# Combination of Deep Learning-based and Handcrafted Features for Blind Image Quality Assessment

Aladine Chetouani, Maurice Quach, Giuseppe Valenzise, Frédéric Dufaux

# COMBINATION OF DEEP LEARNING-BASED AND HANDCRAFTED FEATURES FOR BLIND IMAGE QUALITY ASSESSMENT

*Aladine Chetouani*[1],[2]     *Maurice Quach*[2]     *Giuseppe Valenzise*[2]     *Frédéric Dufaux*[2]

[1]Laboratoire PRISME, Université d'Orléans, Orléans, France
[2]L2S, Centrale Supélec, Université Paris-Saclay, Gif-sur-Yvette, France

## ABSTRACT

This last decade, a plethora of handcrafted-based image quality metrics has been proposed in the literature. Some of them are based on structural analysis, while some others exploit mutual information or perceptual characteristics. Nowadays, deep learning-based methods are widely used in several domains due to its ability to well fit the target directly from the image. In this paper, we study the impact on the performance of combining handcrafted and Deep Learning-based (DL) features, since each of them extracts specific information. Indeed, DL-based image quality assessment methods often extract local information by extracting small patches, while the handcrafted ones provide global information through a global analysis. We analyzed the performance before and after combining the two using bilinear pooling strategy. Experimental results on commonly used datasets show the relevance of combining both approaches.

***Index Terms***— Image Quality, Deep learning-based Features, Handcrafted Features, Bilinear Pooling

## 1. INTRODUCTION

Quality of multimedia contents is a sensitive notion in several applications such as computer vision where the performance depends on the quality of the data, medical imaging where the region-of-interest must be not affected by any distortion, biometrics where the data collected must allow to well identify the person, etc. To answer this request, numerous studies were carried out in order to propose efficient metrics that ensure a high quality of experience. Existing metrics can estimate the quality with reference (FR), without reference (NR) or by extracting only some information from the reference image (RR). In this article, we focus on the NR approach since the latter is more applicable in most real applications.

Traditional NR metrics employ handcrafted features to characterize the distortion. For instance, the perceptual quality of blurred images is often estimated by analyzing edges [1], while the perceptual impact of block-based processing such as JPEG compression is often measured by analyzing the neighborhood of blocks [2]. This kind of metrics is dedicated to specific distortions that limits their use. To overcome this limitation, multi-degradation metrics have been then proposed. The latter are used to estimate the quality of any distorted image (blocking, blurring, ringing, noise, combined degradations, etc.). These metrics are generally based on two steps: feature extraction and feature combination through a machine learning algorithm such as Artificial Neural Networks (ANN) [3] or Support Vector Machine [4].

Nowadays, a large part of the published work focuses on Deep Learning-based methods. This is mainly due to their performance and their ability to automatically learn features from the raw data. In [5], the authors used their own CNN model to estimate the quality of 2D images. The idea is quite simple and provides promising results. An RGB color image is first decomposed into patches of size $32 \times 32 \times 3$. After a local normalization, these patches are used as inputs to a CNN model and the overall quality score is computed by averaging the predicted patch scores. In [6], the method was extended to predict the quality and degradation type simultaneously. The two previous methods use the subjective score of the whole image, often called Mean Opinion Score (MOS), as the target (i.e. desired output) of each patch without considering additional information from the whole image. The authors assumed that the patches extracted from a given image have the same perceptual impact which is not totally in accordance with our Human Visual System (HVS). In order to overcome this assumption, some studies proposed to consider perceptual characteristics. In [7], the authors considered rather saliency information as patch selector. In [8, 9], the author considered rather scanpath. In [10, 11, 12], a similar approach was applied for quality assessment of 3D meshes.

