DEEP BLIND IMAGE QUALITY ASSESSMENT USING DUAL-ORDER STATISTICS

Zihan Zhou, Yong Xu, Yuhui Quan, Ruotao Xu

School of Computer Science and Engineering, South China University of Technology, China

ABSTRACT

Deep convolutional neural networks (CNNs) have become a promising approach to blind image quality assessment (BIQA). Existing CNN-based BIQA methods often employ global average pooling (GAP) to aggregate feature maps into a fixed-size representation for regression, so as to handle input images with varying sizes. However, GAP is only capable of extracting the first-order statistics of feature distributions, which is ineffective for distinguishing complex distortions that cause local degradation or preserve global features. To tackle this problem, we introduce the second-order global covariance pooling (GCP) for aggregating feature maps, leading to a more distortion-sensitive and more discriminative global representation. By incorporating GCP and GAP into a ResNet backbone, we propose an effective deep model for BIQA. The experimental results on five BIQA benchmark datasets, including both the synthetic and authentic ones, have demonstrated the excellent performance of the proposed method.

***Index Terms*—** Blind image quality assessment, Feature pooling, Convolutional neural networks, Deep learning

1. INTRODUCTION

Image quality is the main concern in image processing, and blind image quality assessment (BIQA) techniques are developed to automatically evaluate the perceptual quality of the distorted images. Given an input image, BIQA aims at estimating its perceived quality without accessing any reference information of the image. Such a topic has gained extensive attention from both industries and academics.

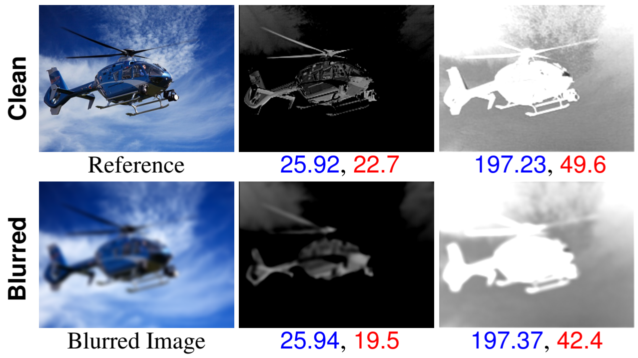
In general, there are two phases in a BIQA approach: quality-aware feature extraction and quality score regression. Traditional approaches usually adopt hand-crafted designs for feature extraction, such as natural scene statistics [1, 2] and variants of local binary patterns [3, 4]. Then the score regression is done by some well-established learning-based model, such as support vector regression, random forest, and Gaussian process. Recently, inspired by the success of CNNs in image recognition, a series of CNN-based BIQA approaches have been proposed, where the two phases are jointly optimized in an end-to-end manner. With the data-driven convolutional-layer-based feature extractor and fully-connected-layer-based score regressor, these approaches have shown impressive performance.

Typical CNNs in computer vision often connect convolutional layers with fully-connected (FC) layers. The FC layers encode feature maps into a single quality-aware representation and act as a regressor. Such a structure can only deal with images of a fixed size. Rescaling, one of the most often-used solutions to this problem in image recognition, is not suitable for IQA, as rescaling can obviously change the visual quality of an image. Therefore, global pooling, which aims to aggregate the score maps or feature maps with different spatial sizes into a single score or a fixed-length global vector, is especially critical for an IQA-oriented CNN model. A line of researches [5, 6, 7] proposed to use score pooling strategies, where the pooling module is designed to aggregate the predicted scores on local regions/patches. However, due to the absence of the real supervision to the visual quality of patches, it is hard to guarantee the accuracy of local quality scores, which limits the performance of such score-pooling-based methods. Another pooling strategy that avoids this issue is to operate global pooling on quality-aware feature maps to generate a fixed-length image representation for regression. Several different pooling mechanisms are adopted in this strategy, such as global averaging pooling (GAP) [8, 9] and other average-based pooling [10, 11, 12].

