

Full-Reference Image Quality Assessment

Subjects: **Computer Science, Artificial Intelligence**

Contributor: Pei-Fen Tsai , Huai-Nan Peng , Chia-Hung Liao , Shyan-Ming Yuan

To improve data transmission efficiency, image compression is a commonly used method with the disadvantage of accompanying image distortion. There are many image restoration (IR) algorithms, and one of the most advanced algorithms is the generative adversarial network (GAN)-based method with a high correlation to the human visual system (HVS). Having a metric to quantify the image quality to the HVS is always a tough task. Image quality assessment (IQA) can be subjective or objective.

image quality assessment (IQA)

full-reference IQA

1. Introduction

Due to the vigorous development of the Internet of Things (IoT), the necessary compression and restoration of images in the transmission process have become increasingly hot topics. With the progress of deep learning models including the convolutional neural network (CNN) [1][2] and the generative and adversarial network (GAN) [3][4], the quality of image restoration [5], image compression, and super-resolution [6] has been greatly improved. Therefore, having a metric to quantify the image quality to the human visual system (HVS) is always a tough task. Image quality assessment (IQA) [7][8][9] can be subjective or objective.

Subjective IQA judged through human eyes is more reliable and accurate for HVS and is named by mean opinion scores (MOS). However, subjective IQA is time-consuming and expensive. Recently, more and more studies on objective IQA models have achieved outstanding results, including full-reference IQA (FR-IQA) [9], reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA), which can be used to predict the image quality score.

FR-IQA compares the distance of the reference image and distorted image as a pair to measure the distorted image quality. NR-IQA predicts the distorted image quality without a reference image. That means NR-IQA has no prior knowledge for original information and limited uses for measuring a distorted image compared to a restored image.

However, in these years, with neural network contribution, NR-IQA has achieved a significant performance on metric distorted images, such as MetalIQ [10], proposed in 2020, and GraphIQ [11] and LIQA [12], proposed in 2022.

2. Full-Reference Image Quality Assessment (FR-IQA)

The FR-IQA algorithm as shown in **Figure 1** inputs the original image, that is, the reference image, and the distorted image, that is, the quality image to be judged. After the image resolution is unified through preprocessing, the features of the images are extracted through feature extraction, and then, the distance between the distorted image and the reference image is calculated to obtain the image quality score of the distorted image.

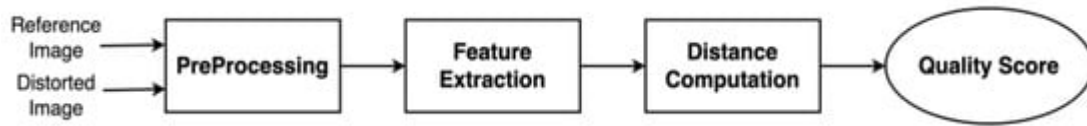


Figure 1. FR-IQA model.

In the distance computation, the mean square error (MSE) and peak signal-to-noise ratio (PSNR) are commonly used methodology. However, metrics of MSE and PSNR cannot obtain accurate image quality score for some images from deep-learning based image restoration algorithms.

To improve the poor performance of MSE and PSNR, structural similarity (SSIM) [13] was proposed to include image luminance, contrast, and structure as an index in FR-IQA to calculate the similarity of the image structure between a reference image and a distorted image. Then, SSIM was extended to a multiscale structural similarity index called MS-SSIM. Structural-based metric models of image luminance, contrast, and structure on multiple scales.

The convolutional neural network (CNN) architecture shows significant achievement in computer vision, including image classification, image segmentation, and object detection, due to its ability to extract different scales of image features. In FR-IQA, the use of CNN architecture to extract multiple levels of features for distance computation was proposed by learned perceptual image patch similarity (LPIPS) [14] and deep image structure and texture similarity (DISTS) [15].

The generative adversarial network (GAN) has been proven to be a good approximator for natural image restoration (IR) [16][17][18][19][20]. In 2021, the perceptual image processing algorithm (PIPAL) dataset [17] was developed with 116 distortions of traditional IR and GAN-based IR as a novel large-scale IQA dataset. The PIPAL dataset had been used in the CVPR NTIRE 2021 and 2022 challenge as a benchmark for IQA algorithms.