In this work, we study the impact on the performance of combining both approaches (i.e. handcrafted and Deep Learning-based (DL) features), since each of them extracts specific information. Indeed, DL-based image quality assessment methods often extract local information, while the handcrafted features provide global information. The idea is to compensate the lack of these patch-based methods by integrating global information from the whole image. To this end, we first extract DL-based feature vectors for each patch of the image as well as a handcrafted feature vector from the whole
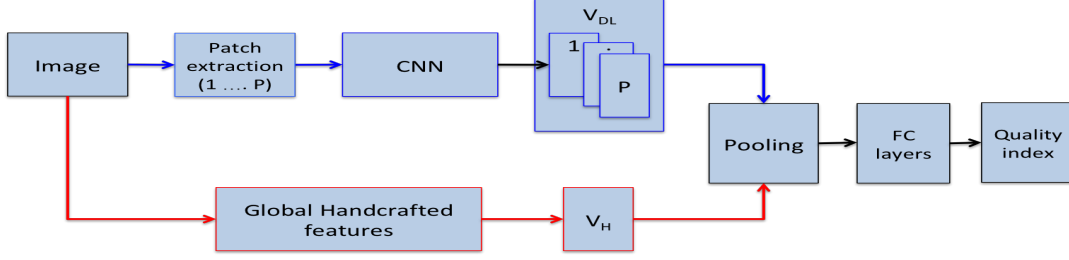
**Fig. 1**. Flowchart of the proposed method.

image. Then, we combine those vectors through the Bilinear Pooling (BP) strategy. The resulting vectors are finally fed to a CNN model to predict the quality. Our method was evaluated on well known databases and the results showed the improvement given by the proposed strategy. It is worth noting that in [13], the author studied the impact of combining DL-based features.

Our paper is organized as follows: In Section 2, we present the proposed method by describing the CNN model and the handcrafted features used. In Section 3, we show the results obtained in terms of correlation with the subjective judgments and we discuss the impact of the combination step. The last section is dedicated to the conclusion.

## 2. PROPOSED METHOD

The pipeline of the proposed method is summarized in Fig. 1. For a given image $\mathbf{I}$, we first extract handcrafted features $V_H(I)$ from the whole image and DL-based feature vectors $V_{DL}(I, p)$ for each patch $\mathbf{p}$ of the image $\mathbf{I}$ (i.e. one $V_{DL}(I, p)$ per patch). The extracted vectors are then pooled using BP strategy and the resulting vectors are fed to Fully Connected (FC) layers. The overall quality index is finally given by averaging the predicted patch scores.

### 2.1. Handcrafted features

A lot of handcrafted features has been used to estimate the quality. Some of them are based on structural information [14], while some others exploit degradation types [4] or even statistical information [15]. In this study, we focus more on statistical features, since damage caused by distortions imply an alteration of the statistics of the image. Here, we employ statistical features based on Gradient Magnitude (**GM**) and Laplacian of Gaussian (**LOG**) that have been widely used in computer vision for image matching and were efficiently used for image quality in [16]. More precisely, we employed the marginal probability distributions of $GM$ and $LOG$ as well as their independency measures. The former allow to get information about the statistical distribution of the image, while the latter give information about the statistical interaction between **GM** and **LOG**. All these features are computed

after applying a joint adaptive normalization that aims to remove content-dependent characteristics. For more details, the reader is referred to [16] where the authors well demonstrated the impact on the performance of each set of features. At the end, 10 features are extracted for each statistical set. So, our handcrafted feature vector $V_H$ is constituted of 40 features, extracted from the whole image.

### 2.2. Deep Learning-based features

Several CNN models have been employed in the literature with different strategies (i.e. from scratch, transfer learning or fine-tuning). In this work, DL-based features are extracted from a modified version of VGG that was fine-tuned [17]. This model, initially pre-trained on ImageNet [18], was proposed in 2014 by the Oxford Visual Geometry Group. Its initial input is an image of size 224x224x3, while its output is composed of 1000 classes. For a given image, $\mathbf{P}$ feature vectors of size 128 are thus extracted from the image $I$ (i.e $[V_{DL}(I, 1), V_{DL}(I, 2), ..., V_{DL}(I, P)]$).

### 2.3. Bilinear Pooling

To combine the handcrafted and DL-based features, Bilinear Pooling (**BP**) is applied [19]. It allows a multiplicative interaction between all features of the two vectors and thus improves the interaction between the local (i.e. DL-based) and global (i.e. handcrafted) features. It consists of applying the outer product of two vectors as follows:

$$VBP_{I,p}(V_{DL}, V_H) = V_{DL}(I, p) \otimes V_H(I) \qquad (1)$$

where $V_{DL}(I, p)$ is the DL-based feature vector of the patch $\mathbf{p}$ of the image $\mathbf{I}$. $V_H(I)$ is the handcrafted feature vector extracted from the image $\mathbf{I}$. $\otimes$ represents the outer product.

