist of sufficient portion of non-homogenous regions for quality evaluation.

C. Variance-Based Weighting for Quality Score Estimation

Besides avoiding homogenous patches for quality score estimation, we should also bias to the predicted scores of the patches with higher variances based on the prediction error property as shown in Fig. 3. It is because high-variance patches are more robust with relatively lower prediction errors. Thus, we propose to use a patch variance weighted average for calculating the final image quality score. For n non-homogenous patches $\{P_1, P_2, \ldots, P_n\}$ that are obtained from the tested image using the adaptive stride patch generation, the overall image quality score Q is calculated by

$$Q = \frac{\sum_{j=1}^{n} Q_j var_{ave}(P_j)}{\sum_{j=1}^{n} var_{ave}(P_j)}$$

where Q_j and $var_{ave}(P_j)$ are predicted quality score and patch variance of P_j, respectively. This simple weighted average can make the final quality score bias to the predicted scores of the patches with higher variances. This can improve the robustness of CNN-based NR-IQA.

IV. Experimental Results

We implemented the proposed CNN-based NR-IQA network as shown in Fig. 1 using Keras with TensorFlow as backend and conducted our experiments based on the LIVE database [21] and the TID2013 database [22]. We randomly choose 80% images to construct the training set and the remaining 20% for the test set. Our CNN model is trained based on the patch sampling strategy of Algorithm 1 with 1500 epochs. In NR-IQA, the performance is evaluated by how good the quality score correlates with subjective

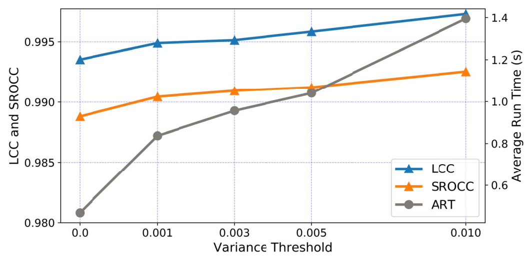


Fig. 5. LCC, SROCC and ART of the quality score estimation against different patch variance thresholds (T_{var}).

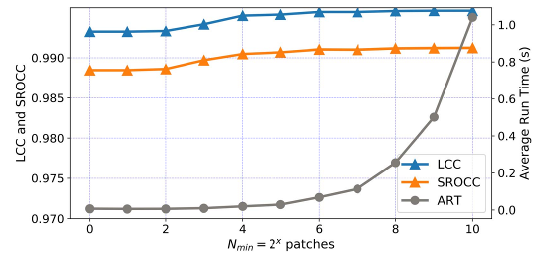


Fig. 6. LCC, SROCC and ART of the quality score estimation against different numbers of N_{min}.

test results, therefore, LCC (Linear Correlation Coefficient) and SROCC (Spearman Rank Order Correlation Coefficient) are used as the performance metrics. In the following experiments, we will first analyze how the variance threshold and minimum number of test patches affect the quality score estimation for parameter selection based on the training set of the LIVE database. We assume that the training set represents the real-world image characteristics. After that we use the test set to evaluate the performance of our algorithm as compared with other well-known methods.

A. Variance Threshold Selection

The patch variance threshold (T_{var}) is an important parameter of the proposed method, which determines complexity of the non-homogenous patches for the CNN training and quality score estimation. Thus, we first analyze how this parameter affect the quality score estimation performance in terms of LCC, SROCC and computational time for score estimation. These results are illustrated in Fig. 5 with T_{var} in the range of 0.0 to 0.010. The curves of LCC and SROCC are the highest values for various settings with only T_{var} is fixed. The trend of these two curves shows that LCC and SROCC are both improving with higher T_{var} but the improvement is becoming insignificant for T_{var} greater than 0.001. This align with our hypothesis that homogenous patches with very low patch variances are unreliable, which cause higher prediction errors. Thus, we cannot use a very low value of T_{var} as too small T_{var} cannot filter out homogenous patches. Based on this observation, we should select T_{var} higher than 0.001. However, too high value of T_{var} will filter out too many patches as some of the useful non-homogenous patches might be removed. Another drawback of large variance threshold is higher computational requirement for the score estimation process. It is because higher T_{var} leads to more densely patches sampling (small stride) in order to obtain sufficient number of patches for the score estimation. The average run time (ART) for a single image quality score estimation against different T_{var} are also shown in Fig. 5, which demonstrate that significantly increase of NRT occur with T_{var} higher than 0.005. Based on these results, 0.005 is chosen for the parameter selection of T_{var}.