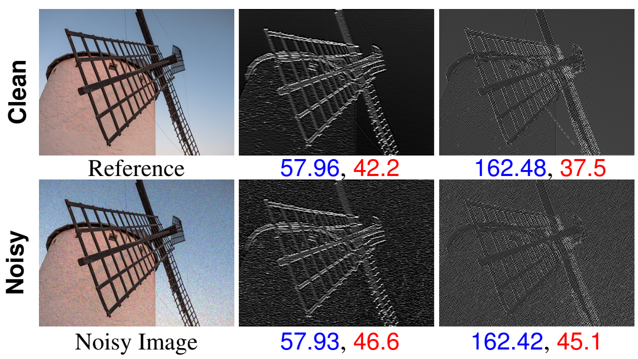
The GAP and average-based pooling mechanisms essentially capture the first-order statistics of a feature distribution. While such average-related statistics can summarize the holistic changes caused by distortions, they are insensitive when handling the distortions that cause little changes to the image features on average. Such distortions, called global-preserving distortions for convenience, are not rare in the real world. For instance, noise degradation which oscillates pixel values is often expected to preserve the average value due to the zero-mean property of the noise. Another often-seen example is image blurring, which also causes little changes to average intensity level due to the sum-to-one property of blur kernels. Since a BIQA-oriented CNN is expected to preserve or propagate the effects caused by distortions for accurate perdition of the image quality score, GAP will be likely to reduce the effectiveness of a BIQA-oriented CNN in

Corresponding author: Ruotao Xu (xrt@scut.edu.cn).

This work is supported by National Natural Science Foundation of China (Grant No. 62106077, 62072188 and 61872151) and Postdoctoral Foundation of China (Grant No. 2020M682705).



(a) Clean-Blurred image pairs



(b) Clean-Noisy image pairs

Fig. 1: Examples of two pairs of images (the first column) and their corresponding paired feature maps (the last two columns) outputted by the conv_3 block of pre-trained ResNet-101. The blue and red digits denote the mean and standard variance values of feature maps respectively.

identifying degradation effects.

See Fig. 1 for an illustration, where the feature maps of noisy/blurred images have little changes in terms of global average, compared to those of their clean versions. Furthermore, for some local distortions, the differences in the feature maps only happen in local regions. The changes, though varying in the corresponding small regions, may be diminished by an average operation and produce similar average values. The above analysis motivates us to exploit other statistics for global feature pooling in a BIQA-oriented CNN so as to identify the degradation effects more effectively.

In this paper, we take a step towards addressing CNN-based BIQA by considering the second-order statistics for global pooling mechanisms. The introduction of the second-order statistics (*e.g.* covariance) to a CNN-based method can lead to a better description for the feature distribution of a distorted image. Consider the average-insensitive distortions such as white noise and Gaussian blur. Although usually preserving the mean, they change the covariance, *i.e.*, noise usually results in higher variances due to its oscillation nature, while blurring often leads to lower variance due to its smoothing effect. Furthermore, second-order statistics are usually more sensitive to outliers. Due to the quadratic calculation, the large changes in a local region can be exposed during the feature pooling. In addition, though not explored in the BIQA realm, the second-order pooling, which captures the self and cross-channel similarities information of feature maps, has proven its ability to generate informative representations and describe complex classification boundary [13] in recognition. Its success in recognition also inspires us to explore the second-order statistics in BIAQ for the complex nature of the human visual system.

To capture the second-order statistics, we introduce the global covariance pooling (GCP) [13] for feature aggregation. Combining GCP with a first-order global pooling implemented by GAP, we propose an end-to-end framework with a dual-order global pooling mechanism for BIQA. The effectiveness of the proposed method is validated on five widely-used IQA benchmark datasets including both synthetic and real-world ones.

2. PROPOSED METHOD

In the proposed method, we construct a two-branch CNN model for blind image quality score prediction. The model accepts an image of an arbitrary size, since a global pooling is employed in both branches. The first branch includes a first-order global pooling module and is expected to process average-sensitive distortions mainly, while the second one, which includes a second-order pooling module, is expected to process average-insensitive distortions. The architecture of the proposed model is illustrated in Fig. 2, whose details are described in the following.

2.1. First-order Global Pooling

Let $\mathcal{I} \in \mathbb{R}^{H_0 \times W_0 \times C_0}$ denote the input image for assessment, and $\mathcal{X} = f(\mathcal{I}) \in \mathbb{R}^{H \times W \times C_1}$ denote the extracted features. Given the extracted feature $\mathcal{X}$ with an arbitrary spatial size $H \times W$, the global pooling mechanism is designed to aggregate $\mathcal{X}$ into a quality-aware global feature $\boldsymbol{y}$, which is then fed into the regressor for quality prediction. In the pooling scheme, the feature tensor $\mathcal{X}$ is treated as a set of samples $\{\boldsymbol{x}_i\}_{i=1}^{N}$ with $N = HW$ and the sample $\boldsymbol{x}_i \in \mathbb{R}^{C_1}$ is the feature vector located at i-th position of $\mathcal{X}$.

