



GADA: Generative Adversarial Data Augmentation for Image Quality Assessment

Pietro Bongini^(✉), Riccardo Del Chiaro, Andrew D. Bagdanov,
and Alberto Del Bimbo

Media Integration and Communication Center, University of Florence,
50134 Florence, FI, Italy

{p.bongini,riccardo.delchiaro,andrew.bagdanov,alberto.delbimbo}@unifi.it

Abstract. We propose a No-reference Image Quality Assessment (NR-IQA) approach based on the use of generative adversarial networks. To address the problem of lack of adequate amounts of labeled training data for NR-IQA, we train an Auxiliary Classifier Generative Adversarial Network (AC-GAN) to generate distorted images with various distortion types and levels of image quality at training time. The trained generative model allow us to augment the size of the training dataset by introducing distorted images for which no ground truth is available. We call our approach Generative Adversarial Data Augmentation (GADA) and experimental results on the LIVE and TID2013 datasets show that our approach – using a modestly sized and very shallow network – performs comparably to state-of-the-art methods for NR-IQA which use significantly more complex models. Moreover, our network can process images in real time at 120 image per second unlike other state-of-the-art techniques.

Keywords: Image Quality Assessment ·
Generative Adversarial Networks · Data augmentation

1 Introduction

In the last few decades images are increasingly a part of everyday life and are used for many purposes. However, images are often not of the best possible quality. This can be caused by many factors, such as the device used for acquisition, the lossy compression algorithm used to store the information (e.g. JPEG), and the entire image acquisition, storage, and transmission process.

Image Quality Assessment (IQA) [21] refers to a range of techniques developed to automatically estimate the perceptual quality of images. IQA estimates should be highly correlated with quality assessments made by multiple human evaluators (commonly referred to as the Mean Opinion Score (MOS) [15, 19]). IQA has been widely applied by the computer vision community for applications like image restoration [8], image super-resolution [20], and image retrieval [24].

IQA techniques can be divided into three different categories based on the available information on the image to be evaluated: full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA). Although FR-IQA and RR-IQA methods have obtained impressive results, the fact that they must have knowledge of the undistorted version of the image (called the *reference image*) for quality evaluation, makes these approaches hard to use in real scenarios. On the contrary, NR-IQA only requires the knowledge of the image whose quality is to be estimated, and for this reason is more realistic (and also more challenging).

In the last few years Convolutional Neural Networks (CNNs) have obtained great results on many computer vision tasks, and their success is partially due to the possibility of creating very deep architectures with millions of parameters, thanks to the computational capabilities of modern GPUs. Massive amounts of data are needed for training such models, and this is a big problem for IQA since the annotation process is expensive and time consuming. In fact, each image must be annotated by multiple human experts, and consequently most available IQA datasets are too small to effectively train CNNs from scratch.

In this paper, we propose an approach to address this lack of large, labeled datasets for IQA. Since obtaining annotated data to train the network is difficult, we propose a technique to generate new images with a specific image quality and distortion type. We learn how to generate distorted images using Auxiliary Classifier Generative Adversarial Networks (AC-GANs), and then use these generated images in order to improve the accuracy of a simple CNN regressor trained for IQA. In Fig. 1 we show patches of images generated with our approach alongside their corresponding patches with real distortions.

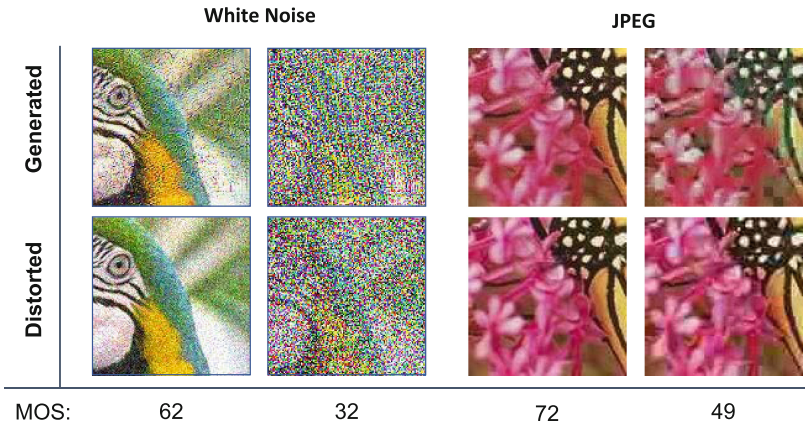


Fig. 1. Patches extracted from images generated by the proposed method compared with the same patches from true distorted images with the same image quality and distortion type.