B. Minimum Number of Test Patches Selection

Another important parameter is the minimum number of test patches (N_{min}). Too little test patches will cause unreliable quality score estimation, while too many test patches will significantly increase the computational requirement. Thus, we also analyzed how N_{min} affect the quality score estimation performance in terms of LCC, SROCC and computational time for quality score estimation. These results are illustrated in Fig. 6 with N_{min} from 1 to 1024, in which $N_{min} = 2^x$. Based on the curves of LCC and SROCC, we can find that higher N_{min} always achieve better accuracy, but the improvements are becoming not very significant for N_{min} larger than 64 ($=2^6$). However, the quality score estimation run time also increase with N_{min}, especially for N_{min} larger than 128 ($=2^7$). To achieve reasonable low computational requirement as well as good accuracy, we select $N_{min} = 128$ patches as another main setting for our experiments.

TABLE I

LCC and SROCC Performance Comparison Under Different Types of Distortion in LIVE Database

LCC	**JP2K**	**JPEG**	**WN**	**BLUR**	**FF**	**ALL**
FSIM [2]	0.910	0.985	0.976	0.978	0.912	0.960
DIIVINE [8]	0.922	0.921	0.988	0.923	0.888	0.917
CORNIA [11]	0.951	0.965	0.987	0.968	0.917	0.935
Kang [12]	0.953	0.981	0.984	0.953	0.933	0.953
Pre-SM [19]	0.979	**0.986**	0.986	0.917	0.914	0.978
Proposed	**0.990**	0.985	**0.992**	**0.989**	**0.964**	**0.987**
SROCC	**JP2K**	**JPEG**	**WN**	**BLUR**	**FF**	**ALL**
FSIM [2]	0.970	0.981	0.967	0.972	0.949	0.964
DIIVINE [8]	0.913	0.910	0.984	0.921	0.863	0.916
CORNIA [11]	0.943	0.955	0.976	0.969	0.906	0.942
Kang [12]	0.952	**0.977**	0.978	0.962	0.908	0.956
Pre-SM [19]	0.969	0.974	0.978	0.956	0.883	0.974
Proposed	**0.973**	0.968	**0.990**	**0.964**	**0.920**	**0.976**

C. Accuracy Performance on LIVE and TID2013 Datasets

To evaluate our method in a non-distortion specific setting, we performed experiments using five distortion types of LIVE databset [21] and a more challenging dataset of TID2013 [22] with 24 types of distortion. The key parameters of the experiments for the proposed algorithm are $T_{var} = 0.005$, $N_{min} = 128$ and $S_{init} = 128$. Table I shows the experimental results in terms of LCC and SROCC for these five distortion types of JPEG2000 compression (JP2K), JPEG compression (JPEG), White Noise (WN), Gaussian blur (BLUR) and fast fading (FF) in LIVE dataset. In which, we compared the proposed method with one FR-IQA algorithm of FSIM [2], two non-CNN-based NR-IQA algorithms (DIIVINE [8] and CORNIA [11]), and two state-of-the-art CNN-based NR-IQA algorithms (Kang [12] and Pre-SM [19]). The objective of this experiment is to see how the proposed algorithm will perform if we only have images with one particular type of distortion. As shown in Table I, the proposed method outperforms all these well-known methods except JPEG distortion with slightly lower than Kang *et al.* [12] and Pre-SM [19] methods. Moreover, our method achieves excellent results (greater than or equal to 0.99) for white noise distortion and outstanding performance on handling JP2K distortion. In addition, Table II lists and compares the proposed method with three popular FR-IQA methods (PSNR, SSIM and FSIM), four non-CNN-based NR-IQA methods (DIIVINE, BLIINDS-II, BRISQUE and CORNIA) and eight CNN-based NR-IQA methods. Table III presents the performance comparison results based on more challenging TID2013 database for

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT IQA METHODS BASED ON LIVE DATABASE

Methods	**LCC**	**SROCC**
PSNR	0.868	0.873
SSIM [1]	0.913	0.906
FSIM [2]	0.960	0.964
DIIVINE [8]	0.916	0.917
BLIINDS-II [9]	0.930	0.931
BRISQUE [10]	0.940	0.942
CORNIA [11]	0.942	0.935
Kang [12]	0.953	0.956
Li [13]	0.956	0.935
VeNICE [14]	0.960	0.950
Sun [15]	0.958	0.959
Bosse [16]	0.972	0.960
Pan [17]	0.969	0.968
Zuo [18]	0.967	0.964
Pre-SM [19]	0.978	0.974
Proposed	**0.987**	**0.976**

TABLE III

PERFORMANCE COMPARISON OF DIFFERENT IQA METHODS BASED ON TID2013 DATABASE

Methods	**LCC**	**SROCC**
SSIM [1]	0.790	0.742
DIIVINE [8]	0.654	0.549
CORNIA [11]	0.613	0.549
RankIQA+FT [20]	-	0.780
Proposed	**0.890**	**0.878**

four well-known methods including the state-of-the-art method of RankIQA [20]. The two tables shown that our method achieves the highest prediction accuracy in terms of both LCC and SROCC among all compared methods. These results demonstrate that removing homogenous patches and the proposed variance-based weighting for quality score estimation can significantly improve the CNN-based NR-IQA performance.