The first-order statistic is important to characterize the feature distribution. In a BIQA task, the first-order statistics, often estimated with average or weighted average, are discriminative for the average-sensitive distortions, such as color shifts and brightness changes. To measure the first-order statistics of feature distribution, we introduce the GAP for feature aggregation, which is simply calculated by the average of features:

$$\mu(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i. \tag{1}$$

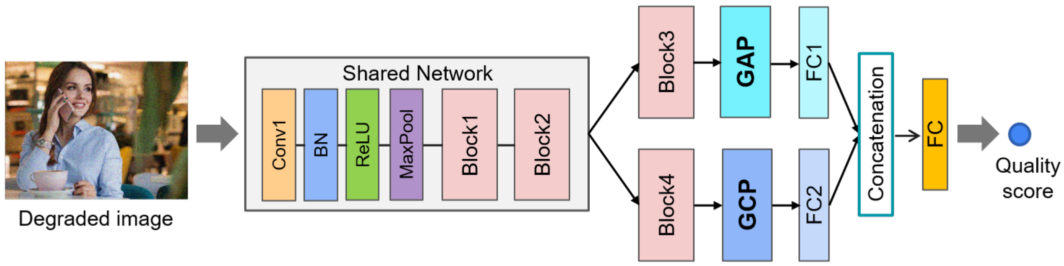


Fig. 2: Framework of the proposed method. The backbone of the network is based on ResNet.

2.2. Second-order Global Pooling

Although GAP can handle the average-sensitive distortions, it usually fails to capture the changes of features caused by average-insensitive distortions, such as blur and zero-mean noise. Furthermore, for some spatial distortions, such as color blocks and non-eccentricity patches, the quality drop only happens in a small region. GAP tends to weaken or even eliminate local changes due to the global average.

The second-order statistics can be beneficial in such cases. We introduce the GCP for feature aggregation, which further captures the second-order statistics of feature distribution. Given the feature samples $\{\boldsymbol{x}_i\}_{i=1}^N$, GCP estimates the covariance of the samples as

$$\Sigma(\mathcal{X}) = \sum_{i=1}^{N} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top, \tag{2}$$

where $\bar{\boldsymbol{x}}$ denotes the mean of samples, *i.e.* $\bar{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{x}_i / N$.

It is known that normalization is beneficial for the robust covariance estimation on high-dimensional low-sample-size data [14]. Encouraged by the promising performance of matrix square-root normalization in classification tasks [13], we adopt it for the GCP:

$$\Sigma^{\frac{1}{2}}(\mathcal{X}) = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{U}^\top, \tag{3}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are the orthogonal matrix of eigenvectors and the diagonal matrix of eigenvalues of $\Sigma(\mathcal{X})$, respectively.

2.3. Dual-Order BIQA Framework

The GAP and GCP use different feature maps as inputs, as they are designed to process different kinds of distortions, *i.e.*, average-sensitive and average-insensitive distortions, respectively. Given the input image $\mathcal{I}$, the feature extraction procedure can be formulated as follows:

$$\mathcal{X}_1 = f_1 \circ f_0 \circ \mathcal{I}, \tag{4}$$
$$\mathcal{X}_2 = f_2 \circ f_0 \circ \mathcal{I}, \tag{5}$$

where $\circ$ denote function composition, $\mathcal{X}_1$, $\mathcal{X}_2$ denote the input features for GAP and GCP modules respectively, f_0 denotes the shared feature extractor (*i.e.* Shared Network in Fig. 2) for low-level features, and f_1, f_2 are average-sensitive and average-insensitive distortion-aware feature extractors respectively (*i.e.* Block3 and Block4 in Fig. 2).

The GAP and GCP aggregate the feature maps into global feature vectors respectively, which are then fed to FC layers for score prediction:

$$s_1 = g_1 \circ \mu \circ \mathcal{X}_1, \tag{6}$$
$$s_2 = g_2 \circ \Sigma^{\frac{1}{2}} \circ \mathcal{X}_2, \tag{7}$$

where g_1, g_2 are the FC layer following GAP, GCP respectively, and s_1, s_2 are the corresponding predicted scores for average-sensitive/insensitive distortions. The final quality score s is calculated as the weighted sum of s_1 and s_2 via a two-neuron FC layer:

$$s = \omega_1 s_1 + \omega_2 s_2, \tag{8}$$

where ω_1 and ω_2 is a pair of learned weights.