In the CVPR NTIRE 2021 challenge, the team that won first implemented the transformer-based IQA named IQT-C [21] to demonstrate the global attention of transformer-to-IQA capability. In the CVPR NTIRE 2022 challenge [22], the team that won first developed an ensemble method with transformer and CNN for feature extraction named attention-based hybrid image quality assessment network [23]. The ensemble deep learning model shows its power in IQA as well.

3. IQA Datasets

Numerous IQA datasets have been proposed over decades. The FR-IQA dataset contains reference images and different types of distortion images as pairs. Each distorted image is evaluated by MOS relative to the reference image by human judgment. The distortion types [24][25] vary to include traditional JPEGs such as image compression, white noise, Gaussian blur, and the traditional algorithm for image restoration.

Some of the well-known FR-IQA datasets are LIVE [26], TID2008 [27], TID2013 [28], and KADID-10k [29], as shown in **Table 1** with the increase in reference images, distortion images, and distortion types yearly.

Table 1. Image Quality Assessment Datasets.

Image Restoration	Dataset	Year	No. Ref	No. Dist.	No. Dist. Type
Traditional Algorithm	LIVE [20]	2004	29	0.8k	5
	TID2013 [21]	2013	25	3k	24
	KADID-10k [23]	2019	81	10.1k	25
Trad. + GAN Algorithm	PIPAL [10]	2020	250	29k	40

PIPAL [30], first introduced in 2020 by Jinjin et al., increased the reference images to 250 and distortion types to 40 and introduced distorted images using the GAN-based restoration algorithm as the current state-of-the-art FR-IQA dataset today. It is also used to evaluate FR-IQA models in distorted and restored images to include the GAN-based algorithm in the CVPR NTIRE 2021 IQA challenge.

4. Feature Extraction Backbone in FR-IQA

Feature extraction backbones are key components in learning-based IQA algorithms as published in recent years. In 2018, Zhang et al. investigated different CNN-based models such as LPIPS [14] to demonstrate that deep features are representative of image semantics for evaluating image quality.

In 2021, Guo et al. proposed the IQA of multiscale features as an image quality multiscale assessment network (IQMA) [24] to demonstrate the power of the feature fusion module with feature pyramid network (FPN) backbone for multiscale feature fusion.

Shi et al. significantly improved WResNet IQA models (W stands for weighted averaging) [31] by modifying the residual block in the feature extraction backbone of the fourth ranking in the NTIRE 2021 IQA challenge.

5. Ensemble Methods in FR-IQA

Ensemble learning has been shown to be effective in performance on deep learning tasks. In FR-IQA, there are several models that use ensemble methodology to obtain high correlation metrics: the IQMA [32] network, which won the 2nd award in the NTIRE 2021 competition, proposed to average predicted quality scores from different levels of feature maps. EGB predicts image quality scores with three regressors and averages the three predicted scores to obtain the final quality score.

Ensemble methodology [33] is commonly used to solve statistical and computational problems while maintaining model generalization. It has been demonstrated successfully in various fields including face recognition [34], emotion recognition [35][36], medical treatment [37][38], etc. In the NTIRE 2021 IQA challenge, the team proposing an IQA metric named the gradient boosting (EGB) ensemble [39], with three regressors for the image quality score, won the second prize for outstanding performance of the IQA ensemble metric.

References

1. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* 1998, 86, 2278–2324.
2. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In *Proceedings of the 2017 International Conference on Engineering and Technology (ICET)*, Antalya, Turkey, 21–23 August 2017; pp. 1–6.
3. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* 2020, 63, 139–144.
4. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* 2018, 35, 53–65.
5. Banham, M.R.; Katsaggelos, A.K. Digital image restoration. *IEEE Signal Process. Mag.* 1997, 14, 24–41.
6. van Ouwerkerk, J. Image super-resolution survey. *Image Vis. Comput.* 2006, 24, 1039–1052.
7. Wang, Z.; Bovik, A.C.; Lu, L. Why is image quality assessment so difficult? In *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, USA, 13–17 May 2002; Volume 4, pp. IV-3313–IV-3316.
8. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* 2020, 63, 1–52.
9. Sheikh, H.; Sabir, M.; Bovik, A. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* 2006, 15, 3440–3451.
10. Zhu, H.; Li, L.; Wu, J.; Dong, W.; Shi, G. MetaIQA: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern*

Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14143–14152.