V. CONCLUSION

In this letter, we reported a special characteristic of CNN-based NR-IQA neural networks, which is small image patches with very low patch variances are not reliable for training and final quality score estimation. In addition, image patch with high variances have much higher chance to achieve better prediction accuracy. Based on this new discovery, we proposed a simple strategy with use of a low patch variance threshold for avoiding homogenous patches in both CNN training and quality score estimation. To bias the score to test patches with higher variances, a variance-based weighting average is also proposed to calculate the final image quality score. Experimental results demonstrated that this new patch variance biased approach can achieve state-of-the-art results on both LIVE and TID2013 databases for NR-IQA. On the other hand, the new discovery of the prediction error characteristic with use of patch variance as homogenous indicator may open a new direction for further development of CNN-based NR-IQA algorithms. It is because we can make use of this characteristic in many different ways to improve the CNN training process as well as the final image quality score calculation.

REFERENCES

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

[2] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[3] H. R. Sheikh, A. C. Bovik, and G. De Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[4] H. R. Sheikh and A. C. Bovik, “A visual information fidelity approach to video quality assessment,” in *Proc. 1st Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, 2005, pp. 23–25.

[5] Z. Wang, H. R. Sheikh, and A. C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2002, p. I-477.

[6] H. R. Sheikh, A. C. Bovik, and L. Cormack, “No-reference quality assessment using natural scene statistics: JPEG2000,” *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[7] T. Brandão, and M. P. Queluz, “No-reference quality assessment of H.264/AVC encoded video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1437–1447, Nov. 2010.

[8] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[9] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[10] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 1098–1105.

[12] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1733–1740.

[13] Y. Li, L.-M. Po, L. Feng, and F. Yuan, “No-reference image quality assessment with deep convolutional neural networks,” in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Beijing, China, Oct. 2016, pp. 685–689.

[14] P. P. Dash, A. Wong, and A. Mishra, “VeNICE: A very deep neural network approach to no-reference image assessment,” in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Toronto, QN, Canada, Mar. 2017, pp. 1091–1096.

[15] C. Sun, H. Li, and W. Li, “No-reference image quality assessment based on global and local content perception,” in *Proc. Vis. Commun. Image Process.*, Chengdu, China, Nov. 2016, pp. 1–4.

[16] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, “A deep neural network for image quality assessment,” in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, Sep. 2016, pp. 3773–3777.

[17] C. Pan, Y. Xu, Y. Yan, K. Gu, and X. Yang, “Exploiting neural models for no-reference image quality assessment,” in *Proc. Vis. Commun. Image Process.*, Chengdu, China, Nov. 2016, pp. 1–4.

[18] L. Zuo, H. Wang, and J. Fu, “Screen content image quality assessment via convolutional neural network,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 2082–2086.

[19] Z. Cheng, M. Takeuchi, and J. Katto, “A pre-saliency map based blind image quality assessment via convolutional neural networks,” in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2017, pp. 77–82.

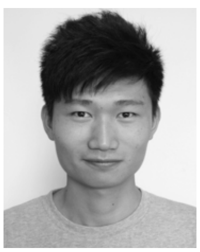
[20] X. Liu, J. van de Weijerand, and A. D. Bagdanov, “RankIQA: Learning from rankings for no-reference image quality assessment,” in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 1040–1049.

[21] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2*. Assessed: Jan. 9, 2019. [Online]. Available: http://live.ece.utexas.edu/research/quality

[22] N. Ponomarenko *et al.*, “Color image database TID2013: Peculiarities and preliminary results,” in *Proc. IEEE 4th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2013, pp. 106–111.

Lai-Man Po (M'92–SM'09) received the B.S. and Ph.D. degrees in electronic engineering from the City University of Hong Kong, Hong Kong, in 1988 and 1991, respectively. He has been with the Department of Electronic Engineering, City University of Hong Kong, since 1991, where he is currently an Associate Professor and the Laboratory Director of the TI Educational Training Centre. He has authored over 140 technical journal and conference papers. His research interests include image and video coding with an emphasis on fast encoding algorithms, new motion compensated prediction techniques, and 3D video processing.