Before combining the vectors, we projected the DL-based features into a lower dimensional space. This makes it possible to reduce the number of learnable parameters and above all, to have a balanced representation of the features (i.e. handcrafted and DL-based). For that, the well-known Count Sketch (CS) projection function is used to obtain a vector $VP_{DL}$ of size $D$ from a vector $\mathbf{V}$ of size $N$ with $D < N$ [20]. Fig. 2 summarizes this procedure.
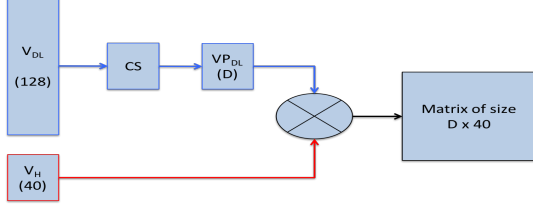
**Fig. 2**. Bilinear Pooling applied between the handcrafted and DL-based features.

The resulting matrix is of size $D \times 40$ that is then vectorized and fed to three FC layers. The first FC layer is followed by a ReLu layer and a dropout layer, while the second FC layer is followed by only a ReLu layer. The third FC layer is a logistic regression layer with one output (i.e. predicted MOS). The model was trained through the Stochastic Gradient Descent (SGD) optimization method using the MSE as loss function. The learning rate and momentum were set to 0.1 and 0.9, respectively. The batch size was fixed to 32 and the number of epochs was equal to 20. The model was saved after each epoch and the training set was shuffled. The model that provided the best result was retained. The overall quality index was finally obtained by averaging all the predicted patch scores.

### 2.4. Datasets

Our experiments were carried-out on four well-known datasets:

- **LIVE - Phase 2 (LIVE2-P2) [21]:** LIVE-P2 dataset is one of the first publicly available datasets and is still commonly used to evaluate existing metrics. It contains 29 pristine images from which 982 degraded images were derived. Five different degradation types were considered: JPEG2000, JPEG, White Noise, Gaussian Blur and Fast Fading. The DMOS (Differential Mean Opinion Score) was used to annotate the degraded images (0 indicates the best quality, while 100 indicates the worst quality).

- **TID 2008 (TID08) [22]:** Constituted of 17 degradation types, TID08 dataset consists of 1700 degraded images obtained from 25 pristine images (i.e. 100 degraded images per degradation) and their corresponding MOS (Mean Opinion Score). The latter varies between 0 and 9 where 0 denotes the worst quality and 9 the best quality.

- **TID 2013 (TID13) [23]:** TID13 dataset is an extended version of the previous one. More degradation types were considered (24 instead of 17) with more degraded images per degradation type (125 instead of 100). A total of 3000 degraded images and their corresponding MOS are provided.

- **CSIQ [24]:** This dataset is composed of 866 degraded images achieved from 30 pristine images and 6 degradation types (JPEG2000, JPEG, White Noise, Contrast, Gaussian Blur and Pink Gaussian Noise). The normalized DMOS are given for each degraded image. This value varies between 0 and 1, where 0 corresponds to the best quality and 1 the worst quality.

## 3. EXPERIMENTAL RESULTS

### 3.1. Evaluation Criteria

The performance evaluation is carried out by computing the Pearson (**PCC**) and Spearman (**SROCC**) correlation coefficients. These values are calculated between subjective scores and the corresponding predicted ones. A correlation of 1 means a perfect prediction, while a correlation of 0 indicates no correlation. The predicted scores are mapped to the subjective ratings using the following non-linear logistic function:

$$Q = \beta_1 \left( \frac{1}{2} - \frac{1}{e^{-\beta_2(Q_p - \beta_3)}} \right) + \beta_4 * Q_p + \beta_5 \qquad (2)$$

where $Q_p$ and $Q$ are the predicted and the mapped scores, respectively.

### 3.2. Individual Evaluation

Our method is first evaluated by randomly splitting the LIVE-P2 database into 60%, 20% and 20% for the training, validation and test sets, respectively. This procedure is repeated 10 times to ensure the generalization ability of our method and the mean correlation across the 10 iterations is computed and used as performance indicator. We compare the performance of each extracted features individually. Table 1 presents the PCC and SROCC obtained. As expected, DL-based features achieve better results (PCC=0.967) than the handcrafted ones (PCC=0.955). The combination of the two allows an improvement of 1.14% and 2.41% compared to the DL-based and handcrafted features, respectively. Therefore, the integration of global information through the handcrafted features has a benefit impact on the performance.