2 Related Works

In this section we briefly review the literature related to No-Reference Image Quality Assessment (NR-IQA) and Generative Adversarial Networks (GANs).

No-Reference Image Quality Assessment. Most traditional NR-IQA can be classified into Natural Scene Statistics (NSS) methods and learning-based methods. In NSS methods, the assumption is that images of different quality vary in the statistics of responses to specific filters. Wavelets, DCT and Curvelets are commonly used to extract the features in different sub-bands. These feature distributions are parametrized, for example with the Generalized Gaussian Distribution. The aim of these methods is to estimate the distributional parameters, from which a quality assessment can be inferred. The authors of [12] propose to extract NSS features in the spatial domain to obtain significant speed-ups. In learning-based methods, local features are extracted and mapped to the MOS using, for example, Support Machine Regression or Neural Networks [3]. Codebook Methods combines different features instead of using local features directly. Datasets without MOS can be exploited to construct the codebook [25,26] by means of unsupervised learning, which is particularly important due to of the small size of existing datasets. Saliency maps can be used to model human vision system and improve precision in these methods.

Deep Learning for NR-IQA. In recent years several works have used deep learning for NR-IQA. These techniques requires large amounts of data for training and IQA datasets are especially lacking in this regard. Therefore, to address this problem different approaches have been proposed. Kang et al. [6] use small patches of the original images to train a shallow network and thus enlarging the initial dataset. A similar approach was presented in [7] where the authors use a multi-task CNN to learn the type of distortion and the image quality at the same time. Bianco et al. [1] used a pre-trained DCNN fine tuned with an IQA dataset to extract features, and then train a Support Vector Regression model that maps extracted features to quality scores. Liu et al. in [11] use a learning from rankings approach. They train a Siamese Network to rank images in term of image quality and subsequently the information represented in the Siamese network is transferred, trough fine-tuning, to a CNN that predicts the quality score. Another interesting work is from Lin et al. [10] who use a GAN to generate a hallucinated reference image corresponding to a distorted version and then give both the hallucinated reference and the distorted image as input to a regressor that predicts the image quality.

In our work we present a novel approach to address the scarcity of training data: we train an Auxiliary Classifier Generative Adversarial Network (ACGAN) [14] to produce distorted images given a reference image together with a specific quality score and a category of distortion. In this way we can produce new labeled examples that we can use to train a regressor.

Auxiliary Classifier GANs. In the last few years GANs have been widely used in different areas of computer vision. The Auxiliary Classifier GAN (ACGAN) [14] is a variant of the Generative Adversarial Network (GAN) [4] which

uses label conditioning. This kind of network produces convincing results. Our aim is to use this architecture to generate distorted images conditioned to a distortion category and image quality value. Since the main objective of the work is NR-IQA and the performance of the quality regressor is highly related to the generated image, it is crucial that the generator produce convincing distortions.

3 Generative Adversarial Data Augmentation for NR-IQA

In this section we describe our approach to perform data augmentation for NR-IQA datasets. We first show the general steps that characterize our technique, and then describe the use of AC-GAN in this context.