Dr. Po is a member of the Technical Committee on Multimedia Systems and Applications and the IEEE Circuits and Systems Society. He was the Chairman of the IEEE Signal Processing Hong Kong Chapter in 2012 and 2013. He was an Associate Editor of *HKIE Transactions* in 2011 to 2013. He also served on the Organizing Committee, of the IEEE International Conference on Acoustics, Speech and Signal Processing in 2003, and the IEEE International Conference on Image Processing in 2010.



Mengyang Liu received the B.E. degree in optoelectronic engineering from the Shanghai University of Electric Power, Shanghai, China, in 2014, and the M.Sc. degree with a dissertation in electronic and information engineering from the City University of Hong Kong, Hong Kong, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His research interests include image and video processing, video indexing, computer vision, and machine learning.

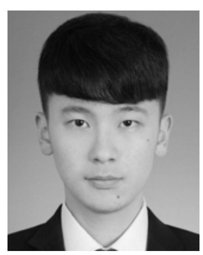
Wilson Y. F. Yuen received the B.Sc. degree (Hons.) in applied computing from the University of Hertfordshire, U.K., and the M.Sc. degree in information engineering from The Chinese University of Hong Kong. He was the Chief Information Officer at Commercial Radio Hong Kong, and taught post-graduate courses at The Hong Kong Polytechnic University, Hong Kong City University, and Hong Kong Baptist University. In 2010, he founded TFI Digital Media Ltd., a Hong Kong-based technology company specializing in video-related technologies. Meanwhile, he also provides advisory services to multiple family offices that are listed on Forbes Hong Kong's Top 50, with interests across real estate, telecom, and hospitality industries.

He has established himself as an Expert in information engineering and user experience design and a Pioneer in multimedia and game development education for over 20 years. His passion toward innovative technology is well-recognized and he was named one of Debrett's Hong Kong 100 Most Influential People (one out of ten individuals under technology and digital category).

He currently serves as a Professor of practice with the Academy of Film, Baptist University of Hong Kong, as an Advisory Committee Member with the City University Department of Electronic Engineering, as a Customer Advisory Board Member with HP Enterprise OEM, and a Mentor and a Vetting Committee member of PolyU Micro Fund and the Committee of Qualifications Framework Industry Training Advisory Committee under the Hong Kong Education Bureau. He is also an Honorary Advisor to the Professional Information Security Association and the Global Design Ambassador of the Interaction Design Foundation.

Yuming Li received the B.Eng. and M.Eng. degrees from the Huazhong University of Science and Technology, in 2011 and 2013, respectively, and the Ph.D. degree from the City University of Hong Kong in 2016. He is currently an Engineer with MiniEye. His research interests include image and video processing, multiscale analysis, deep learning, and autonomous driving.

Xuyuan Xu received the B.E. and Ph.D. degrees in electronic engineering from the City University of Hong Kong, in 2010 and 2014, respectively. He is currently with Tencent Video, Tencent Holdings Ltd., as a Senior Engineer. His research interests include 3D video coding, 3D view synthesis, audio/video fingerprint, and object detection.



Chang Zhou received the B.Sc. degree from Donghua University, Shanghai, China, in 2016, and the master's degree from the City University of Hong Kong, Hong Kong, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His research interests are in computer vision and deep learning.

Peter H. W. Wong (SM'08) received the B.Eng. degree (Hons.) in computer engineering from the City University of Hong Kong in 1996, and the M.Phil. and Ph.D. degrees in electrical and electronic engineering from the Hong Kong University of Science and Technology (HKUST), in 1998 and 2003, respectively. He was a Post-Doctoral Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, from 2003 to 2005. He was with the Applied Science and Technology Research Institute Company Ltd., as a Member of the Professional Staff from 2005 to 2007. From 2007 to 2008, he was a Visiting Assistant Professor with the Department of Electronic and Computer Engineering, HKUST. From 2008 to 2015, he was with Visual Perception Dynamics Labs (Mobile) Ltd., where he worked on RGBW display technology. He is currently the Chief Software Engineer of TFI Digital Media Ltd., where he is working on high dynamic range and wide color gamut video processing. He has authored about 40 technical journal and conference papers. He is the inventor of ten U.S. patents.

Kin Wai Lau received the B.E. degree (Hons.) in information engineering from the City University of Hong Kong in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong. He is currently with TFI Digital Media Ltd., as a Software Engineer. His research interests include image and video processing, video/image retrieval, computer vision, and deep learning.

Hon-Tung Luk received the B.S. degrees in mathematics and in information engineering and the Ph.D. degree in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2010 and 2015, respectively. He is currently with TFI Digital Media Ltd., as a Senior Software Engineer. His research interests include high dynamic range video, video color and contrast enhancement, video quality assessment, and video fingerprintings.