**Table 1**. Mean PCC and SROCC of the handcrafted and DL-based features on the LIVE-P2. The best results are represented in bold.

| Features | PCC | SROCC |
|---|---|---|
| *Handcrafted* | 0.955 | 0.951 |
| *DL* | 0.967 | 0.971 |
| *BP(DL,Handcrafted)* | **0.978** | **0.977** |

Fig. 3 shows the PCC of the CNN model (red line) for each iteration as well as the PCC obtained after combination

of the two features (blue line) across the 10 iterations. As can be seen, the combination allows an improvement for each iteration. The mean, the maximum and the minimum PCC improvements are equal to 0.010, 0.013 and 0.008 respectively.
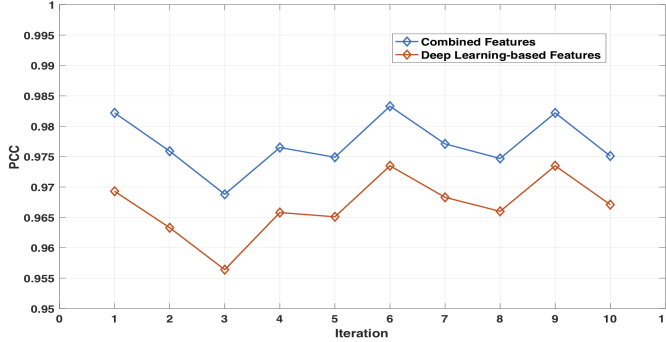


**Fig. 3**. PCC of the DL-based features (red line) and the combined ones (blue line).

A one-way ANalysis Of VAriance (ANOVA) test is carried out to show the statistical significance of improvement gain. Fig. 4 presents the boxplot of the predicted and subjective scores. This plot depicts the distribution of our data around their median values. It can be seen that the predicted scores of the combined features and subjective scores have close distributions with similar median and variance values. The predicted scores of the DL-based features has a higher variance. We compute the p-value between the predicted scores using the combined features and the DL-based features. The value is less than the significance level 0.05, which means that the improvement gain is statistically significant.
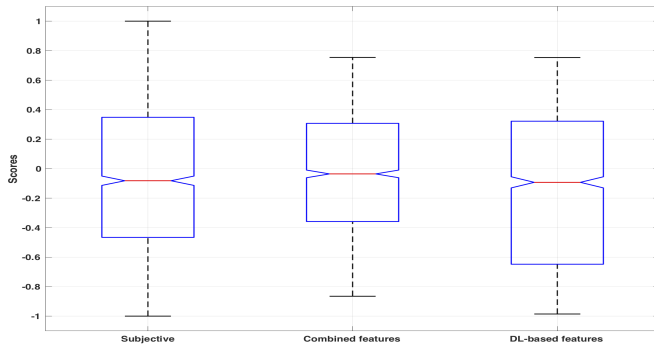


**Fig. 4**. Boxplot for Anova test of the subjective scores and the predicted scores obtained by the CNN and those obtained by the combination of both kinds of features.

We compare bilinear pooling strategy to some common strategies: Concatenation, Summation and Multiplication. Table 2 shows the correlations obtained for each of them. As can be seen, bilinear pooling outperforms all the compared

**Table 2**. Comparison of some pooling strategies using the LIVE-P2 dataset.

| Pooling strategy | PCC | SROCC |
|---|---|---|
| **Concatenation** | 0.970 | 0.969 |
| **Summation** | 0.967 | 0.971 |
| **Multiplication** | 0.970 | 0.971 |
| **Bilinear Pooling** | **0.978** | **0.977** |

**Table 3**. Performance comparison on LIVE-P2 and CSIQ datasets. Mean PCC and SROCC of each dataset across 10 random splitting. The first three best methods are marked in bold and DL-based methods are in italic. The overall best result is with a grey background.