3.1 Overview of Proposed Approach

The main idea of this work is to generate new distorted images with a specific image quality level and distortion type to partially solve the problem of the poverty of annotated data for IQA. We use an AC-GAN to generate new distorted images. Once the generator has learned to produce distorted images convincingly we use it to generate new examples to augment the training set as we train a deep convolutional regressor to estimate IQA. The pipeline of our technique is as follows:

1. **Training the AC-GAN.** Using patches of the training images we train an AC-GAN. The generator learns to generate distorted images with a given distortion class and quality level starting from reference images. The regressor, which aims is to predict the image quality, is trained with both generated and real distorted images using the adversarial GAN loss.
2. **Generative data augmentation.** Once the training of the AC-GAN converges, the generator is able to produce convincing distortions and we can stop its training. We continue training the discriminator branch, augmenting the training data via the trained generator. The regressor is trained with both real distorted images from the training set and images artificially distorted using the generator.
3. **Fine-Tuning of the regressor.** Once convergence is reached in step 2 we perform a final phase of fine-tuning: the regressor is trained with only real distorted images from the IQA training set.

3.2 Auxiliary Classifier GANs for NR-IQA

An Auxiliary Classifier Generative Adversarial Network is a GAN variant in which it is possible to condition the output on some input information. In the AC-GAN every generated sample has a corresponding class label, $c \sim p_c$, in addition to the noise z . This information is given in input to the generator which produces fake images $X_{\text{fake}} = G(c, z)$. The discriminator not only distinguishes

between real and generated examples but predicts also the class label of the examples. The sub-network that classifies the input is called the *classifier*. The objective function is characterized by two components: a log-likelihood on the correct discrimination L_S and a log-likelihood on the correct class L_C :

$$L_S = E[\log P(S = \text{real} \mid X_{\text{real}})] + E[\log P(S = \text{fake} \mid X_{\text{fake}})] \quad (1)$$

$$L_C = E[\log P(C = c \mid X_{\text{real}})] + E[\log P(C = c \mid X_{\text{fake}})] \quad (2)$$

The discriminator is trained to maximize $L_S + L_C$ and the generator is trained to minimize $L_C - L_S$.

Our approach is slightly different from a standard AC-GAN: the latter expects only noise and class label as input, but in our case we want to generate an output image that is a *distorted* version of a reference one, so we also need to feed the reference image and force a reconstruction with an $L1$ loss. Moreover, we want to distort the reference image so that the output matches a target *image quality*, so we feed also this value as input. Because we would like to reconstruct a distorted version of the reference image given in input, we can write the additional $L1$ loss as it follows:

$$\mathcal{L}_{L1} = E[||y - G(z, x, c, q)||_1]$$

where y is the distorted ground truth image, z is a random Gaussian noise vector, x is the reference image, c is the distortion class and v is the image quality.

The goal of this work is to predict the quality score of images, so we introduce a regressor network whose aim is to predict the quality score of input images. The loss used to train this component is a mean squared error (MSE) between the predicted quality score and the ground truth:

$$L_E = E[(q - \hat{q})^2] \quad (3)$$

where q and \hat{q} are the ground truth and the prediction of the image quality score, respectively.

The expectations for all losses defined here are taken over minibatches of either generated or labeled training samples.

3.3 The GADA Architecture

In Fig. 2 we give a schematic representation of the proposed model. The components of the GADA network are as follows.

Generator. The Generator follows the general auto-encoder architecture. It takes as input a high quality reference image, a distortion class, and a target image quality. The input information is encoded through three convolutional layers (one with 64 feature maps and two with 128). Before up-sampling we concatenate a noise vector z to the latent representation, together with an embedding of the distortion category and image quality. We use skip connections [5, 16] in the generator, which allows the network to generate qualitatively better results.