Given a set of images $\{\mathcal{I}_i\}_{i=1}^D$ and their subjective scores measured by human $\{s_i^*\}_{i=1}^D$. Let $\{s_i\}_{i=1}^D$ denote the scores predicted by the proposed model. We use the Huber loss for training due to its stronger robustness to outliers over the commonly-used mean square error loss, which is defined as

$$\ell = \sum_{i=1}^{D} \ell_\delta(s_i, s_i^*), \tag{9}$$

where ℓ_δ is the paramterized Huber loss defined by

$$\ell_\delta(s, s^*) = \begin{cases} \frac{1}{2}(s - s^*)^2, & \text{for } |s - s^*| \le \delta \\ \delta\left(|s - s^*| - \frac{1}{2}\delta\right), & \text{otherwise} \end{cases}. \tag{10}$$

And δ is a parameter to choose the way to penalty outliers. In implementation, we set $\delta = 1/9$ as suggested in [11].

3. EXPERIMENTS

3.1. Experimental Setups

Implementation Details. Following [11, 15], we select ResNet-101 [16] as our network backbone due to its effectiveness in feature extraction. In detail, we employ the first

3 building blocks (conv1, conv2_x, conv3_x) in ResNet-101 as the shared feature extraction network $f_0(\cdot)$, the 4-th building block (conv4_x) for the feature extraction networks $f_1(\cdot)$, $f_2(\cdot)$ in two branches. These blocks are all initialized with weights of ResNet-101 trained on ImageNet. During training, the weights of $f_0(\cdot)$ is frozen, while other blocks are optimized with the loss in (9). A two-stage training strategy is adopted. The network without the GCP branch is first trained. Then, the whole network and the paired parameters ω_1, ω_2 are jointly trained with frozen $f_1(\cdot)$.

Datasets. In order to validate the performance of the proposed method, five publicly available natural image quality databases are employed in our experiments, including three artificially distorted sets (LIVE [17], TID2013 [18] and Kadid-10K [19]) and two realistically distorted sets (LIVE-C [20] and KonIQ-10K [21]). The images in their original resolution are fed into the model as input to test how well the model generalizes to pictures of arbitrary sizes. During training, we follow [8, 11] to randomly sample 80% of the images in each database for training and leave the rest for testing. Regarding the synthetically-distorted datasets, we split training and test sets according to the reference images such that content is not intersected between the two sets. Dynamic data augmentation comprising horizontal flip, vertical flip, and rotation of $\pm 3°\text{C}$ is randomly applied to the training images. Since rotation produces extra black borders, we have removed the excess area by cropping.

Evaluation Criteria. Two commonly used evaluation metrics for performance comparison are adopted, including Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SRCC). The SRCC measures the prediction monotonicity, PLCC measures the linear correlation. An effective IQA metric is expected to yield high values of PLCC and SRCC.

Compared Methods. Deep BIQA models have been frequently reported to outperform traditional knowledge-driven BIQA methods, such as NIQE [22] and ILNIQE [23]. Thus we only compared the proposed model against recently State-of-the-Art DNN-based BIQA methods, including PQR [24], deepIQA [5], DBCNN [25], MetaIQA [26], CaHDC [12], HyperNet [9], SiamIQA [27] and AIGQA [28]. The experimental results of compared methods are based on implementations obtained from the respective authors or just copied from the original papers.

3.2. Performance Evaluation

The results on synthetically distorted databases are reported in Table 1. It can be seen that the proposed method significantly outperforms the compared methods on datasets TID2013, Kadid-10K, and achieves comparative performance on dataset LIVE. Note that all methods achieve nearly perfect results (over 0.95 on PLCC, SRCC) on LIVE. In comparison, TID2013 and Kadid-10K which contain more samples and complex distortions, are usually more challenging and valuable for performance evaluation.

Table 1: Median SRCC and PLCC results across ten sessions on the test sets of the synthetically distorted IQA databases. **Bold** on digits denote the best result for each criteria.

