A Survey of Learning Based No Reference Image Quality Assessment

Botao An [1, a], Hongwei Zhou [2, b], Peiran Peng [1, c], Lei Zhang [2, d], Shubo Ren [2, e], Jianan Li [1, f], Tingfa Xu[*1, g 1]

[1] School of Optics and Photonics, Beijing Institute of Technology, China;
[2] Institute of Telecommunication and Navigation Satellites, China Academy of Space Technology, China.

ABSTRACT.

Digital images are captured by various fixed and mobile cameras, compressed with traditional and novel techniques, transmitted through different communication channels, and stored in various storage devices. Distortions can occur at each stage of the image acquisition, processing, transmission and storage pipeline, resulting in loss of perceptual information and degradation of quality. Therefore, image quality assessment is becoming increasingly important in monitoring image quality and ensuring the reliability of image processing systems. And as the most widely applicable and usable of the image quality assessment fields, a large number of learning-based no-reference quality assessment studies have been conducted in recent years. In this survey, we provide an up-to-date and comprehensive review of these studies. Specifically, this paper presents recent advances in the field of deep learning-based no-reference quality assessment and provides an overview of benchmark databases for deep learning-based no-reference quality assessment tasks as well as assessment metrics and the backbone networks commonly used in quality assessment tasks.

Keywords: Survey; Image quality assessment; No-reference image quality assessment；

1. INTRODUCTION

With the proliferation of smartphones and digital products, billions of images are uploaded to the Internet daily, necessitating the need for quality assessment. Objective quality assessment plays a crucial role in handling this massive volume of data. For a given image to be evaluated, the task of Image Quality Assessment is to give a quantitative score to measure its image quality within a certain error range. This margin of error is relevant to the human visual system. In other words, the NR-IQA model needs to be able to assess image quality to a degree comparable to subjective scoring using the human visual system. Image Quality Assessment (IQA) can be categorized into three types: full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA). NR-IQA is particularly suitable for practical applications as it does not require a corresponding high-quality reference image for comparison.

Traditional NR-IQA models rely on hand-crafted or learned low-level features to evaluate image quality. However, recent advancements in convolutional neural networks (CNNs) have led to the development of CNN-based NR-IQA methods, outperforming traditional feature-based approaches and achieving remarkable performance in various computer vision tasks. Nevertheless, there are still many challenges to overcome, such as datasets that are difficult to support training and models that are not powerful enough to match the human vision system.

This paper provides a systematic overview of learning-based No-Reference Image Quality Assessment (NR-IQA). It discusses the purpose, requirements, and challenges in NR-IQA models. The analysis in Section 2 focuses on common NR-IQA databases, metrics for evaluating IQA models, and datasets specific to NR-IQA. In Section 3, thought-provoking learning-based NR-IQA methods are introduced, highlighting their features and limitations, aiming to inspire innovation in NR-IQA methods.

[1a] bitanbotao@bit.edu.cn, [b] hwzhou@sina.com, [c] 3220215082@bit.edu.cn, [d] leizhang200@163.com, [e] renshubo@163.com, [f] lijianan@bit.edu.cn, [g*] ciomxtf1@bit.edu.cn

The International Conference Optoelectronic Information and Optical Engineering (OIOE2024), edited by Yang Yue, Lu Leng, Proc. of SPIE Vol. 13513, 135132S · © 2025 SPIE · 0277-786X · © The Authors.

2. DATABASES AND EVALUATION METRICS

2.1 Databases

The selection of the training database significantly impacts the predictive ability of NR-IQA models. A larger and more diverse database with various distortion types and levels improves the model's generalization. Commonly used NR-IQA databases can be categorized into synthetic and authentic distortion databases. Authentic databases consist of real-world distorted images without reference images, specifically designed for NR-IQA. Further details of these databases are shown in Table 1, and sample images is shown in Figure 1.

Figure. 1 Different distortion categories in the KADAD-10k database

2.1.1 LIVE Image Quality Assessment database (LIVE)

The LIVE database [6] was presented in 2006.The reference images for the entire database are derived from a collection of 29 high resolution and high quality colour images from the internet and photographic CD-ROMs, The LIVE database was degraded using five computer distortion operations for each of the reference images at 5 levels, resulting in 779 distorted images. Each distortion type contains 5 or 4 distortion levels.