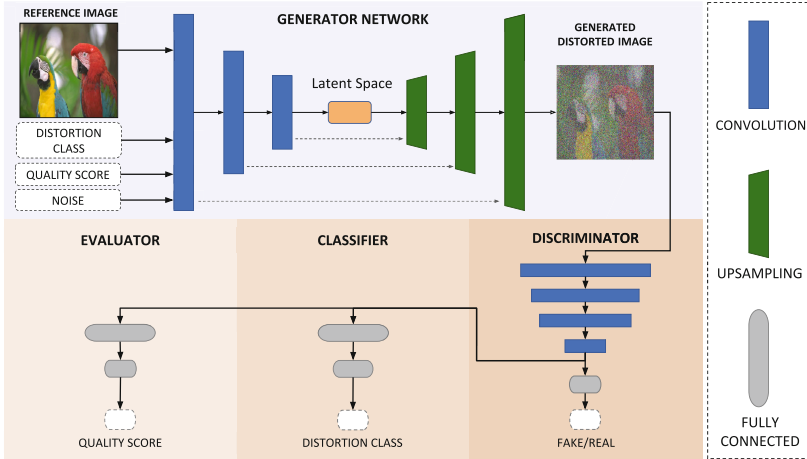


Fig. 2. A schematic representation of the proposed network.

Discriminator. The Discriminator takes as input a distorted image and through three convolutional layers (one with 64 feature maps, and two with 128 to mimic the encoder) followed by a 1×1 convolution extracts 1024 feature maps (that are also fed to the classifier and the regressor). A single fully-connected layer reduces these feature maps to a single value and a sigmoid activation outputs the prediction of the provenance of the input image (i.e. real or fake). This output is used to compute the loss defined in Eq. 1.

Classifier. The Classifier takes as input the feature maps described for the Discriminator. This network consists of two fully-connected layers. The first layer has 128 units and the second has a number of units equal to the number of distortion categories and is followed by a softmax activation function. The output of this module is used in the classifier loss for the AC-GAN as defined in Eq. 2.

Evaluator. The Evaluator takes as input the feature maps described for the Discriminator and should accurately estimate the image quality of the input image. This module consists of two fully-connected layers, the first with 128 and the second with a single unit. The MSE loss defined in Eq. 3 is computed using the output of this module.

4 Experimental Results

In this section we describe experiments conducted to evaluate the performance of our approach. We first introduce the datasets used for training and testing our network, then we describe the protocols adopted for the experiments.

4.1 Datasets

For our experiments we used the standard LIVE [18] and TID2013 [15] datasets for IQA. LIVE contains 982 distorted versions of 29 reference images. Original images are distorted with five different types of distortion: JPEG compression (JPEG), JP2000 compression (JP2K), white noise (WN), gaussian blur (GB) and fast fading (FF). The ground truth quality score for each image is the Difference Mean Opinion Score (DMOS) whose value is in the range $[0, 100]$. TID2013 consist of 3000 distorted images versions of 25 reference images. The original images are distorted with 24 different types of distortions. The Mean Opinion Score of distorted images varies from 0 to 9.

4.2 Experimental Protocols

We analyze the performance of our model using the standard IQA metrics. For each dataset we randomly split the reference images (and their corresponding distorted versions) in 80% used for training and 20% used for testing, as described in [6, 28]. This process is repeated ten times. For each split we train from scratch and compute the final scores on the test set.

Training Strategy. At each training epoch, we randomly crop each image in the training-set using patches of 128×128 pixels and feed it to the model. For all the three phases we train using these crops with a batch size of 64. During the first one we use Adam optimizer with a learning rate of $1e^{-4}$ for the discriminator and $5e^{-4}$ for the generator, classifier and evaluator. During the second and third phases we divide the learning rate by 10.

Testing Protocol. At test time We randomly crop 30 patches from each test image as suggested in [1]. We then pass all 30 crops through the discriminator network (with only the evaluator branch) to estimate IQA. The average of the predictions for the 30 crops gives the final estimated quality score.

Evaluation Metrics. We use two evaluation metrics commonly used in IQA context: the Linear Correlation Coefficient (LCC) and Spearman Correlation Coefficient (SROCC). LCC is a measure of the linear correlation between the ground truth and the predicted quality scores. Given N distorted images, the ground truth of i -th image is denoted by y_i , and the predicted score from the network is \hat{y}_i . The LCC is computed as:

$$LCC = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_i^N (y_i - \bar{y})^2} \sqrt{\sum_i^N (\hat{y}_i - \bar{\hat{y}})^2}} \quad (4)$$

where \bar{y} and $\bar{\hat{y}}$ are the means of the ground truth and predicted quality scores, respectively.