| | | LIVE-P2 | | CSIQ | |
|---|---|---|---|---|---|
| | | PCC | SROCC | PCC | SROCC |
| **FR-IQA** | **PSNR** | 0.856 | 0.866 | 0.800 | 0.806 |
| | **SSIM [14]** | 0.906 | 0.913 | 0.861 | 0.876 |
| | **FSIM [25]** | 0.960 | 0.964 | 0.919 | **0.931** |
| | *DeepIQA [26]* | **0.981** | **0.982** | **0.891** | 0.871 |
| | *DIQa-FR [27]* | **0.980** | **0.970** | — | — |
| **NR-IQA** | **DIIVINE [28]** | 0.917 | 0.916 | 0.900 | 0.880 |
| | **BLIINDS-II [29]** | 0.930 | 0.931 | **0.930** | 0.910 |
| | **BRISQUE [15]** | 0.942 | 0.940 | 0.797 | 0.756 |
| | **CORNIA [30]** | 0.935 | 0.942 | 0.781 | 0.714 |
| | *IQA-CNN [5]* | 0.953 | 0.956 | **0.954** | **0.948** |
| | *IQA-CNN++ [6]* | 0.950 | 0.950 | — | — |
| | *SOM [31]* | 0.962 | 0.964 | — | — |
| | *CNN-Prewitt [32]* | 0.966 | 0.958 | — | — |
| | *Image-wise CNN [33]* | 0.963 | 0.964 | 0.791 | 0.812 |
| | *BIECON [34]* | 0.960 | 0.958 | 0.823 | 0.815 |
| | *DIQA [35]* | 0.972 | **0.970** | 0.880 | 0.844 |
| | *DIQa-NR [27]* | 0.972 | 0.960 | — | — |
| | *Our method* | **0.978** | **0.977** | **0.970** | **0.949** |

methods. Concatenation, Summation and Multiplication achieve close results. These results are certainly due to the fact that bilinear pooling allows a multiplicative interaction between all features and thus combines the advantages of the Concatenation and Multiplication strategies. Indeed, Concatenation permits an interaction between all features, while Multiplication permits only a multiplicative interaction element by element.

### 3.3. Comparison with the state-of-the-art

Our method is compared to the state-of-the-art including handcrafted-based FR and NR metrics (PSNR, SSIM [14], FSIM [25], DIIVINE [28], BLIINDS-2 [29], BRISQUE [15] and CORNIA [30]) as well as CNN-based FR and NR metrics (DeepIQA [26], IQA-CNN [5], IQA-CNN+/IQA-CNN++ [6], SOM [31], CNN-Prewitt [32], BIECON [34], DIQA [35], DIQa-FR & DIQa-NR [27] and Image-wise CNN [33]). Two well-known databases are used (LIVE-P2 and CSIQ). The protocol described in Section 3.2 is applied on both databases. Table 3 shows the results obtained. The first three best methods are marked in bold and CNN-based methods are in italic. The overall best result is represented

with a grey background. On LIVE-P2 database, our method achieves the third best result outperforming all the compared NR methods and being competitive to CNN-based FR methods. DeepIQA and DIQa-FR obtain the best results on this database with a PCC equal to 0.981 and 0.980, respectively. However, the latter necessitate the pristine image (FR-IQA). In addition, CNN-based methods outperform all the handcrafted metrics for both approaches (FR-IQA and NR-IQA), which confirms the contribution of such methods. On CSIQ database, our method achieves the best result with high correlation (PCC=0.970). The second best result is obtained by a CNN-based method (IQA-CNN). The latter use only patches of the whole image to predict the quality and thus not integrate information from the whole image.

### 3.4. Cross-database Evaluation

Finally, we evaluate the generalization ability of our method to predict the quality whatever the dataset. For that, we apply a cross-database evaluation by training our method on LIVE-P2 dataset and testing it on CSIQ, TID08 and TID13 datasets. Table 4 shows the correlations for the three datasets. As can be seen, our method predicts well the quality for all datasets. The best result is obtained for CSIQ (0.896), but still close to those obtained for TID13 (0.894).

**Table 4**. Cross-dataset evaluation. The model was trained on LIVE-P2 dataset and tested on CSIQ, TID08 and TID13 datasets.

|       | PCC   | SROCC |
|-------|-------|-------|
| **CSIQ**  | 0.896 | 0.878 |
| **TID08** | 0.881 | 0.863 |
| **TID13** | 0.894 | 0.873 |

## 4. CONCLUSION

In this paper, we studied the impact of combining handcrafted and DL-based features. For that, bilinear pooling was employed after decreasing the size of the DL-based features through a Count Sketch (CS) projection. We compared the performance of both kinds of features individually and after their combination. We also evaluated the impact of the CS projection on the performance as well as the statistical differences between the predicted scores of the DL-based features and those obtained after combination. The best configuration was retained and compared to the state-of-the-art. Experimental results showed the contribution of handcrafted features.