2.1.2 Categorical Subjective Image Quality database(CSIQ)

The CSIQ [7] database was created in 2009 and contains 30 original images and 866 synthetic distorted images. Each distortion type was subjected to degradation operations at four to five different distortion levels to obtain 866 distorted versions of the original images.

2.1.3 Tampere Image database 2008 (TID2008)

TID2008 [8] includes 25 reference images and 1700 distorted images. The reference image was obtained from the Kodak Lossless True Color Image Suite by cropping a total of 17 distortion types.

2.1.4 Tampere Image database 2013 (TID2013)

TID2013 [9] is extended from TID2008 by increasing the number of distortion levels to 5, and the number of distortion types to 24. Therefore, 3000 distorted images are generated from 25 pristine images.

2.1.5 LIVE In The Wild (LIVEC)

LIVEC database [10] includes variations in luminance/color activity and smoothness, providing a broader spectrum of perceptible impairments. By incorporating complex distortion mixing inherent in image acquisition, processing, and transmission, the LIVEC database serves as a more representative and comprehensive training source for NR-IQA models.includes variations in luminance/color activity and smoothness, providing a broader spectrum of perceptible impairments. By incorporating complex distortion mixing inherent in image acquisition, processing, and transmission, the LIVEC database serves as a more representative and comprehensive training source for NR-IQA models.

2.1.6 Konstanz Artificially Distorted Image quality database (KADID-10k)

KADID-10k [11], created in 2019, it includes 81 pristine images degraded by 25 distortions across 5 levels. Each distorted image has obtained 30 reliable degradation category ratings through crowdsourcing. KADID-10k is three times larger than TID2013, containing a total of 25 types of distortions. These distortions can be grouped into blurs, color-related distortions, compression distortions, noise-related distortions, brightness changes, spatial distortions, sharpness, and contrast.

2.1.7 Konstanz Authentic Image Quality database (KonIQ-10k)

KonIQ-10k [12] was created in 2020, selected 10,073 images from the YFCC100M database (a large public multimedia database containing 10 million images), all of which are from the real world, and the sampling process used a depth brightness, colour, contrast and sharpness can be widely and evenly distributed. KonIQ-10k produced from these images is better in terms of data fidelity, size and diversity than previous manually produced databases, and the models trained with this kind of database are more applicable.

Table 1. Comparison of existing benchmark IQA databases

database	Year	Content	No. of distorted images	Distortion type	No. of Distortion types	No. of Rated images	Subjective Subjectivestudy environment
LIVE	2006	29	779	artificial	5	779	lab
CSIQ	2009	30	866	artificial	6	866	lab
TID2008	2008	25	1700	artificial	17	1700	lab
TID2013	2013	25	3000	artificial	24	3000	lab
LIVEC	2016	1169	1169	authentic	N/A	1169	crowdsourcing
KADID-10k	2019	81	20125	artificial	25	10125	crowdsourcing
KonIQ-10k	2020	10073	10073	authentic	N/A	10073	crowdsourcing

2.2 Metrics

A better performing image quality assessment algorithm will have a quality assessment score that is highly consistent with the subjective assessment quality score, and IQA has a number of evaluation metrics. In order to measure the consistency between method test results and subjective evaluations, Video Quality Expert Group (VQEG) has proposed metrics that can verify the closeness between objective and subjective evaluation results: PLCC, SROCC and KROCC, which are currently the most commonly used metrics for evaluating the performance of IQA algorithms.

2.2.1 Pearson linear correlation coefficient (PLCC)

PLCC is used to assess the accuracy of the IQA model predictions. PLCC evaluates the correlation between the subjective score (MOS) and the objective score after non-linear regression.

$$PLCC = \frac{\sum_{i=1}^{N}(p_i-\bar{p})(s_i-\bar{s})}{\sqrt{\sum_{i=1}^{N}(p_i-\bar{p})^2(s_i-\bar{s})^2}} \qquad (1)$$

where s_i and p_i denote the subjective score and the converted objective score after nonlinear mapping of the i-th image, $\bar{s}$ and $\bar{p}$ are the averages of al s_i and p_i .

2.2.2 Spearman rank order correlation coefficient (SROCC)

SROCC is used to measure the monotonicity of IQA algorithm predictions.