Given N distorted images, the SROCC is:

$$SROCC = 1 - \frac{6 \sum_{i=1}^N (v_i - p_i)^2}{N(N^2 - 1)}, \quad (5)$$

where v_i is the *rank* of the ground-truth IQA score y_i in the ground-truth scores, and p_i is the *rank* of \hat{y}_i in the output scores for all N images. The SROCC measures the monotonic relationship between ground-truth and estimated IQA.

4.3 Generative Data Augmentation with AC-GAN

As described in Sect. 3.1 our approach consists of three phases: a first one where we train the generator, a second phase where we perform data augmentation, and the final fine-tuning phase of the evaluator over the original training-set. As a first experiment, we calculated the performance obtained after each of the three different phases and compared with the performance of a direct method which consists of training *only* the evaluator and classifier branches of the discriminator directly on labeled training data (e.g. no adversarial data augmentation). We trained and tested the proposed method and the direct baseline on the LIVE dataset as described in Sect. 4.2, but for this preliminary experiment we used crops of 64×64 pixels and a shallower regression network.

In Table 1 we give the LCC and SROCC values computed for the baseline and after each of the three phase of our approach. We note first that each phase of our training procedure results in improved LCC and SROCC, which indicates that generative data augmentation and fine-tuning both add to performance. At the end of phase 3 the LCC and SROCC results surpass the direct approach by $\sim 2\%$, confirming the effectiveness of GADA with respect to direct training.

Table 1. Comparison of baseline and each phase of the GADA approach in LCC and SROCC. In the first block results for the direct baseline method (directly training the evaluator with only labeled IQA data) are shown. In the second block results for our method are shown after each of the three phases: training of the AC-GAN (Phase 1), generator data augmentation (Phase 2), and evaluator fine-tuning (Phase 3).

		JP2K	JPEG	WN	GBLUR	FF	ALL
Baseline	LCC	0.950	0.964	0.973	0.938	0.933	0.943
	SROCC	0.938	0.931	0.977	0.939	0.898	0.935
Phase 1	LCC	0.944	0.952	0.967	0.920	0.912	0.933
	SROCC	0.933	0.930	0.980	0.926	0.889	0.930
Phase 2	LCC	0.958	0.958	0.974	0.939	0.924	0.942
	SROCC	0.941	0.933	0.988	0.945	0.891	0.939
Phase 3	LCC	0.959	0.973	0.993	0.953	0.935	0.962
	SROCC	0.955	0.941	0.990	0.953	0.912	0.955

4.4 Comparison with the State-of-the-Art

Here we compare GADA with state-of-the-art results from the literature.

Results on LIVE. We trained on LIVE dataset following the protocol described in Sect. 4.2. The results are shown in Table 2. Each column of the table represents the partial scores for a specific distortion category of LIVE dataset. Our method seems to be very effective on this dataset despite the fact that many other approaches process larger patches (e.g. 224×224 , the input size of the VGG16 network) and capture more context information. We observe from the table that our model performs very well on Gaussian noise (GN) and JPEG2000 (JP2K). We obtain worse results for Fast Fading (FF), which is probably due to the fact that FF is a local distortion and we process patches of small dimension, so for each crop the probability of picking a distorted region is not 1.

Table 2. Comparison between GADA and the state-of-the-art on LIVE.