	LIVE		TID2013		Kadid-10K	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
PQR [24]	**0.965**	**0.971**	-	-	-	-
deepIQA [5]	0.954	0.963	0.761	0.787	0.628	0.647
DBCNN [25]	0.946	0.959	0.816	0.865	-	-
MetaIQA [26]	-	-	-	-	0.767	0.774
CaHDC [12]	**0.965**	0.964	0.862	0.878	-	-
HyperNet [9]	0.962	0.966	-	-	-	-
SiamIQA [27]	0.961	-	0.855	-	-	-
AIGQA [28]	0.963	0.957	0.871	0.893	0.864	0.863
Ours	0.957	0.961	**0.946**	**0.948**	**0.952**	**0.953**

The results on authentically databases are reported in Table 2. It is shown that our method performs significantly better than all compared methods on the large-scale dataset KonIQ-10K and is comparative on LIVE-C. Note that KonIQ-10K contains over 10 thousand realistically and complexly distorted images, which are usually more challenging for BIQA evaluation.

Table 2: Median SRCC and PLCC results across ten sessions on the test sets of the authentically distorted IQA databases. **Bold** on digits denote the best result for each criteria.

	LIVE-C		KonIQ-10K	
	SRCC	PLCC	SRCC	PLCC
PQR [24]	0.857	0.882	0.881	0.884
deepIQA [5]	0.671	0.686	0.797	0.805
DBCNN [25]	0.851	0.869	0.875	0.884
MetaIQA [26]	0.802	0.835	0.851	0.887
CaHDC [12]	0.738	0.744	-	-
HyperNet [9]	**0.859**	**0.882**	0.906	0.917
SiamIQA [27]	0.851	-	0.894	-
AIGQA [28]	0.751	0.761	-	-
Ours	0.834	0.852	**0.911**	**0.923**

To evaluate the generalization ability of our method, we conduct a cross-dataset experiment by using the KonIQ-10K as the training set and LIVE-C as the test set. Only compared methods with available results are reported. The results are shown in Table 3, which indicates that our proposed model has a good generalization capability, especially for predicting the quality for pictures of arbitrary resolution and of real-world complex distortions.

3.3. Ablation Study

To analyze the effectiveness of utilizing GCP with GAP module, we conduct an ablation experiment to validate the effec-

Table 3: SRCC comparison between CNN-based BIQA models trained on KonIQ-10K and tested on LIVE-C without fine-tuning. **Bold** on digits denote the best result.

PQR [24]	DBCNN [25]	HyperNet [9]	**Ours**
0.772	0.755	0.785	**0.809**

tiveness of components in the proposed model. The following model setting is constructed: 1) The model without the GCP branch, named GAP. 2) The model without the GAP branch, named GCP. 3) The model with both branches, *i.e.* the proposed model, name GAP+GCP. The results on TID2013, LIVE-C, KonIQ-10K are reported in Table 4. It can be seen that the combination of GAP and GCP shows better performance than other single pooling strategies for both realistic and artificial distortions, which indicates effectiveness of the dual-order statistic global pooling for BIQA.

Table 4: Median SRCC and PLCC results across ten sessions of ablation study on the test sets of three IQA databases. **Bold** on digits denote the best result for each criteria.

Setting	TID2013		LIVE-C		KonIQ-10K	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
GAP	0.929	0.917	0.801	0.824	0.909	0.912
GCP	0.916	0.911	0.784	0.798	0.893	0.906
GAP+GCP	**0.946**	**0.948**	**0.834**	**0.852**	**0.911**	**0.923**

To further investigate the effectiveness of the proposed dual-order pooling mechanism on different types of distortions, we select several often-seen types of distorted images from Kadid-10K dataset and summarize the testing SRCC values in Table 5. The results are compared with the GAP-only baseline to demonstrate the improvements brought by our dual-order pooling. From Table 5, obvious improvements (over 0.025) can be observed on the average-insensitive distortions, *e.g.* different types of blur, noise, and local spatial distortions. However, for the average-sensitive distortions, *e.g.* brighten, darken and mean shift, the proposed pooling mechanism only brings slight improvements (less than 0.020).

4. CONCLUSION

In this work, we introduced the global covariance pooling module for the deep BIQA, so as to characterize the second-order statistics of feature distribution and produce a quality-aware global feature to encode the average-insensitive distortions. Incorporating with the global average pooling, we construct a two-branch deep neural network for quality score prediction of degraded images with various distortion types. Experimental results on five benchmark IQA datasets have demonstrated the efficiency of the proposed method as well the dual-order pooling mechanism. In future, we would like to investigate the potentials of higher-order statistics in IQA tasks. Furthermore, inspired by the effectiveness of GCP, we will make a step further by seeking better characterization for the functional relationship that exists between channel consistencies and subjective quality scores.