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \qquad (2)$$

where d_i represents the difference between the i-th images 's ranks in subjective and objective evaluations, and N is the number of testing images.

2.2.3 Kendall rank order correlation coefficient (KROCC)

KROCC is used as well as SROCC to measure the monotonicity of the IQA model predictions.

$$\text{KROCC} = \frac{2(N_c-N_d)}{N(N-1)} \qquad (3)$$

where N_c and N_d express the numbers of concordant and discordant pairs in the testing data. N is the number of testing images as well.

Of the three evaluation criteria mentioned above, SROCC and KROCC measure prediction monotonicity, PLCC evaluates linearity and consistency. the higher the SROCC, KROCC and PLCC scores, the better the correlation with subjective scores.

3. NR-IQA METHOD

This section reviews the NR-IQA learning based methods in recent years. We provide a specific description of these NR-IQA models.

3.1 RankIQA

As the demand for larger and deeper convolutional neural networks for IQA increases, the limited availability of annotated IQA databases poses a challenge. To address this issue, Xialei Liu et al. proposed RankIQA [1], which leverages non-IQA databases without subjective ratings for training. They generated distorted images with varying levels of distortion intensity and assigned quality ranking labels to train a Siamese network capable of ranking the quality of distorted image batches. RankIQA achieved a 5% improvement over the state-of-the-art method on the CSIQ database.

3.2 DBCNN

To address the performance gap between synthetic and authentic distorted image databases, Weixia Zhang et al. proposed the DBCNN [2]. DBCNN incorporates two branches, one for synthetic distortion and the other for authentic distortion, to better handle different types of distortions. The synthetic distortion branch is pre-trained using a large-scale dataset, while the authentic distortion branch utilizes a pre-trained VGG-16 model. The extracted features from both branches are fused using bilinear pooling and fed into the fully connected layer for quality prediction. DBCNN achieves state-of-the-art performance on both synthetic and authentic IQA databases. The addition of a dedicated branching model for authentic distortion and feature fusion contribute to the improved performance, and further customization of the true distortion model branches is expected to enhance the model even more.

3.3 HyperIQA

To address the challenge of IQA for real distorted images, a novel approach called HyperIQA [3] was proposed. It consists of semantic feature extraction, perceptual rule building, and quality prediction steps. By leveraging a hyper network, the model generates weights and biases for the quality prediction module based on high-level semantic features, which improves the accuracy of quality assessment by considering the overall content of the image. HyperIQA incorporates multi-scale fusion of features to capture more distortion information and prevents confusion between different image content types. Experimental results demonstrate the effectiveness of the method, as the hyper network generates consistent weights for images with similar content categories, enhancing adaptability and accuracy in evaluating image quality across diverse images.

3.4 TReS

To address the limitations of convolutional neural networks (CNNs) in capturing non-local feature relationships, Alireza Golestaneh et al. proposed TReS [4], which replaces CNNs with a vision transformer (ViT) as the feature extractor in NR-IQA. TReS performs feature fusion using the output of each bottleneck of ViT and employs a transformer encoder to model non-local dependencies among the extracted multi-scale features. Additionally, the authors introduced the concept of learning to rank using Triplet loss, aiming to differentiate image quality details by making the scores of the best-quality images closer to those of the second-best images and further from the worst-quality images. This approach is particularly effective for images with reasonable distortion levels commonly found in IQA databases, as it enables better representation learning and detail differentiation.

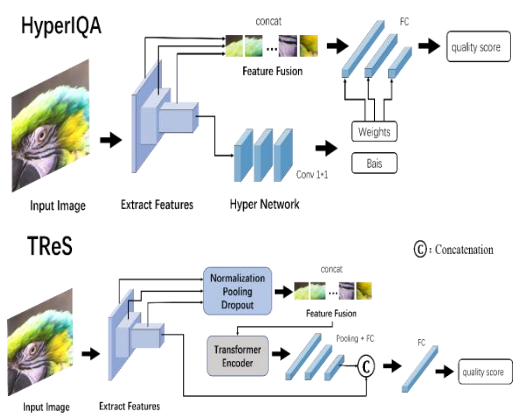

Figure. 2 Comparison of NR-IQA methods (HyperIQA and TReS)