	LCC						SROCC					
	JP2K	JPEG	GN	GB	FF	ALL	JP2K	JPEG	GN	GB	FF	ALL
DIVINE [13]	.922	.921	.988	.923	.888	.917	.913	.91	.984	.921	.863	.916
BLIINDS-II [17]	.935	.968	.980	.938	.896	.930	.929	.942	.969	.923	.889	.931
BRISQUE [12]	.923	.973	.985	.951	.903	.942	.914	.965	.979	.951	.887	.940
CORNIA [26]	.951	.965	.987	.968	.917	.935	.943	.955	.976	.969	.906	.942
CNN [6]	.953	.981	.984	.953	.933	.953	.952	.977	.978	.962	.908	.956
SOM [28]	.952	.961	.991	.974	.954	.962	.947	.952	.984	.976	.937	.964
BIECON [9]	.965	.987	.970	.945	.931	.962	.952	.974	.980	.956	.923	.961
PQR [27]	–	–	–	–	–	.971	–	–	–	–	–	.965
DNN [2]	–	–	–	–	–	.972	–	–	–	–	–	.960
RankIQA+FT [11]	.975	.986	.994	.988	.960	.982	.970	.978	.991	.988	.954	.981
Hall.-IQA [10]	.977	.984	.993	.990	.960	.982	.983	.961	.984	.983	.989	.982
NSSADNN [23]	–	–	–	–	–	.984	–	–	–	–	–	.986
GADA (ours)	.977	.978	.994	.968	.943	.973	.963	.948	.991	.958	.917	.964

Table 3. Comparison between GADA and the state-of-the-art on TID2013 (SROCC).

Method	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13
BLIINDS-II [17]	0.714	0.728	0.825	0.358	0.852	0.664	0.780	0.852	0.754	0.808	0.862	0.251	0.755
BRISQUE [12]	0.630	0.424	0.727	0.321	0.775	0.669	0.592	0.845	0.553	0.742	0.799	0.301	0.672
CORNIA-10K [26]	0.341	-0.196	0.689	0.184	0.607	-0.014	0.673	0.896	0.787	0.875	0.911	0.310	0.625
HOSA [22]	0.853	0.625	0.782	0.368	0.905	0.775	0.810	0.892	0.870	0.893	0.932	0.747	0.701
RankIQA+FT [11]	0.667	0.620	0.821	0.365	0.760	0.736	0.783	0.809	0.767	0.866	0.878	0.704	0.810
NSSADNN [23]	–	–	–	–	–	–	–	–	–	–	–	–	–
HALLUCINATED IQA [10]	0.923	0.880	0.945	0.673	0.955	0.810	0.855	0.832	0.957	0.914	0.624	0.460	0.782
GADA (ours)	0.932	0.897	0.943	0.825	0.949	0.920	0.919	0.790	0.881	0.775	0.886	0.435	0.702
Method	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	ALL	
BLIINDS-II [17]	0.081	0.371	0.159	-0.082	0.109	0.699	0.222	0.451	0.815	0.568	0.856	0.550	
BRISQUE [12]	0.175	0.184	0.155	0.125	0.032	0.560	0.282	0.680	0.804	0.715	0.800	0.562	
CORNIA-10K [26]	0.161	0.096	0.008	0.423	-0.055	0.259	0.606	0.555	0.592	0.759	0.903	0.651	
HOSA [22]	0.199	0.327	0.233	0.294	0.119	0.782	0.532	0.835	0.855	0.801	0.905	0.728	
RankIQA+FT [11]	0.512	0.622	0.268	0.613	0.662	0.619	0.644	0.800	0.779	0.629	0.859	0.780	
NSSADNN [23]	–	–	–	–	–	–	–	–	–	–	–	0.844	
HALLUCINATED IQA [10]	0.664	0.122	0.182	0.376	0.156	0.850	0.614	0.852	0.911	0.381	0.616	0.879	
GADA (ours)	0.206	0.200	0.196	0.739	0.688	0.950	0.679	0.937	0.895	0.843	0.889	0.790	

TID2013. We follow the same test procedure for TID2013 and report our SROCC results in Table 3. We see that for 11 of the 24 types of distortion we

obtain the best results. For local and challenging distortions like #14, #15 and #16 the performance of our model is low, and again we hypothesize that the small size and uniform sampling of patches could be a limitation especially for extremely local distortions.