11. Sun, S.; Yu, T.; Xu, J.; Zhou, W.; Chen, Z. GraphIQA: Learning Distortion Graph Representations for Blind Image Quality Assessment. *IEEE Trans. Multimedia* 2022.
12. Liu, J.; Zhou, W.; Li, X.; Xu, J.; Chen, Z. LIQA: Lifelong Blind Image Quality Assessment. *IEEE Trans. Multimedia* 2022, 1–16.
13. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* 2004, 13, 600–612.
14. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.
15. Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 44, 2567–2581.
16. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017.
17. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, Munich, Germany, 8–14 September 2018.
18. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.
19. Zhang, W.; Liu, Y.; Dong, C.; Qiao, Y. RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, 27–28 October 2019.
20. Cai, H.; He, J.; Qiao, Y.; Dong, C. Toward interactive modulation for photo-realistic image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 19–25 June 2021; pp. 294–303.
21. Cheon, M.; Yoon, S.-J.; Kang, B.; Lee, J. Perceptual image quality assessment with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 19–25 June 2021; pp. 433–442.
22. Gu, J.; Cai, H.; Dong, C.; Ren, J.S.; Timofte, R.; Gong, Y.; Lao, S.; Shi, S.; Wang, J.; Yang, S.; et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 951–967.
23. Lao, S.; Gong, Y.; Shi, S.; Yang, S.; Wu, T.; Wang, J.; Xia, W.; Yang, Y. Attentions Help CNNs See Better: Attention-based Hybrid Image Quality Assessment Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1139–1148.
 24. Larson, E.C.; Chandler, D.M. Most apparent distortion: A dual strategy for full-reference image quality assessment. In Image Quality and System Performance VI; SPIE: Sydney, Australia, 2009; Volume 7242, pp. 270–286.
 25. Larson, E.C.; Chandler, D.M. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* 2010, 19, 011006.
 26. Sheikh, H. LIVE Image Quality Assessment Database Release 2. 2005. Available online: <http://live.ece.utexas.edu/research/quality/subjective.htm> (accessed on 5 June 2021).
 27. Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.* 2009, 10, 30–45.
 28. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* 2015, 30, 57–77.
 29. Lin, H.; Hosu, V.; Saupe, D. KADID-10k: A large-scale artificially distorted IQA database. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–3.
 30. Jinjin, G.; Haoming, C.; HaoYu, C.; Xiaoxing, Y.; Ren, J.S.; Chao, D. PIPAL: A Large-Scale Image Quality Assessment Dataset for Perceptual Image Restoration. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2020; pp. 633–651.
 31. Shi, S.; Bai, Q.; Cao, M.; Xia, W.; Wang, J.; Chen, Y.; Yang, Y. Region-adaptive deformable network for image quality assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 324–333.
 32. Guo, H.; Bin, Y.; Hou, Y.; Zhang, Q.; Luo, H. Iqma network: Image quality multi-scale assessment network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 443–452.
 33. Zhou, Z.-H. Ensemble Methods: Foundations and Algorithms; CRC Press: Boca Raton, FL, USA, 2012.

34. Mu, X.; Lu, J.; Watta, P.; Hassoun, M.H. Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 2168–2171.
35. Rieger, S.A.; Muraleedharan, R.; Ramachandran, R.P. Speech based emotion recognition using spectral feature extraction and an ensemble of kNN classifiers. In Proceedings of the 9th International Symposium on Chinese Spoken Language Processing, Singapore, 12–14 September 2014; pp. 589–593.
36. Krajewski, J.; Batliner, A.; Kessel, S. Comparing Multiple Classifiers for Speech-Based Detection of Self-Confidence - A Pilot Study. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3716–3719.
37. Savio, A.; García-Sebastián, M.; Chyzyk, D.; Hernandez, C.; Graña, M.; Sistiaga, A.; de Munain, A.L.; Villanúa, J. Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Comput. Biol. Med.* 2011, 41, 600–610.
38. Ayerdi, B.; Savio, A.; Graña, M. Meta-ensembles of classifiers for Alzheimer's disease detection using independent ROI features. In International Work-Conference on the Interplay Between Natural and Artificial Computation, Mallorca, Spain, 10–14 June 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 122–130.
39. Hammou, D.; Fezza, S.A.; Hamidouche, W. Egb: Image quality assessment based on ensemble of gradient boosting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 541–549.

Retrieved from <https://encyclopedia.pub/entry/history/show/120957>