As future work, we plan to study other handcrafted features (i.e. quality indexes) as well as other pooling strategies using a feature selection method. We will also try to improve the generalization ability of our method, since the performances obtained are not high as expected.

## 5. REFERENCES

[1] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings. International Conference on Image Processing*, Sep. 2002, vol. 3, pp. III–III.

[2] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings. International Conference on Image Processing*, Sep. 2002, vol. 1, pp. I–I.

[3] A. Chetouani, A. Beghdadi, S. Chen, and G. Mostafaoui, "A free reference image quality measure using neural networks," in *International Workshop on Video Processing and Quality Metrics*, 2010.

[4] A. K. Moorthy and A. C. Bovik, "A modular framework for constructing blind universal quality indices," in *IEEE Signal Processing Letters*. IEEE, 2009.

[5] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

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

[7] S. Jia and Y. Zhang, "Saliency-based deep convolutional neural network for no-reference image quality assessment," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 14859–14872, 2018.

[8] A. Chetouani, "A blind image quality metric using a selection of relevant patches based on convolutional neural network," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1452–1456.

[9] Aladine Chetouani and Leida Li, "On the use of a scanpath predictor and convolutional neural network for blind image quality assessment," *Signal Processing: Image Communication*, vol. 89, pp. 115963, 2020.

[10] I. Abouelaziz, A. Chetouani, M. El Hassouni, L.J. Latecki, and H. Cherifi, "Convolutional neural network for blind mesh visual quality assessment using 3d visual saliency," in *2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018*, 2018, pp. 3533–3537.

[11] Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, Longin Jan Latecki, and Hocine Cherifi, "No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling," *Pattern Recognition*, vol. 100, pp. 107174, 2020.

[12] Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, LJ Latecki, and Hocine Cherifi, "3D visual saliency and convolutional neural network for blind mesh quality assessment," *Neural Computing and Applications*, 2019.

[13] Aladine Chetouani, "Image quality assessment without reference by mixing deep learning-based features," in *IEEE International Conference on Multimedia and Expo, ICME 2020, London, UK, July 6-10, 2020*. 2020, pp. 1–6, IEEE.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," in *IEEE Transactions on Image Processing*. IEEE, 2012, vol. 21(12), p. 4695–4708.

[16] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C. Bovik, and Xiangchu Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, pp. 4850–4862, 2014.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR, abs/1409.1556*, 2014.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[19] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji, "Bilinear cnn models for fine-grained visual recognition," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, ICCV '15, pp. 1449–1457.

[20] Ninh Pham and Rasmus Pagh, "Fast and scalable polynomial kernels via explicit feature maps," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2013, KDD '13, pp. 239–247, ACM.

[21] H.R. Sheikh, L. Cormack Z.Wang, and A.C. Bovik, "Live image quality assessment database release 2," in *http://live.ece.utexas.edu/research/quality*.

[22] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[23] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo, "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57 – 77, 2015.

[24] Eric C. Larson and Damon M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electronic Imaging*, vol. 19, pp. 011006, 2010.

[25] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.

[26] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1969–1977.

[27] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, Jan 2018.

[28] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," 2011.

[29] M.A. Saad, A. C. Bovik, and C. Charrier, "Dct statistics model-based blind image quality assessment," in *Proceedings of the IEEE International Conference on image Processing (ICIP)*. IEEE, 2011, pp. 3093–3096.

[30] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised Feature Learning Framework for No-reference Image Quality Assessment," in *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR 2012)*, 2012, pp. 1098–1105.

[31] P. Zhang, W. Zhou, L. Wu, and Houqiang Li, "Som: Semantic obviousness metric for image quality assessment," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 2394–2402.

[32] Jie Li, Lian Zou, Jia Yan, Dexiang Deng, Tao Qu, and Guihui Xie, "No-reference image quality assessment using prewitt magnitude based on convolutional neural networks," *Signal, Image and Video Processing*, vol. 10, pp. 609–616, 2016.

[33] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 130–141, Nov 2017.

[34] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of selected topics in signal processing*, vol. 11, no. 1, pp. 206–220, 2017.

[35] J. Kim, A. Nguyen, and S. Lee, "Deep cnn-based blind image quality predictor," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 11–24, Jan 2019.