5 Conclusions

In this paper we proposed a new approach called GADA to resolve the problem of lack of training data for No-reference Image Quality Assessment. Our approach uses a modified Auxiliary Classifier GAN. This technique allows us to use the generator to generate new training examples and to train a regressor which estimates the image quality score. The results obtained on LIVE and TID2013 datasets show that our performance is comparable with the best methods of the state-of-the-art. Moreover, the very shallow network used for the regressor can process images with an high frame rate (about 120 image per second). This is in stark contrast to state-of-the-art approaches which typically use very deep models like VGG16 pre-trained on ImageNet.

We feel that the GADA approach offers a promising alternative to laboriously annotating images for IQA. Significant improvements can likely be made, especially for highly local distortions, through saliency-based sampling of image patches during training.

References

1. Bianco, S., Celona, L., Napolitano, P., Schettini, R.: On the use of deep learning for blind image quality assessment. arXiv preprint [arXiv:1602.05531](https://arxiv.org/abs/1602.05531) (2016)
2. Bosse, S., Maniry, D., Wiegand, T., Samek, W.: A deep neural network for image quality assessment. In: Proceedings of ICIP, pp. 3773–3777. IEEE (2016)
3. Chetouani, A., Beghdadi, A., Chen, S., Mostafaoui, G.: A novel free reference image quality metric using neural network approach. In: Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics, pp. 1–4 (2010)
4. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017
6. Kang, L., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733–1740 (2014)
7. Kang, L., Ye, P., Li, Y., Doermann, D.: Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 2791–2795. IEEE (2015)
8. Katsaggelos, A.K.: Digital Image Restoration. Springer, Heidelberg (2012)
9. Kim, J., Lee, S.: Fully deep blind image quality predictor. IEEE J. Sel. Top. Signal Process. **11**(1), 206–220 (2017)

10. Lin, K.Y., Wang, G.: Hallucinated-IQA: no-reference image quality assessment via adversarial learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 732–741 (2018)
11. Liu, X., van de Weijer, J., Bagdanov, A.D.: RankIQA: learning from rankings for no-reference image quality assessment. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1040–1049 (2017)
12. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012)
13. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Trans. Image Process.* **20**(12), 3350–3364 (2011)
14. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier GANs. In: Proceedings of the 34th International Conference on Machine Learning, vol. 70. pp. 2642–2651 (2017). [JMLR.org](http://jmlr.org)
15. Ponomarenko, N., et al.: Color image database TID2013: peculiarities and preliminary results. In: 2013 4th European Workshop on Visual Information Processing (EUVIP), pp. 106–111. IEEE (2013)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* **21**(8), 3339–3352 (2012)
18. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: Live image quality assessment database. <http://live.ece.utexas.edu/research/quality>
19. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **15**(11), 3440–3451 (2006)
20. Van Ouwierkerk, J.: Image super-resolution survey. *Image Vis. Comput.* **24**(10), 1039–1052 (2006)
21. Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, p. IV–3313. IEEE (2002)
22. Xu, J., Ye, P., Li, Q., Du, H., Liu, Y., Doermann, D.: Blind image quality assessment based on high order statistics aggregation. *IEEE Trans. Image Process.* **25**(9), 4444–4457 (2016)
23. Yan, B., Bare, B., Tan, W.: Naturalness-aware deep no-reference image quality assessment. *IEEE Trans. Multimedia* **PP**, 1 (2019)
24. Yan, J., Lin, S., Kang, S.B., Tang, X.: A learning-to-rank approach for image color enhancement. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2987–2994. IEEE (2014)
25. Ye, P., Doermann, D.: No-reference image quality assessment using visual codebooks. *IEEE Trans. Image Process.* **21**(7), 3129–3138 (2012)
26. Ye, P., Kumar, J., Kang, L., Doermann, D.: Unsupervised feature learning framework for no-reference image quality assessment. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1098–1105. IEEE (2012)
27. Zeng, H., Zhang, L., Bovik, A.C.: A probabilistic quality representation approach to deep blind image quality prediction (2017)
28. Zhang, P., Zhou, W., Wu, L., Li, H.: SOM: semantic obviousness metric for image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2394–2402 (2015)