3.5 MANIQA

To address the challenge of handling fine and complex textures in image recovery using GAN-based algorithms, Sidi Yang et al. proposed the MANIQA [5]. MANIQA utilizes an attention block to capture feature dependencies and improve the accuracy of model evaluation. The authors employ a Transposed Attention Block, consisting of a multi-headed self-attentive module, to enhance information interaction in the channel dimension. They also incorporate a Scale Swin Transformer Block to facilitate spatial attention and increase the interaction between different regions of global and local images. The final patch-weighted quality prediction branch generates quality scores and corresponding weights for each patch, which are aggregated to obtain the final quality scores. MANIQA's network architecture largely eliminates CNNs and achieves state-of-the-art performance through extensive utilization of self-attentive mechanisms. Moreover, MANIQA adopts a similar approach to HyperIQA, using feature information to generate weights for patch-based quality prediction.

4. FUTURE TRENDS

NR-IQA has made tremendous progress over the past decade again, and some NR-IQA models are already well on their way to scoring distorted images similarly to the human visual system, but there are still many exciting challenges that need to be addressed. In this section, we address some of the hard problems and enhancements in the field of NR-IQA.

4.1 Dataset Making

NR-IQA models currently rely on identifying distortion features and weighting them separately, which limits their ability to achieve human-like image quality assessment. To address this, providing additional information about the details and reasons for ratings in IQA datasets can aid models in learning human image quality rating. Additionally, improving the performance and generalization of NR-IQA methods requires more diverse and extensive training data covering various scenes, distortion types, and image features.

4.2 Considering Scene Perception

The NR-IQA method usually performs image quality assessment under the condition of no scene information. However, there is a close relationship between image quality and application scenes. Introducing scene information into the model and considering the quality characteristics of images in specific scenes can improve the performance of the model in practical applications.

5. CONCLUSION

NR-IQA is a vital task for evaluating image quality without reference images. Deep learning methods have shown significant progress in NR-IQA, but practical application remains challenging. There is a growing interest in NR-IQA, aiming to provide reliable evaluation in various domains such as image enhancement, compression, mobile applications, video communication, and retrieval. This paper reviews classical and novel NR-IQA methods based on deep learning, highlighting their ability to learn quality features and accurately assess reference-free images. Evaluation metrics like SROCC and PLCC are discussed, emphasizing the dataset's impact on method performance. The overview summarizes the current status, development trends, and future research directions, emphasizing algorithm improvement and optimization to enhance model performance and broaden application scenarios, advancing image quality assessment and computer vision.

REFERENCES

[1] Liu, Xialei, Joost Van De Weijer, and Andrew D. Bagdanov. "Rankiqa: Learning from rankings for no-reference image quality assessment." Proceedings of the IEEE international conference on computer vision. 2017.
[2] Zhang, Weixia, et al. "Blind image quality assessment using a deep bilinear convolutional neural network." IEEE Transactions on Circuits and Systems for Video Technology 30.1 (2018): 36-47.
[3] Su, Shaolin, et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
[4] Golestaneh, S. Alireza, Saba Dadsetan, and Kris M. Kitani. "No-reference image quality assessment via transformers, relative ranking, and self-consistency." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
[5] Yang, Sidi, et al. "Maniqa: Multi-dimension attention network for no-reference image quality assessment." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[6] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statisticalevaluation of recent full reference image quality assessmentalgorithms. TIP, 15(11):3440–3451, 2006.
[7] E. C. Larson and D. M. Chandler. Most apparent distortion:full-reference image quality assessment and the role of strategy. Journal of Electronic Imaging, 19(1):011006, 2010.

[8] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian,M. Carli, and F. Battisti. Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. AdvModern Radioelectron, 10:30–45, 2004.

[9] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian,L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, and F. Battisti. Color image database tid2013: Peculiarities and preliminary results. In European Workshop on Visual Information Processing, pages 106–111, 2013.

[10] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. IEEE Transactions on Image Processing, 25(1):372–387,2015.

[11] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k:A large-scale artificially distorted iqa database. In 2019 Eleventh International Conference on Quality of Multimedia Experience, 2019.

[12] Hosu, Vlad, et al. "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment." IEEE Transactions on Image Processing 29 (2020): 4041-4056.