Table 5: Median SRCC results across ten sessions of ablation study on Kadid-10K in terms of different distortion types.

Distortion types		GAP	GAP+GCP	*Gain*
Blur	Gaussian blur	0.948	**0.981**	0.032
	Motion blur	0.926	**0.975**	0.049
Noise	White noise	0.928	**0.973**	0.044
	Impulse noise	0.939	**0.976**	0.036
Spatial distortion	Pixelate	0.929	**0.974**	0.045
	Color block	0.943	**0.973**	0.029
Brighten change	Brightness	0.954	**0.972**	0.019
	Darken	0.952	**0.970**	0.018
	Mean shift	0.957	**0.974**	0.017

5. REFERENCES

[1] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[2] Deepti Ghadiyaram and Alan C Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. vision*, vol. 17, no. 1, pp. 32–32, 2017.

[3] Min Zhang, Chisako Muramatsu, Xiangrong Zhou, Takeshi Hara, and Hiroshi Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Letters*, vol. 22, no. 2, pp. 207–210, 2014.

[4] Qiaohong Li, Weisi Lin, and Yuming Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Letters*, vol. 23, no. 4, pp. 541–545, 2016.

[5] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, 2018.

[6] Jie Gu, Gaofeng Meng, Shiming Xiang, and Chunhong Pan, "Blind image quality assessment via learnable attention-based pooling," *Pattern Recognit.*, vol. 91, pp. 332–344, 2019.

[7] Da Pan, Ping Shi, Ming Hou, Zefeng Ying, Sizhe Fu, and Yuan Zhang, "Blind predicting similar quality map for image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 6373–6382.

[8] Hossein Talebi and Peyman Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, 2018.

[9] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 3664–3673.

[10] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik, "From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 3575–3585.

[11] Yicheng Su and Jari Korhonen, "Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling," in *Proc. Int. Conf. Image Process.*, 2020, pp. 191–195.

[12] Jinjian Wu, Jupo Ma, Fuhu Liang, Weisheng Dong, Guangming Shi, and Weisi Lin, "End-to-end blind image quality prediction with cascaded deep neural network," *IEEE Trans. Image Process.*, vol. 29, pp. 7414–7426, 2020.

[13] Peihua Li, Jiangtao Xie, Qilong Wang, and Wangmeng Zuo, "Is second-order information helpful for large-scale visual recognition?," in *Proc. IEEE Conf. Comput. Vision*, 2017, pp. 2070–2078.

[14] Qilong Wang, Jiangtao Xie, Wangmeng Zuo, Lei Zhang, and Peihua Li, "Deep CNNs meet global covariance pooling: Better representation and generalization," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2582–2597, 2021.

[15] Aladine Chetouani, "Image quality assessment without reference by mixing deep learning-based features," in *Proc. IEEE Conf. Multimedia Expo*, 2020, pp. 1–6.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.

[17] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006.

[18] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Process. Image Comm.*, vol. 30, pp. 57–77, 2015.

[19] Hanhe Lin, Vlad Hosu, and Dietmar Saupe, "KADID-10k: A large-scale artificially distorted IQA database," in *Proc. IEEE Conf. Quality Multimedia Exp.*, 2019, pp. 1–3.

[20] Deepti Ghadiyaram and Alan C Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, 2015.

[21] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4041–4056, 2020.

[22] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[23] Lin Zhang, Lei Zhang, and Alan C Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, 2015.

[24] Hui Zeng, Lei Zhang, and Alan C Bovik, "A probabilistic quality representation approach to deep blind image quality prediction," *CoRR*, vol. abs/1708.08190, 2017.

[25] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, 2018.

[26] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi, "MetaIQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 14143–14152.

[27] Weixia Zhang, Kede Zhai, Guangtao Zhai, and Xiaokang Yang, "Learning to blindly assess image quality in the laboratory and wild," in *Proc. IEEE Intel. Conf. Image Process.*, 2020, pp. 111–115.

[28] Jupo Ma, Jinjian Wu, Leida Li, Weisheng Dong, Xuemei Xie, Guangming Shi, and Weisi Lin, "Blind image quality assessment with active inference," *IEEE Trans. Image Process.*, vol. 30, pp. 3650–3663, 2021.